

3rd International Electronic Conference on Metabolomics

15-30 November 2018

chaired by Prof. Peter Meikle, Dr. Thusitha W. Rupasinghe, Prof. Susan Sumner, Dr. Katja Dettmer-Wilde

sponsored by



metabolites

WebSpecmine:

A website for metabolomics data analysis and mining

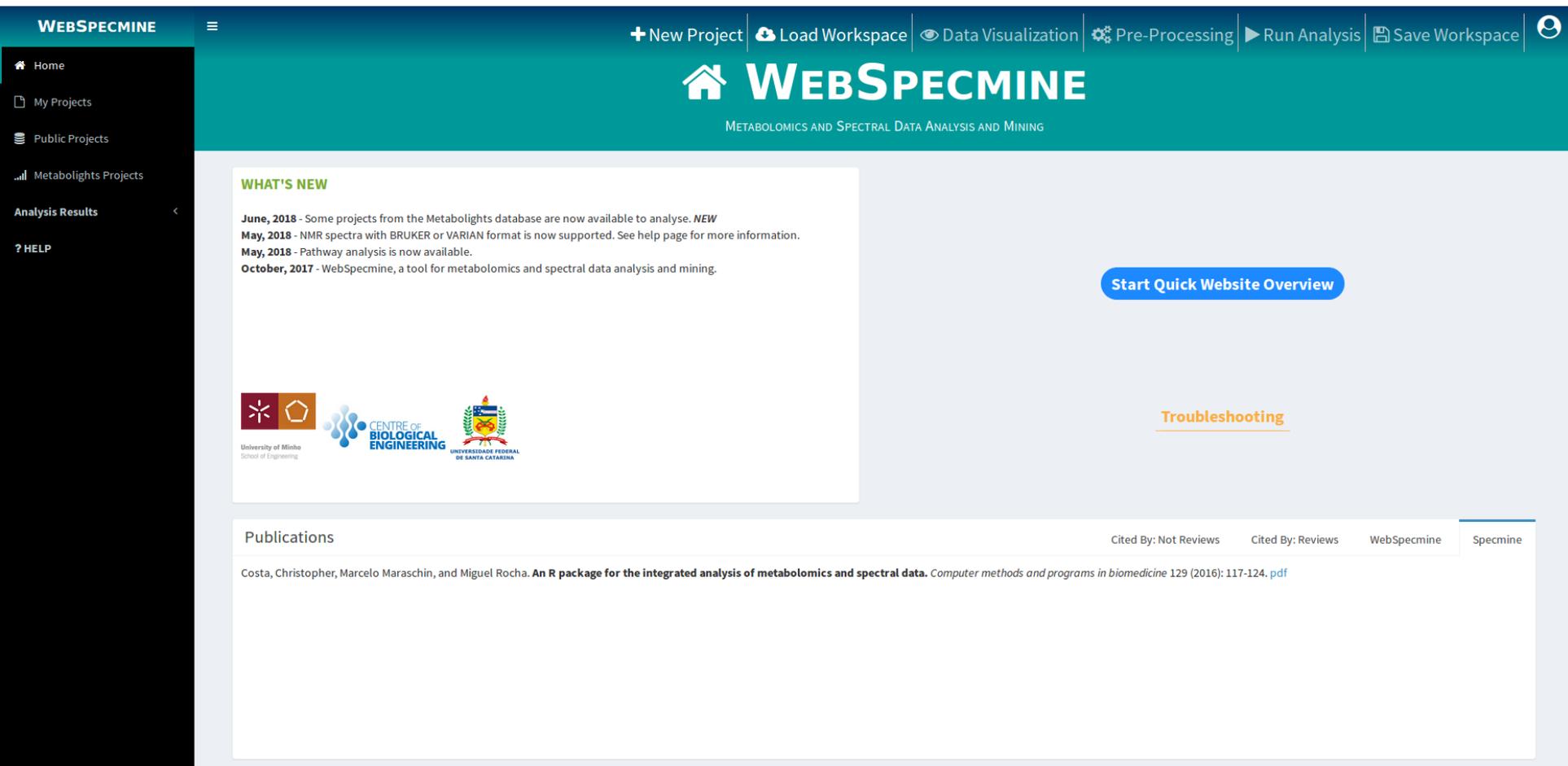
Sara Cardoso^{1,*}, Telma Afonso¹, Marcelo Maraschin², and Miguel Rocha^{1,*}

¹ CEB - Centre Biological Engineering, University of Minho, Campus of Gualtar, Braga, Portugal;

² Plant Morphogenesis and Biochemistry Laboratory, Federal University of Santa Catarina, Florianopolis, SC, Brazil.

* Corresponding authors: saracardoso501@gmail.com; mrocha@di.uminho.pt

WebSpecmine: A website for metabolomics data analysis and mining



WEBSPERMINE

+ New Project | Load Workspace | Data Visualization | Pre-Processing | Run Analysis | Save Workspace

WEBSPERMINE

METABOLOMICS AND SPECTRAL DATA ANALYSIS AND MINING

WHAT'S NEW

June, 2018 - Some projects from the Metabolights database are now available to analyse. *NEW*

May, 2018 - NMR spectra with BRUKER or VARIAN format is now supported. See help page for more information.

May, 2018 - Pathway analysis is now available.

October, 2017 - WebSpecmine, a tool for metabolomics and spectral data analysis and mining.

[Start Quick Website Overview](#)

[Troubleshooting](#)

University of Minho School of Engineering | CENTRE of BIOLOGICAL ENGINEERING | UNIVERSIDADE FEDERAL DE SANTA CATARINA

Publications

Cited By: Not Reviews | Cited By: Reviews | WebSpecmine | Specmine

Costa, Christopher, Marcelo Maraschin, and Miguel Rocha. **An R package for the integrated analysis of metabolomics and spectral data.** *Computer methods and programs in biomedicine* 129 (2016): 117-124. [pdf](#)



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

Introduction

- ✓ Analysing metabolomics data correctly and efficiently is nowadays very important in biological and biomedical research.

However

Most people that want to perform this analysis may not have the programming skills needed

A website to perform metabolomics data analysis is an important asset

Introduction

Some of the Existing Websites



LC/GC-MS Raw Spectra
 LC/GC-MS Peak Lists
 NMR Peak Lists
 Metabolites' Concentrations
 (Quantitative Data)

Univariate Analysis
 T-Test; ANOVA;
 Fold Change

Multivariate Analysis
 PCA; Clustering;
 Machine Learning
 (only PLS-DA);
 Feature Selection
 (only Random
 Forests and SVM)

Other Features
 Correlation
 Analysis;
 Metabolite
 Identification (only
 for MS); Pathway
 Analysis



LC/GC-MS Raw Spectra

T-Test; ANOVA;
 Non-Parametric
 Tests;

PCA; Clustering;
 Machine Learning
 (only LDA, PLS-DA
 and Random
 Forests)

Metabolite
 Identification;
 Pathway Analysis;
 User Account



3rd International Electronic Conference
 on Metabolomics
 15-30 November 2018

sponsors:



metabolites

Introduction

What is missing in the existing websites?

✗ A wide variety of techniques and data formats supported



Spectral Data (Raman, UV-Vis and IR) is missing

✗ A wide variety of pre-processing methods



Mostly just normalization, scaling, missing values treatment

✗ A wide variety of analysis methods



There should be more model options for machine learning, for example

✗ Flexible Pipeline



Most of the time, users have to follow a strict pipeline

✗ User Account



So that data and results can be stored and shared

What was our main goal, then?

- ✓ Create an easy-to-use and freely available website that provides a wide variety of methods and data types for analysis, and ways to store and share metabolomics data and the results generated.

SOLUTION:

 **WEBSPECMINE**

METABOLOMICS AND SPECTRAL DATA ANALYSIS AND MINING



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: overview

Metabolomics data Supported

- ✓ NMR
- ✓ LC/GC-MS
- ✓ Infrared, UV-Visible, and Raman Spectra
- ✓ Concentrations Data (Quantitative Data)

Data Pre-Processing

Metabolomics data Analysis Available

- ✓ Univariate Statistical Analysis
- ✓ Unsupervised Multivariate Statistical Analysis
- ✓ Supervised Multivariate Statistical Analysis
- ✓ Metabolite Identification
- ✓ Pathway Analysis

User Account

- ✓ Store data and results privately
- ✓ Share data across users

Tutorials and User Guide



WebSpecmine: Supported data

LC/GC-MS

Raw Spectra Data Formats

- ✓ .mzData
- ✓ .mzXML
- ✓ .netDF

Peak Lists Data Formats

- ✓ CSV
- ✓ TSV



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Supported data

NMR

Raw Spectra Data Formats

- ✓ BRUKER
- ✓ VARIAN

Peak Lists Data Formats

- ✓ CSV
- ✓ TSV



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Supported data

Spectral Data:
Raman, IR and UV-Vis

Spectra Data Formats

- ✓ CSV
- ✓ (J)DX
- ✓ SPC
- ✓ MS EXCEL (.xlsx)



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Supported data

Concentrations (Quantitative) Data

✓ CSV/TSV File:

Patient ID	1.6-Anhydro-beta-D-glucose	1-Methylnicotinamide	2-Aminobutyrate	2-Hydroxyisobutyrate	2-Oxoqlutarate
PIF_178,	40.85,	65.37,	18.73,	26.05,	71.52
PIF_087,	62.18,	340.36,	24.29,	41.68,	67.36
PIF_090,	270.43,	64.72,	12.18,	65.37,	23.81
NETL_005_V1,	154.47,	52.98,	172.43,	74.44,	1199.91
PIF_115,	22.2,	73.7,	15.64,	83.93,	33.12

Metabolites Names or Identifiers

Samples'
Names

Concentrations values of each
metabolite in each sample



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Supported data

Metadata

- ✓ All types of data should have a metadata file associated
- ✓ CSV/TSV File:

Names of the metadata classes

```
sample, Muscle.loss  
PIF_178, cachexic  
PIF_087, cachexic  
PIF_090, cachexic  
NETL_005_V1, cachexic  
PIF_115, cachexic
```

Samples'
Names

Metadata values for each
metadata class in each sample

WebSpecmine: User Account

Why a User Account?

- ✓ Main website functionalities are accessible without a user account
- ✓ But you will have to create an account if you want to:
 - Save and Share data and results
 - Leave an analysis in 'stand-by'



WebSpecmine: User Account

Creation of a User Account

USER REGISTRATION

To register into our website, you must send an e-mail asking us to create an account for you in WebSpecmine.

We will create the account as soon as possible and send an email with your credentials.

Send e-mail to:
webspecmine@gmail.com

I already have an account: [Login](#)

Close

+ New Project | Load Workspace | Data Visualization | Pre-Processing | Run Analysis | Save Workspace

 **WEBSPECMINE**

METABOLOMICS AND SPECTRAL DATA ANALYSIS AND MINING

User Account Options

Currently not logged in
[Login](#) - or - [Register](#)

To have one, users have to send an email, asking to create an account, and an email with the credentials will be sent as soon as possible.

Email: webspecmine@gmail.com



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: User Account

Data Projects: What is?

A project is a study, or group of studies, which contains the data and metadata for each study, as well as reports from the results obtained

Projects can be:

Private  Public



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: User Account

Your Projects

The projects stored in an account are accessible through **My Projects** sidebar tab

The screenshot displays the 'My Projects' sidebar tab in the WebSpecmine user account. On the left, a 'LIST OF PROJECTS' section shows a table with 8 entries. The 'Propolis' project is selected. Below the table are navigation buttons for 'Create project', 'Delete project', and 'Edit project'. The main content area shows details for the 'Propolis' project, including a 'View project' menu with options like 'Data', 'Metadata', 'Reports', and 'Description'. A 'DATA FOLDERS:' section shows 'NMR Peaks Data' is selected. Below that, a 'FILES IN NMR PEAKS DATA FOLDER:' section lists various CSV files for selection. A 'Data type:' field is set to 'nmr-peaks', and a 'Folder description:' provides a citation for the project.

LIST OF PROJECTS

Show 10 entries Search:

	Project Name
1	Bananas
2	Cachexia
3	Cassava Carotenoids
4	Cassava PPD
5	IP3R in Breast Cancer (MTBLS326 - identified metabolites)
6	Mice Spinal Cord
7	OVCAR-3 (MTBLS152 - identified metabolites)
8	Propolis

Showing 1 to 8 of 8 entries Previous 1 Next

This project is public

View project:

- Data
- Metadata
- Reports
- Description

DATA FOLDERS:

NMR Peaks Data UV-Vis Data 2014 UV-Vis Data 2014_2015

FILES IN NMR PEAKS DATA FOLDER:

Data type: nmr-peaks

Folder description: Marcelo Maraschin et al. Metabolic Profiling and Classification of Propolis Samples from Southern Brazil: An NMR-Based Platform Coupled with Machine Learning. Journal of Natural Products, 2016.

Select all

- AC_au.csv
- AC_sm.csv
- AC_sp.CSV
- AC_wi.csv
- AN_au.csv
- AN_sm.csv
- AN_sp.CSV
- AN_wi.csv
- BR_au.csv
- BR_sm.csv

WebSpecmine: User Account

Public Projects

Everyone that accesses the website can see all public projects, at the **Public Projects** sidebar tab

Community projects

Show entries

Search:

Name	Author	Datatypes
1 Propolis	Admin	nmr-peaks, uvv-spectra
2 Mice Spinal Cord	Admin	lcms-spectra
3 Cachexia	Admin	concentrations
4 Cassava PPD	Admin	ir-spectra
5 Bananas	Admin	nmr-peaks
6 Cassava Carotenoids	Admin	uvv-spectra
7 OVCAR-3 (MTBLS152 - identified metabolites)	Admin	concentrations
8 IP3R in Breast Cancer (MTBLS326 - identified metabolites)	Admin	concentrations

Showing 1 to 8 of 8 entries

Previous Next

[Import Project](#) [Refresh](#)

Project description:

Banana peels are well recognized as a source of important bioactive compounds, such as phenolics, carotenoids, biogenic amines, among others. As such, they have recently started to be used for industrial purposes. However, its composition seems to be strongly affected by biotic or abiotic ecological factors. Thus, this study aimed to investigate banana peels chemical composition, not only to get insights on eventual metabolic changes caused by the seasons, in southern Brazil, but also to identify the most relevant metabolites for these processes. To achieve this, the Nuclear magnetic resonance (NMR) technique was applied. Sara Cardoso et al. A Chemometrics Approach for Nuclear Magnetic Resonance Data to Characterize the Partial Metabolome Banana Peels from Southern Brazil. Journal of integrative bioinformatics 14.4 (2017)

View project files in:

Data Metadata Reports

Data Folders:

NMR Peaks Data

Files in NMR Peaks Data folder:

- Apr2011.csv
- Aug2011.csv
- Feb2011.csv
- July2010.csv
- July2011.csv
- June2011.csv
- Mar2011.csv
- may2011.csv
- Nov2010.csv

[View selected file\(s\)](#)



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: User Account

Public Projects

Everyone that accesses the website can see all public projects, at **Public Projects** sidebar tab

Community projects

Show entries

Search:

Name	Author	Datatypes
1 Propolis	Admin	nmr-peaks, uvv-spectra
2 Mice Spinal Cord	Admin	lcms-spectra
7 OVCA3-3 (MTBLS152 - identified metabolites)	Admin	concentrations
8 IP3R in Breast Cancer (MTBLS326 - identified metabolites)	Admin	concentrations

Showing 1 to 8 of 8 entries

Previous Next

Project description:

Banana peels are well recognized as a source of important bioactive compounds, such as phenolics, carotenoids, biogenic amines, among others. As such, they have recently started to be used for industrial purposes. However, its composition seems to be strongly affected by biotic or abiotic ecological factors. Thus, this study aimed to investigate banana peels chemical composition, not only to get insights on eventual metabolic changes caused by the seasons, in southern Brazil, but also to identify the most relevant metabolites for these processes. To achieve this, the Nuclear magnetic resonance (NMR) technique was applied. Sara Cardoso et al. A Chemometrics Approach for Nuclear Magnetic Resonance Data to Characterize the Partial Metabolome Banana Peels from Southern Brazil. Journal of integrative bioinformatics 14.4 (2017).

NMR Peaks Data

- Feb2011.csv
- July2010.csv
- July2011.csv
- June2011.csv
- Mar2011.csv
- may2011.csv
- Nov2010.csv

To analyse a public project, you would have to copy it to your account and analyse it from there, so that the original project is not compromised.



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: User Account

Workspace:
Users can leave their analysis in
'stand-by' and continue later

Users can leave an analysis at any time, by saving the workspace, and continue next time

Save Workspace Data

All the data and results regarding the current data folder from the current project can be saved on your account, so it can be uploaded at a different time and the analysis continued.

Current Project:

Propolis

Current Data Folder from the project above:

UV-Vis Data 2014

Save Workspace

Load Workspace Data

Choose an available workspace:

Project Name : Data Folder Name

NMR Peaks

bananas : nmr_peaks

Close



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Select Data for Analysis

For Logged In Users

1

Select the Project, the data folder from that project where the data to analyse is, and the metadata file from that project that corresponds to the data selected

Choose Files for Analysis

PROJECT

Choose the project where the data to analyse is:

- Bananas
- Cachexia
- Cassava Carotenoids
- Cassava PPD
- IP3R in Breast Cancer (MTBLS326 - identified metabolites)
- Mice Spinal Cord
- OVCAR-3 (MTBLS152 - identified metabolites)
- Propolis

DATA FOLDER

Choose the data folder that has the data files to analyse:

- NMR Peaks Data
- UV-Vis Data 2014
- UV-Vis Data 2014_2015

DATA TYPE: nmr-peaks

METADATA FILE

Choose the file with the metadata information of the data folder selected:

- propolis_nmr_metadata.csv
- propolis_uvv_metadata_2014_2015
- propolis_uvv_metadata_2014.csv

> Next

Close

WebSpecmine: Select Data for Analysis

For Logged In Users

1

2

Set the options required to correctly read the data and metadata files

Choose Files for Analysis

OPTIONS

DATA FILES OPTIONS

Data files have a header row with the names of the data variables

Separator character of the data files

Comma
 White Space

Character used in data files for decimal points

Dot
 Comma

METADATA FILE OPTIONS

Metadata file has a header column with the name of the metadata variables

Metadata file has a header row with the name of the samples

Separator character of the metadata file

Comma
 White Space

OPTIONAL INFORMATION:

Short description of the data

Short label for the x values

< Previous

Next >

Close



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Select Data for Analysis

For Logged In Users

1

2

3

After finishing the setting of data and metadata options, the user can submit the data for analysis

Choose Files for Analysis

OPTIONS

ALIGNMENT OF PEAKS OPTIONS

There are two methods available to perform alignment of peaks. The specmine algorithm does not allow overlapping of windows, being the size of the window equal to the step. The MetaboAnalyst method allows overlapping of windows, being the step half the size of the window. The step size for the MetaboAnalyst method has a default of 0.015 for NMR peaks and 0.125 for GC/LC-MS peaks. The bandwidth, used in this method, has the values 10, 30 and 5 for NMR, LC/MS and GC/MS peaks, respectively.

Method:

Specmine Algorithm

MetaboAnalyst Algorithm

Size of the step, in ppms:

0.03

< Previous

Submit For Analysis

Close



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Select Data for Analysis

For Logged Out Users

The screenshot shows a web interface titled '+ New Project' with a close button (x) in the top right corner. Below the title is a navigation bar with five tabs: 'MS Spectra', 'NMR Spectra', 'NMR or MS peaks lists' (which is selected), 'UV-vis, IR or Raman spectra', and 'Concentrations'. The main content area is divided into three sections: 'DATA OPTIONS', 'METADATA OPTIONS', and 'OPTIONAL INFORMATION:'.
1. 'DATA OPTIONS': Includes a 'Data Folder' section with a 'Browse...' button and 'No file selected' text. Below it is 'Type of the data:' with radio buttons for 'NMR Peaks' (selected), 'LC-MS Peaks', and 'GC-MS Peaks'. There is also a checked checkbox for 'Data files have a header row with the names of the data variables'. At the bottom of this section is 'Separator character of the data files'.
2. 'METADATA OPTIONS': Includes a 'Metadata File' section with a 'Browse...' button and 'No file selected' text. Below it are two checked checkboxes: 'Metadata file has a header column with the name of the metadata variables' and 'Metadata file has a header row with the name of the samples'. At the bottom of this section is 'Separator character of the metadata file' with radio buttons for 'Comma' (selected) and 'White Space'.
3. 'OPTIONAL INFORMATION:': Includes a 'Short description of the data' text input field and a 'Short label for the x values' text input field.
At the bottom center of the form is a green button with a right-pointing arrow and the text '> Next'. At the bottom right is a 'Close' button.

- ✓ The Procedure is similar, but the data files and metadata files have to be submitted, as they are not stored in the website.
- ✓ The data submitted will only be temporarily stored, while the analysis is in action.

Once the user selects the data, the data analysis pages will be accessible



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Visualization

1

The website provides a way to visualize the data

WEBSPECMINE | Choose Files | Load Workspace | **Data Visualization** | Pre-Processing | Run Analysis | Save Workspace

DATA VISUALIZATION

Data Summary

- Data Table
- Metadata Table
- Samples' Statistics
- Variables Statistics
- Boxplots of the Variables
- Spectra Plot

Dataset being used

OriginalData

Analysis Results

? HELP

Dataset Visualization Report (html):

Download Save

The data you are exploring in this tab is the data selected in the sidebar section 'Dataset being used'.

If a metadata variable is not available to choose in the boxplots and/or spectra plots, it means that it needs to be converted to a factor (Pre-Processing page).

Dataset summary:
Valid dataset
Description:
Type of data: uvv-spectra
Number of samples: 165
Number of data points 521
Number of metadata variables: 5
Label of x-axis values:
Label of data points:
Number of missing values in data: 0
Mean of data values: 0.1759754
Median of data values: 1e-04
Standard deviation: 0.5662896
Range of values: 0 4.499
Quantiles:
0% 25% 50% 75% 100%
0.0000 0.0001 0.0001 0.0220 4.4990



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Visualization

1

2

Data Summary

Data Summary

Data Table

Metadata Table

Samples' Statistics

Variables Statistics

Boxplots of the Variables

Peaks Plot

Dataset summary:
Valid dataset
Description:
Type of data: nmr-peaks
Number of samples: 13
Number of data points 173
Number of metadata variables: 1
Label of x-axis values: ppm
Label of data points: Intensity
Number of missing values in data: 795
Mean of data values: 0.02179374
Median of data values: 0.0019
Standard deviation: 0.04556014
Range of values: 1e-04 0.3023
Quantiles:

0%	25%	50%	75%	100%
0.000100	0.000700	0.001900	0.010875	0.302300

Dataset Visualization Report (html):

Download

Save

The data you are exploring in this tab is the data selected in the sidebar section 'Dataset being used'.

If a metadata variable is not available to choose in the boxplots and/or spectra plots, it means that it needs to be converted to a factor (Pre-Processing page).



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Visualization

1

2

3

Data and Metadata Tables

The screenshot displays the WebSpecmine interface. On the left is a sidebar with navigation options: Data Summary, Data Table, Metadata Table, Samples' Statistics, Variables Statistics, Boxplots of the Variables, and Peaks Plot. The main area is divided into two sections. The top section, titled 'Data Table of OriginalData dataset', contains a table with columns for various months (Apr2011, Aug2011, Feb2011, July2010, July2011, June2011, Mar2011, may2011, Nov2010, Oct2010, Oct2011, Sept2010, Sept2011) and rows of numerical data. The bottom section, titled 'Metadata Table of OriginalData dataset', contains a table with a 'Seasons' column and rows for the same months. A search bar is located at the top right. A 'Dataset Visualization Report (html)' section with 'Download' and 'Save' buttons is present in the bottom left. A tooltip menu is overlaid on the 'Data Table' sidebar option, listing all navigation options. A red text note at the bottom left explains that unavailable metadata variables in boxplots or spectra plots need to be converted to factors.

Search:

Data Table of OriginalData dataset.

	Apr2011	Aug2011	Feb2011	July2010	July2011	June2011	Mar2011	may2011	Nov2010	Oct2010	Oct2011	Sept2010	Sept2011
0.74					0.0001								
0.89	0.0013	0.0019	0.001	0.0004	0.0011	0.0012	0.0017	0.0005		0.001	0.001	0.0014	
0.92	0.0005	0.0014	0.0007	0.0006	0.002	0.0009	0.0022	0.0006		0.0003	0.0006	0.001	
0.96	0.0001												
0.99	0.0001												
1.02													
1.05	0.0001												
1.08													
1.14	0.0001												
1.17	0.0001												
1.2	0.0001												
1.23	0.0001												
1.25	0.0001												

Metadata Table of OriginalData dataset.

	Seasons
Apr2011	Summer
Aug2011	Autumn
Feb2011	Summer
July2010	Autumn
July2011	Autumn
June2011	Autumn
Mar2011	Summer
may2011	Summer

Dataset Visualization Report (html):

[Download](#) [Save](#)

The data you are exploring in this tab is the data selected in the sidebar section 'Dataset being used'.

If a metadata variable is not available to choose in the boxplots and/or spectra plots, it means that it needs to be converted to a factor (Pre-Processing page).

Showing 1 to 17

WebSpecmine: Data Visualization

1

2

3

4

Samples' and Variables' Statistics

Data Summary

Data Table

Metadata Table

Samples' Statistics

Variables Statistics

Boxplots of the Variables

Peaks Plot

Search:

Samples' Statistics Table of OriginalData dataset.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
Apr2011	0.0003	0.0009	0.0026	0.0314876288659794	0.0414	0.2515	76
Aug2011	0.0001	0.0006	0.00165	0.0149777777777778	0.005125	0.1948	65
Feb2011	0.0001	0.0006	0.0018	0.0225092307692308	0.0073	0.2565	43
July2010	0.0003	0.000					
July2011	0.0001	0.000					
June2011	0.0002	0.000					
Mar2011	0.0003	0.001					
may2011	0.0002	0.0006					
Nov2010	0.0007	0.001					
Oct2010	0.0002	0.000					
Oct2011	0.0001	0.000					
Sept2010	0.0003	0.000					
Sept2011	0.0001	0.000					

Showing 1 to 13 of 13 entries

Data Summary

Data Table

Metadata Table

Samples' Statistics

Variables Statistics

Boxplots of the Variables

Peaks Plot

Variables' Statistics Table of OriginalData dataset.

	Min.	1st Qu.	Median
X0.74	0.0001	0.0001	0.0001
X0.89	0.0004	0.001	0.000
X0.92	0.0003	0.0006	0.000
X0.96	0.0007	0.000875	0.000
X0.99	0.0006	0.0006	0.000
X1.02	0.0004	0.00055	0.000
X1.05	0.0003	0.0006	0.000
X1.08	0.0004	0.000525	0.000

Dataset Visualization Report (html):

[Download](#) [Save](#)

The data you are exploring in this tab is the data selected in the sidebar section 'Dataset being used'.

Dataset Visualization Report (html):

[Download](#) [Save](#)

The data you are exploring in this tab is the data selected in the sidebar section 'Dataset being used'.

If a metadata variable is not available to choose in the boxplots and/or spectra plots, it means that it needs to be converted to a factor (Pre-Processing page).



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Visualization

1

2

3

4

5

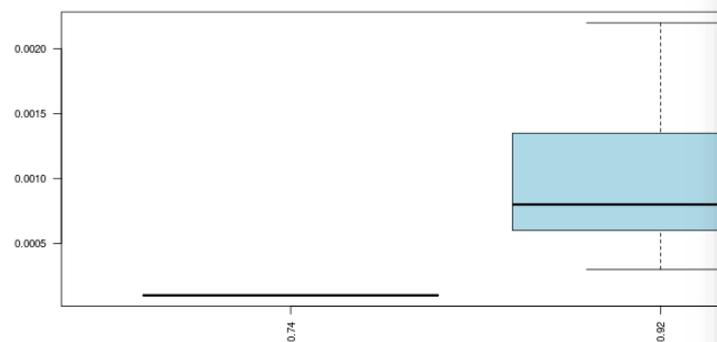
Boxplots of the Variables

One or more variables

One data variable over a metadata variable

Choose the variables to be shown in the boxplot

0.74, 0.92, 1.02



One or more variables

One data variable over a metadata variable

Choose the variable to be shown in the boxplot

1.02

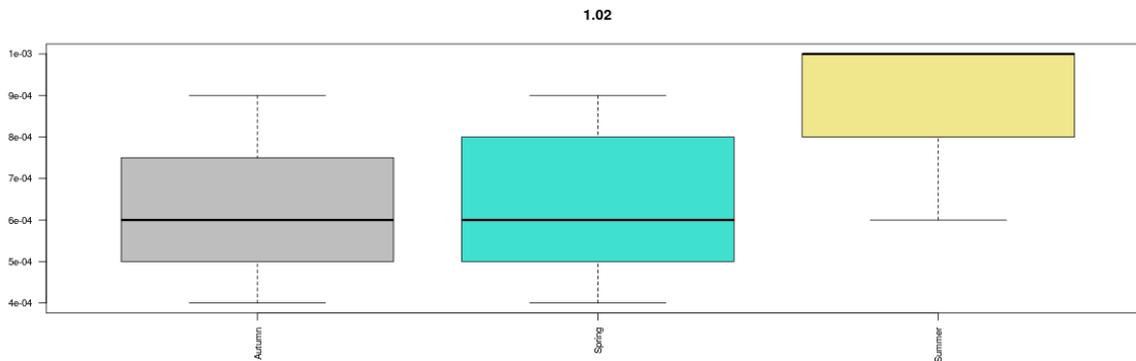
Choose the metadata variable to plot

Seasons

choose the 3 colors for each class of metadata variable chosen, to color the corresponding box

Gray, Turquoise, Khaki

Construct Plot



WebSpecmine: Data Visualization

1

2

3

4

5

6

Plot for Peaks Data

Data Summary

Data Table

Metadata Table

Samples' Statistics

Variables Statistics

Boxplots of the Variables

Peaks Plot

Metadata variable to color the plot:

Seasons

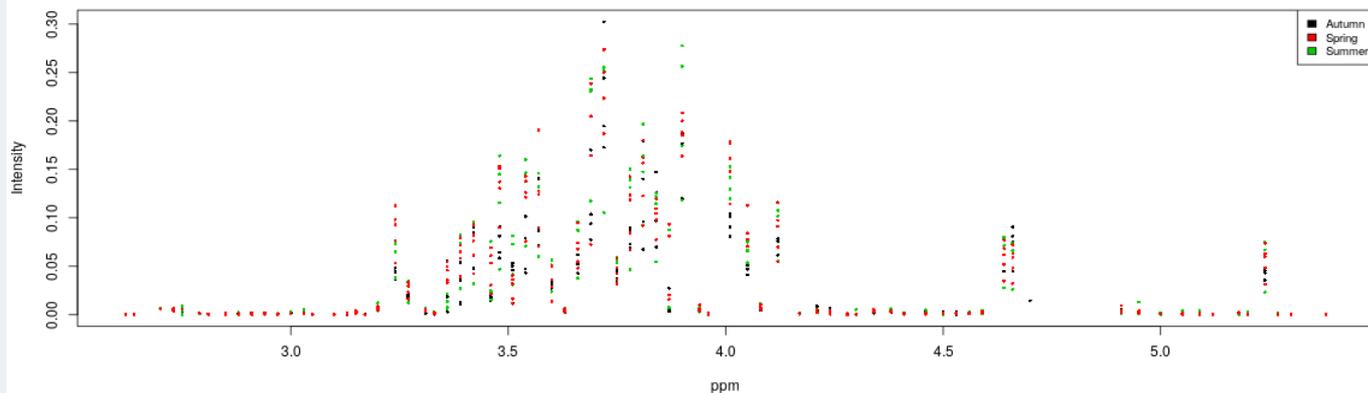
Show axis range



Select samples to appear on the plot:

Show all samples.

Reverse x-axis



Dataset Visualization Report (html):

Download

Save

The data you are exploring in this tab is the data selected in the sidebar section 'Dataset being used'.

If a metadata variable is not available to choose in the boxplots and/or spectra plots, it means that it needs to be converted to a factor (Pre-Processing page).



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Visualization

1

2

3

4

5

6

Plot for Spectra

Data Summary

Data Table

Metadata Table

Samples' Statistics

Variables Statistics

Boxplots of the Variables

Spectra Plot

Metadata variable to color the plot:

seasons

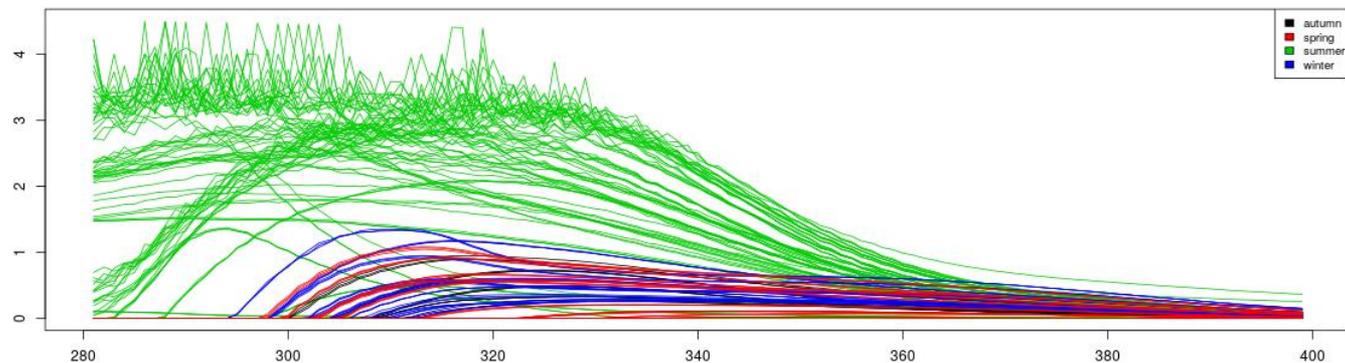
Select samples to appear on the plot:

Show all samples.

Show axis range



Reverse x-axis



Dataset Visualization Report (html):

Download

Save

The data you are exploring in this tab is the data selected in the sidebar section 'Dataset being used'.

If a metadata variable is not available to choose in the boxplots and/or spectra plots, it means that it needs to be converted to a factor (Pre-Processing page).



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Pre-Processing

1

The website provides a wide variety of pre-processing methods, that can be performed in the desired order

WEBSPECMINE

Choose Files | Load Workspace | Data Visualization | **Pre-Processing** | Run Analysis | Save Workspace

PRE-PROCESSING

Missing Values

Your data has no missing values.

Data Transformation

Choose one method to transform the data:

Logarithmic
 Cubic Root

Transform

Scaling

Choose one method to scale the data:

Auto
 Pareto
 Range

Scale

Detect NMR Peaks

Not applicable to this data type.

Subset Dataset

Subset by Variables | Subset by samples

Subset by interval of data values
 Subset by specific data values

Choose the variable range to keep on the new dataset:

280 800

Select the data variables to keep on the new dataset

280

Subset



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Pre-Processing

1

2

Methods that are available for all types of data

Missing Values

You have 795 missing values in your dataset. Choose one of these methods to treat the values:

- Mean
- Value given
- Median
- K-Nearest Neighbours
- Linear approximation

Value:

0.00005

Impute Missing Values

Scaling

Choose one method to scale the data:

- Auto
- Pareto
- Range

Scale

Data Normalization

Choose one method to normalize the data:

- Sum
- Median

Constant:

1000

Normalize

Data Transformation

Choose one method to transform the data:

- Logarithmic
- Cubic Root

Transform

Flat Pattern Filter

Choose one Filter Function:

- Interquartile Range
- Relative Standard Deviation
- Standard Deviation
- Median Absolute Deviation
- Mean
- Median

Choose how to filter the values:

- Percentage
- Threshold
- Calculate automatically number of variables to remove

Choose the percentage of the number of variables to filter:



Filter



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Pre-Processing

1

2

Methods that are available for all types of data

Mean-Centering

Mean-Center

Remove data by NAs

Remove:

Samples Data variables

According to what do you want to remove samples?

- Number of NAs in samples
 Percentage of NAs in samples
 NAs in metadata

Insert the maximum percentage of NAs that a sample can have:



Remove

Remove data

Remove:

Samples Data variables Metadata variables

Choose the sample(s) to remove:

Remove

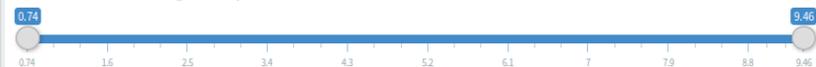
Subset Dataset

Subset by Variables

Subset by samples

- Subset by interval of data values
 Subset by specific data values

Choose the variable range to keep on the new dataset:



Select the data variables to keep on the new dataset

0.74

Subset



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Pre-Processing

1

2

Methods that are available for all types of data

Aggregate samples

Samples can be aggregated according to the classes of a certain metadata variable. Samples in the same class will be aggregated together.

Choose the metadata variable by which samples will be aggregated:

Seasons

Aggregate samples' values by:

Mean Median Sum Maximum value Minimum value

Metadata variables to remove when aggregating the samples, if wanted. If not wanted, do not select any option:

No metadata variables will be removed

Aggregate

Low-level data fusion

Only the samples from the new data provided that have the same name as samples in the current dataset will be joined.

MS Spectra NMR Spectra NMR or MS peaks lists UV-vis, IR or Raman spectra Concentrations

Note that only the formats .mzXML, .netCDF, mzData are supported.

When reading the data, the peak detection will be performed.

Data Folder

Browse... No file selected

Type of the data:

LC-MS Spectra GC-MS Spectra

Options for the feature (peak) detection in the chromatographic time domain:

Join With Current Dataset



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Pre-Processing

1

2

3

Methods only for spectral data

Correction

Choose the correction to use on your spectral data:

- Baseline
- Offset
- Background

Correct

First derivative

Apply

Multiplicative Scatter Correction

Correct

Smoothing interpolation

Choose the smoothing interpolation type

- Bin
- Loess
- Savitzky-Golay

Reducing factor:

2

Apply



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Pre-Processing

1

2

3

4

Method only for NMR Spectra

Detect NMR Peaks

Baseline threshold value (Minimum value to consider a peak):

50000

Options to align peaks after their detection:

There are two methods available to perform alignment of peaks. The specmine algorithm does not allow overlapping of windows, being the size of the window equal to the step. The MetaboAnalyst method allows overlapping of windows, being the step half the size of the window. The step size for the MetaboAnalyst method has a default of 0.015 for NMR peaks and 0.125 for GC/LC-MS peaks. The bandwidth, used in this method, has the values 10, 30 and 5 for NMR, LC/MS and GC/MS peaks, respectively.

Method:

- Specmine Algorithm
 MetaboAnalyst Algorithm

Size of the step, in ppms:

0.03

Detect Peaks

WebSpecmine: Pre-Processing

1

2

3

4

5

After processing the data, a name to the new dataset has to be given

Name for the new dataset

Write the name you would like to give to the processed dataset, without spaces.

Finish

i To perform an analysis on the new dataset, the user will have to choose it on the sidebar panel

WebSpecmine: Data Analysis

WEBSPECMINE | Choose Files | Load Workspace | Data Visualization | Pre-Processing | **Run Analysis** | Save Workspace

▶ RUN ANALYSIS

To Start the analysis of your Metabolomic Data, choose one of the analysis boxes below.
Boxes in grey represent unavailable boxes.
(This occurs when the dataset data type is unsupported or the dataset has missing values (treat them on "Pre-Processing" tab)).

- Univariate Analysis**
 - T-Test
 - One-way and multifactor ANOVA
 - Kruskal-Wallis and Komolgorov-Smirnov tests
 - Fold Change analysis[Univariate Analysis](#)
- Principial Component Analysis (PCA)**
 - Perform principal component analysis
 - Both classical and robust approaches available[PCA](#)
- Clustering Analysis**
 - Two types of clustering analysis available:
 - Hierarchical Clustering
 - K-Means Clustering[Cluster Analysis](#)
- Machine Learning**
 - Train models with the data available.
 - Predict new samples with the models trained previously or a model saved in user's account.[Machine Learning](#)
- Feature Selection**
 - There are two methods available for Feature Selection:
 - Recursive Feature Elimination.
 - Selection by Filter[Feature Selection](#)
- Metabolite Identification**
 - Identification of metabolites only available for datasets obtained from the following techniques:
 - LC-MS technique
 - NMR Peaks[Metabolite Identification](#)
- Regression Analysis**
 - Available analysis:
 - Regression analysis
 - Correlation analysis[Regression Analysis](#)
- Pathway Analysis**
 - Available for:
 - Metabolites identified through 'Metabolite Identification' box
 - Concentrations data whose variables names are in HMDB OR KEGG codes[Pathway Analysis](#)



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Univariate Analysis

Example for T-Test

T-Test

- One-Way Analysis Of Variance (ANOVA)
- Multi-Factor Analysis Of Variance (ANOVA)
- Kruskal-Wallis Test
- Kolmogorov-Smirnov Test
- Fold Change Analysis

T-Test

Give a name to the analysis:

Select the metadata variable to use:

P-value threshold

Submit

 Analysis options for a T-Test

WebSpecmine: Data Analysis

Univariate Analysis

Example for T-Test

WebSpecmine interface showing T-Test Analysis results. The interface includes tabs for "Numerical Results" and "Plot". The "T-Test Analysis" section contains buttons for "Download CSV", "Download HTML Report", "Save CSV", and "Save HTML Report". A search bar is present. The results are displayed in a table with columns for "p.value", "-log10", and "fdr". The table shows 10 entries, with the first entry having a p-value of 1.28876620629546e-7, -log10 of 6.88982586040062, and fdr of 0.000181152093484064.

Showing 1 to 10 of 50 entries

Previous 1 2 3 4 5 Next

i Types of results available for this type of analysis: numerical results



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:

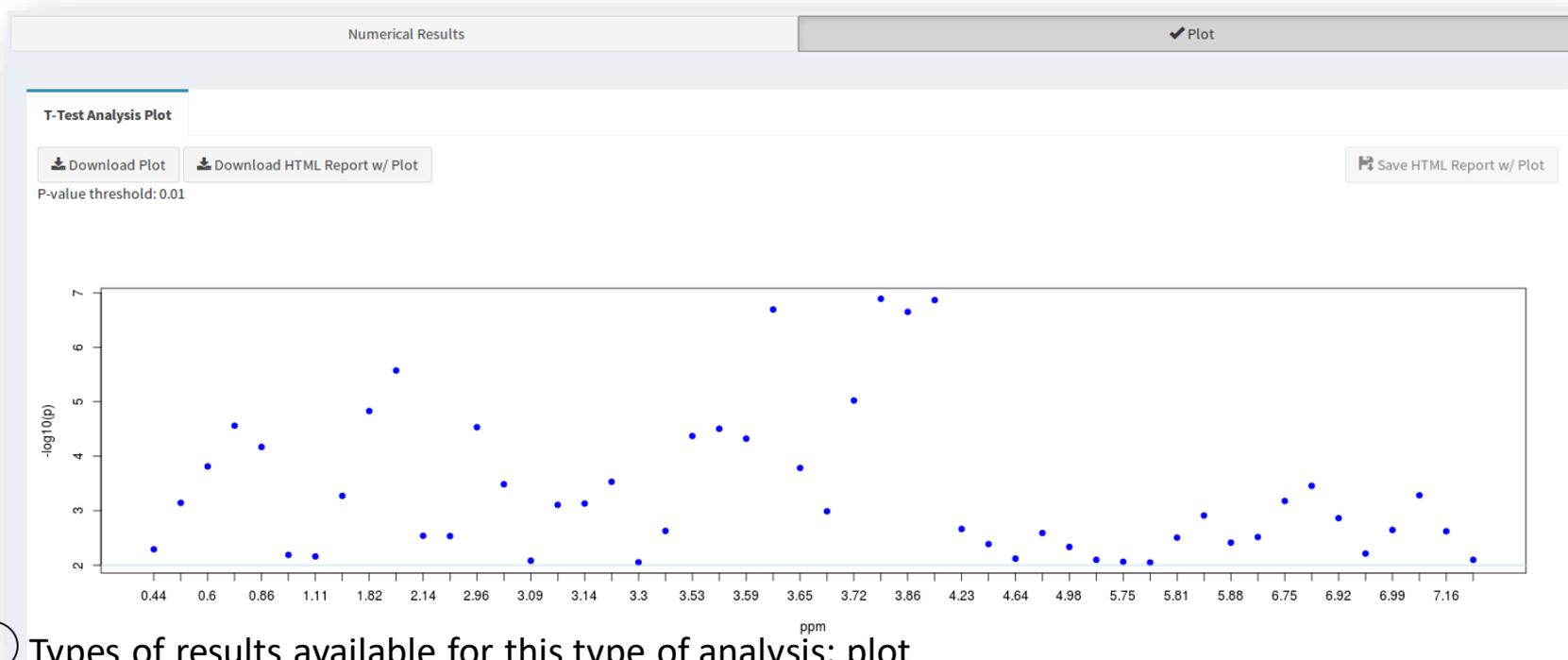


metabolites

WebSpecmine: Data Analysis

Univariate Analysis

Example for T-Test



i Types of results available for this type of analysis: plot

WebSpecmine: Data Analysis

Univariate Analysis

Other Analysis

T-Test

One-Way Analysis Of Variance (ANOVA)

Multi-Factor Analysis Of Variance (ANOVA)

Kruskal-Wallis Test

Kolmogorov-Smirnov Test

Fold Change Analysis

① There are other Univariate Analysis methods available

① The types of results available for each analysis is similar to those showed for T-Test



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Principal Components Analysis (PCA)

Normal PCA

Robust PCA

Normal PCA

Robust PCA

Robust PCA

Give a name to the analysis:

robust_PCA

Center method:

Mean Median

Scale method:

Standard deviation ratio

Mean absolute deviation

Number of components:

10

Submit

Normal PCA

Give a name to the analysis:

PCA

Scale variables

Center variables

Submit

 Analysis options for both normal and robust PCAs

WebSpecmine: Data Analysis

Principal Components Analysis (PCA)

WebSpecmine interface showing PCA results. The interface includes tabs for Numerical Results, Make plots, and Visualize plots. The Numerical Results tab is active, displaying the Component Importance section.

Component Importance: Scores Matrix, Variable Loadings

Download CSV, Download HTML Report, Save CSV, Save HTML Report

(The Report includes results from all three tabs)

Show 10 entries Search:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
Standard deviation	7.2859	5.3192	4.3309	3.8682	3.4049	3.2045	2.7954	2.7161	2.6459	2.3899	2.2180	1.7933	0.0000
Proportion of Variance	0.3068	0.1636	0.1084	0.0865	0.0670	0.0594	0.0452	0.0426	0.0405	0.0330	0.0284	0.0186	0.0000
Cumulative Proportion	0.3068	0.4704	0.5788	0.6653	0.7323	0.7917	0.8368	0.8795	0.9200	0.9530	0.9814	1.0000	1.0000

Showing 1 to 3 of 3 entries Previous 1 Next

i Types of results available for this type of analysis: numerical results



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Principal Components Analysis (PCA)

Numerical Results Make plots Visualize plots

Scree K-means Pairs Scores Plot 3D Scores Plot 2D K-means Plot 2D Biplot **PLOT TYPE**

Give a name to the plot:

Number of components:

Relative font size of legend:

Legend position (colors of each class on the metadata variable):

"Individual percent" color:

"Cumulative percent" color:

Your plot will be displayed in the corresponding "Analysis results" tab for the selected PCA result.

i Types of results available for this type of analysis: plot results



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

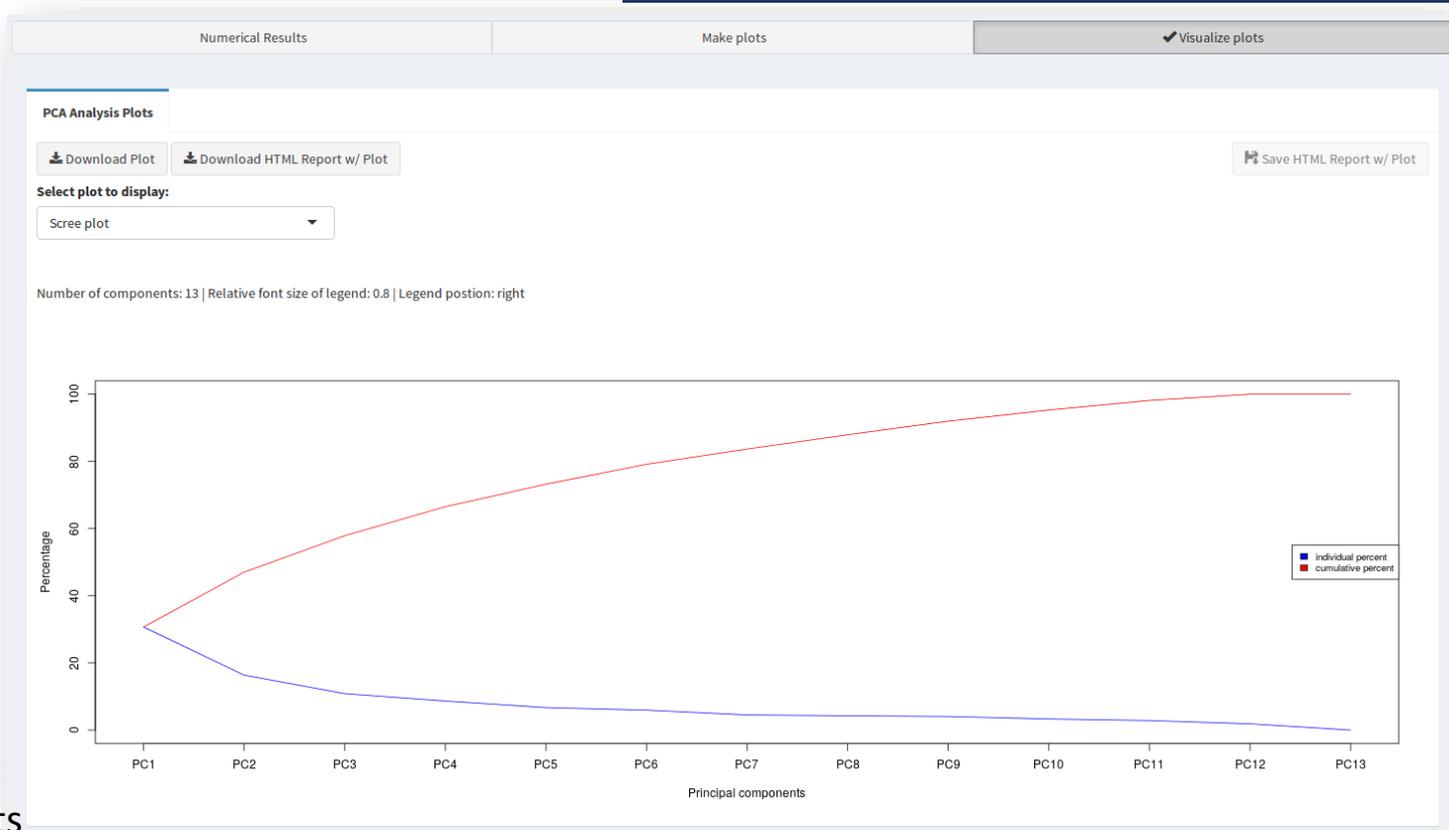
sponsors:



metabolites

WebSpecmine: Data Analysis

Principal Components Analysis (PCA)



i Types of results available for this type of analysis: plot results

WebSpecmine: Data Analysis

Clustering Analysis

Hierarchical Clustering

Hierarchical Clustering

K-Means Clustering

 Analysis options

Hierarchical Clustering

Give a name to the analysis:

Hier_clustering

Distance measure

Euclidean Manhattan Pearson Spearman

Agglomeration method

Complete Ward Single Average Mcquitty Median Centroid

Hierarchical cluster analysis on

Samples Variables

Submit



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Clustering Analysis

Hierarchical Clustering

Numerical Results

Dendrogram

Hierarchical Clustering Analysis

Download CSV Download HTML Report w/ plot

Save CSV Save HTML Report

To be able to download a report, a dendrogram plot must have been generated in 'Dendrogram' section, accessible above.

Heights

[1]	244.5360	292.9697	346.4199	429.7078	451.2250	569.9818
[7]	674.3466	676.6284	793.1968	843.1993	893.5235	983.4365
[13]	973.8300	1024.2079	1060.4265	1109.7344	1148.0559	1202.4623
[19]	1394.7463	1429.7784	1524.0785	1590.5965	1595.2086	1758.0043
[25]	1825.3550	1880.2430	1882.7233	1904.9573	1915.3433	2001.0597
[31]	2055.3826	2100.3535	2208.0812	2282.5717	2299.1636	2309.8700
[37]	2340.5440	2341.1069	2448.1359	2623.0522	2730.1312	2852.9477
[43]	2929.5713	3187.6903	3290.5813	3306.7768	3313.9506	3515.2404
[49]	3663.4035	3807.9177	3873.4079	3983.4114	4121.1914	4453.5449
[55]	4504.7492	4923.0165	5055.4852	5177.5019	5326.3170	5544.8603
[61]	6777.5414	6913.3660	7056.8954	7353.1417	7805.7842	8310.4664
[67]	8432.6939	9799.1861	10246.4801	10486.6391	12715.7338	14523.0217
[73]	15771.4438	20934.6178	22019.2733	35438.4962		

Numerical Results

Dendrogram

Hierarchical Clustering Analysis Dendrogram

Select the metadata variable to color the leaves:

Muscle.loss

Relative font size of samples names:

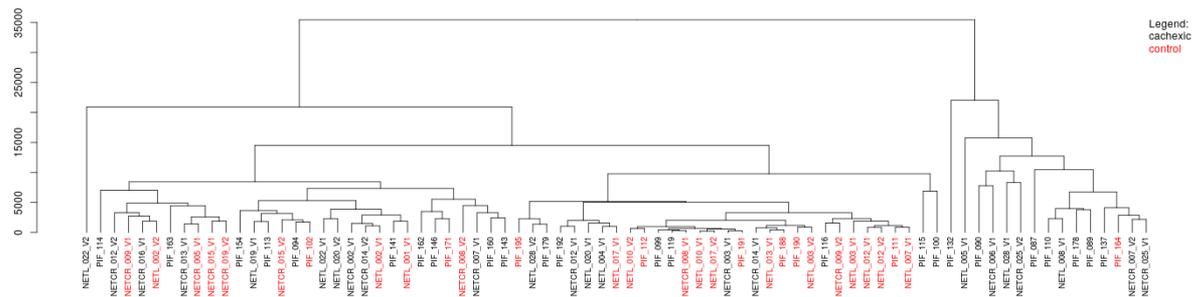
0.8

Name the leafs by:

- Names of the respective samples
 Metadata value of the respective samples

Plot

Download Dendrogram



Types of results available for this type of analysis



3rd International Electronic Conference on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Clustering Analysis

K-Means Clustering

Hierarchical Clustering

K-Means Clustering

 Analysis options

K-means Clustering

Give a name to the analysis:

K-Means_Clustering

K-means cluster analysis on

Samples Variables

Number of clusters

2

Submit



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Clustering Analysis

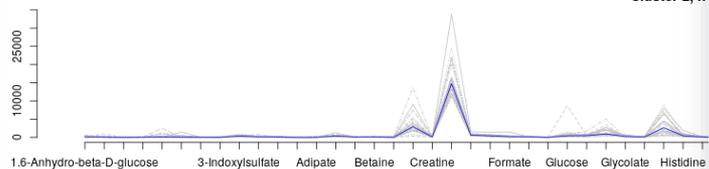
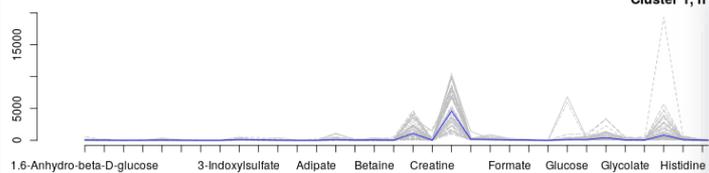
K-Means Clustering

Numerical Results

Plot

K-means Clustering Analysis Plot

Download Plot



Numerical Results

Plot

Sample's cluster Samples per cluster Centers Cluster Size

Download CSV Download HTML Report

Save CSV Save HTML Report

(Downloaded and/or saved files include results from all four tabs)

Show 10 entries

Search:

Sample	Cluster
PIF_178	2
PIF_087	2
PIF_090	2
NETL_005_V1	2
PIF_115	1
PIF_110	2
NETL_019_V1	1
NETCR_014_V1	1
NETCR_014_V2	1
PIF_154	1

Showing 1 to 10 of 77 entries

Previous 1 2 3 4 5 ... 8 Next

(Shows to which cluster were samples assigned)

 Types of results available for this type of analysis



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Machine Learning

Train Models

Train Models Predict New Samples

TRAIN MODELS OPTIONS

Give a name to the analysis:
trained_models

Choose the models to train:
Partial Least Squares (pls)

Column in the metadata where the class to predict is:
Seasons

Parameter Optimization:
 Choose the number of different values that will be generated and tested for each parameter of the selected models
 Choose the specific values to test in each parameter of the selected models

Number of different values to test in each model parameter
10

Model validation:
Choose one validation method:
 Resampling Cross-Validation Repeated Cross-validation Leave One Out Cross-Validation
 Leave Group Out Cross-Validation

Number of Repeats
10

Metric to test the models performance
 Accuracy
 ROC

[Go back to the Analysis Boxes](#) [Train Models](#)

 Analysis options



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:   *metabolites*

WebSpecmine: Data Analysis

Machine Learning

Train Models

Train Models Predict New Samples

TRAIN MODELS OPTIONS

Give a name to the analysis:
trained_models

Choose the models to train:
Partial Least Squares (pls)

Parameter Optimization:
 Choose the number of different values that will be generated and tested for each parameter of the selected models
 Choose the specific values to test in each parameter of the selected models

Number of different values to test in each model parameter
10

Column in
Season

Model
Choose
 R
 R

Num
10

Metri
 A
 R

Available models:

- PLS
- Decision Tree
- Rule-Based Classifier
- SVMs with Linear Kernel
- Random Forests
- Linear Discriminant Analysis
- Neural Networks

[Go back to the Analysis Boxes](#) [Train Models](#)

 Analysis options

WebSpecmine: Data Analysis

Machine Learning

Train Models

Partial Least Squares

Best Model Results Full Results Variables' Importance 3D PCA Plot Loadings Plot Loadings Table

Download Plot Image

Save Plot Image

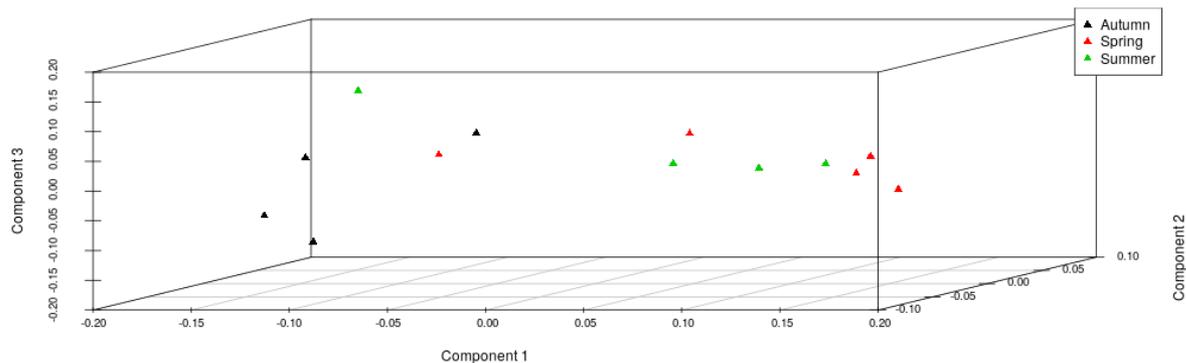
Choose the 3 Principal Components (PCs) to show in the plot:

1 2 3

Where to place the legend on the plot:

Top Right

Plot



i Types of results available for this type of analysis



3rd International Electronic Conference on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Machine Learning

Predict New Samples

Train Models

Predict New Samples

PREDICT SAMPLES OPTIONS

Please note that the dataset currently being used, chosen in the tab 'Dataset being used', must be the same one used for the model training.

Give a name to the analysis:

Submit Files

Choose one of the final models obtained to do the prediction:

Partial Least Squares - trained_models

 Analysis options

[Go back to the Analysis Boxes](#)

Predict



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Machine Learning

Predict New Samples

Predicted samples report (html):

[Download](#) [Save](#)

 Types of results available for this type of analysis

Search:

Samples' Names	Predicted Class
Apr2011	Summer
Aug2011	Autumn
Feb2011	Summer
July2010	Summer
July2011	Autumn
June2011	Autumn
Mar2011	Summer
may2011	Summer
Nov2010	Spring
Oct2010	Spring

Showing 1 to 13 of 13 entries

[Download HTML](#) [Download CSV](#) [Download EXCEL](#)

[Save HTML](#) [Save CSV](#) [Save EXCEL](#)



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Feature Selection

Give a name to the analysis:

feature_selection

Choose the method for feature selection:

- Recursive Feature Elimination (RFE)
 Selection by Filter

Column in the metadata where the class to predict is:

- Cultivar
 Transgene

Choose the Function for model fitting, prediction and variable importance/filtering:

- Random Forests Linear Regression Bagged Trees
 Linear Discriminant Analysis Naive-Bayes

Do Feature Selection

For Model validation:

Choose one validation method:

- Resampling Cross-Validation Repeated Cross-validation Leave One Out Cross-Validation Leave Group Out Cross-Validation

Number of Resampling Iterations:

10

Indicate the number of features for each group of test. If you do not want to indicate this, default values will be used.

[Go back to the Analysis Boxes](#)

 Analysis options



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

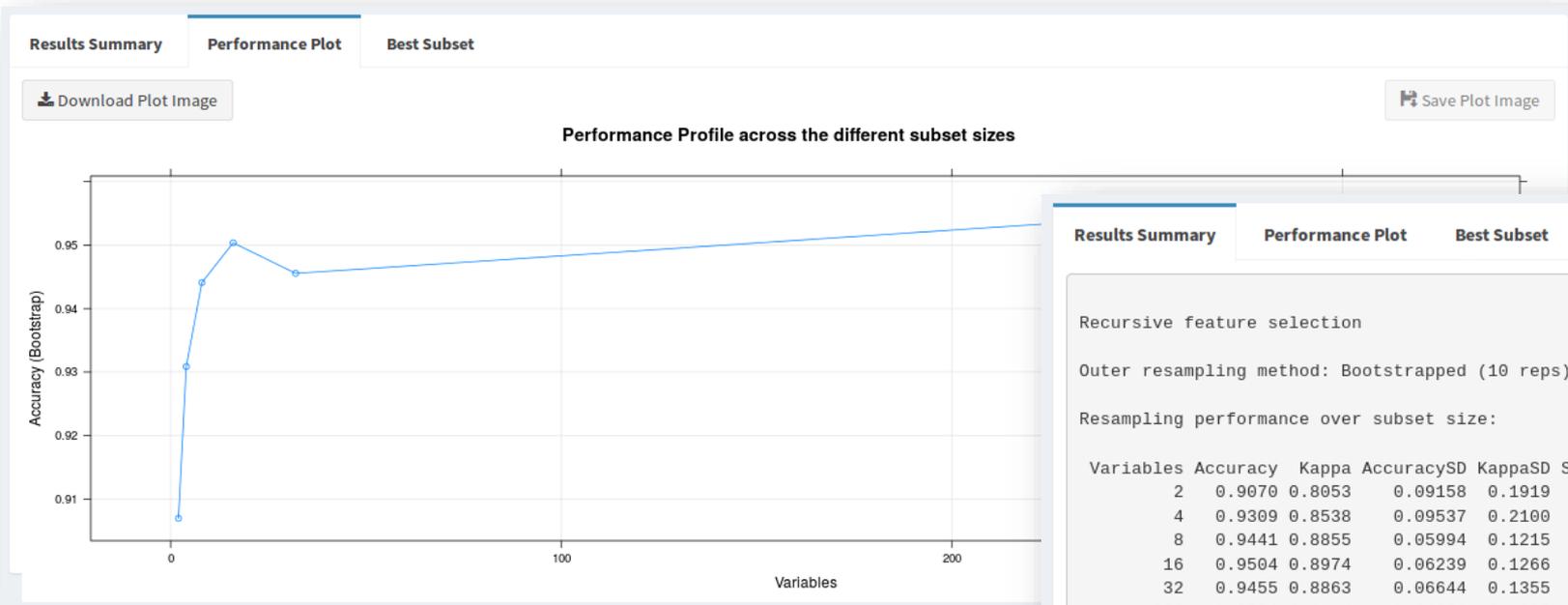
sponsors:



metabolites

WebSpecmine: Data Analysis

Feature Selection



Results Summary Performance Plot Best Subset

Recursive feature selection

Outer resampling method: Bootstrapped (10 reps)

Resampling performance over subset size:

Variables	Accuracy	Kappa	AccuracySD	KappaSD	Selected
2	0.9070	0.8053	0.09158	0.1919	
4	0.9309	0.8538	0.09537	0.2100	
8	0.9441	0.8855	0.05994	0.1215	
16	0.9504	0.8974	0.06239	0.1266	
32	0.9455	0.8863	0.06644	0.1355	
323	0.9574	0.9114	0.06359	0.1305	*

The top 5 variables (out of 323):
X3.84, X3.72, X3.62, X4.17, X3.56

Types of results available for this type of analysis



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Regression Analysis

Regression Analysis

Linear Regression Analysis

Correlation Analysis

Linear Regression Analysis

Give a name to the analysis:

Linear_Regression

Choose the metadata variables to use:

Cultivar, Transgene

Formula specifying the model:

Cultivar*Transgene

Submit

 Analysis options



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Regression Analysis

Regression Analysis

✓ Numerical Results P-value plots

P-values **Coefficients** **R-squared**

[Download CSV](#) [Download HTML Report](#) [Save CSV](#) [Save HTML Report](#)

(The Report includes results from all three tabs)

Show entries Search:

	(Intercept) ↓	CultivarThompson Seedless ↓	TransgeneGM2 ↓	TransgeneWT ↓	CultivarThompson Seedless:TransgeneWT ↓
-1.92	1	1	0.176354038104037	1	0.611342312124036
-1.88	0.9999999999999999	1	0.185740970543552	1	0.585736633894036
-1.85	0.9999999999999999	0.9999999999999999	0.193105310149427	1	0.568097851859607
-1.81	1	1	0.191282246037117	1	0.572289411099725
-1.77	0.574367532826187	0.699724577051295	0.303725168775618	0.690970332523226	0.560685526247247
-1.73	0.543073653057677	0.676143556233214	0.662140030427401	0.666790342645711	0.256880954632918
-1.71	0.710678646436599	0.79904278600031	0.335443055224965	0.793001913620993	0.41783277610987
-1.67	1	1	0.22834286452847	1	0.50367870449755
-1.64	1	1	0.159328473319151	1	0.670529160035819
-1.6	1	1	0.26910764540268	1	0.453835271472448

Showing 1 to 10 of 323 entries Previous 2 3 4 5 ... 33 Next

i Types of results available for this type of analysis: numerical results



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:

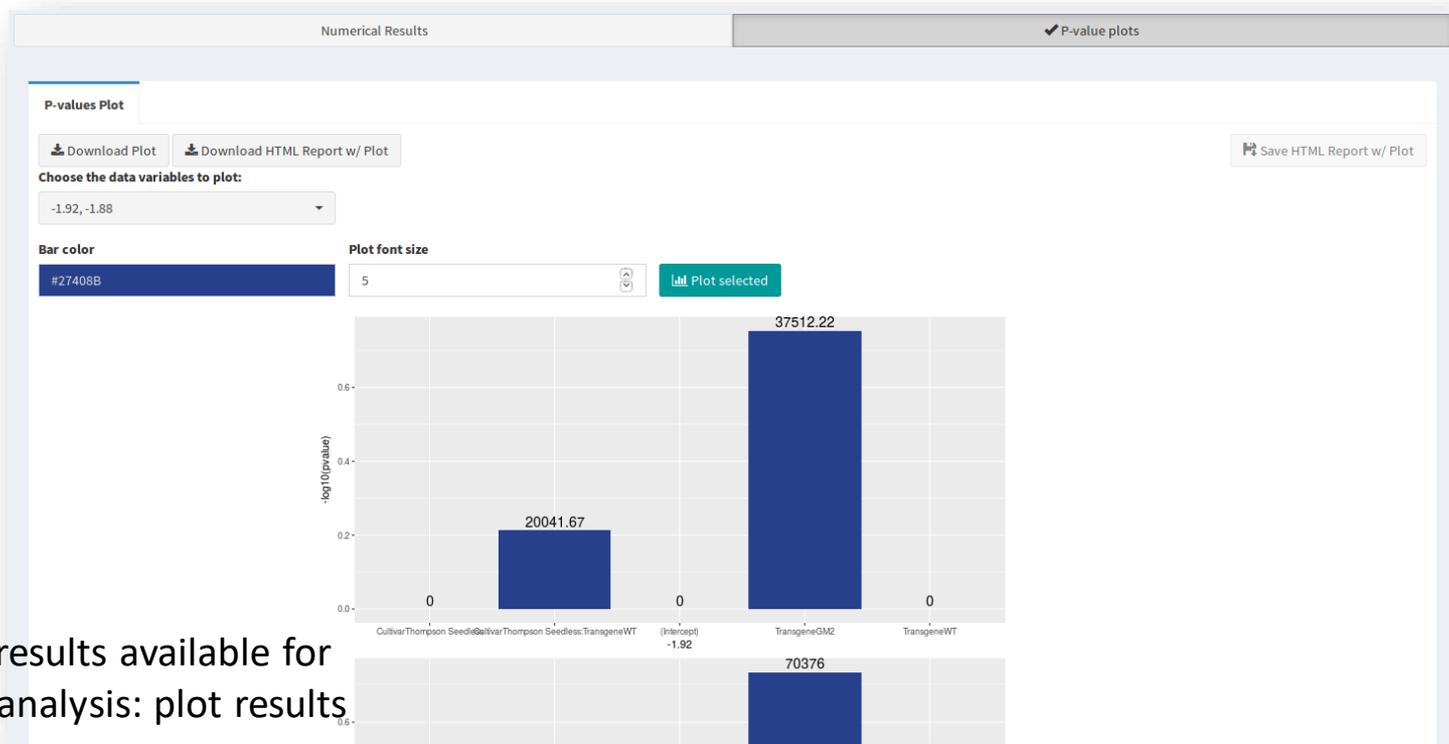


metabolites

WebSpecmine: Data Analysis

Regression Analysis

Regression Analysis



i Types of results available for this type of analysis: plot results

WebSpecmine: Data Analysis

Regression Analysis

Correlation Analysis

Linear Regression Analysis

Correlation Analysis

Correlation Analysis

Give a name to the analysis:

Correlation method:

Pearson Kendall Spearman

Calculate correlation between:

Samples Variables

Color palette used for heatmap:

 Use reversed colors of palette?

Perform correlations test to the whole dataset?

Please note the larger the dataset the more time it takes to perform the analysis.

Alternative hypothesis:

Two-sided

Greater (positive association)

Less (negative association)

Submit

[Go back to the Analysis Boxes](#)

 Analysis options



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Regression Analysis

Correlation Analysis

✓ Numerical Results Heatmap

Correlation Matrix Correlation Test Analysis

Download CSV Download HTML Report Save CSV Save HTML Report

Show 10 entries Search:

	SIL-GM1_1	SIL-GM1_2	SIL-GM1_3	SIL-GM1_4	SIL-GM1_5	SIL-GM1_6	SIL-GM1_7	SIL-GM1_8	SIL-GM1_9	SIL-GM2_1
SIL-GM1_1	1	0.873436836891152	0.864697790258547	0.67619559238544	0.947630256344774	0.921339737711138	0.874036973202246	0.862911728339762	0.899033321541915	0.542199105822603
SIL-GM1_2	0.873436836891152	1	0.978700297355265	0.522094759038555	0.922305552735914	0.971321200436765	0.981865919458206	0.839643910603088	0.947870548447443	0.332791597581709
SIL-GM1_3	0.864697790258547	0.978700297355265	1	0.565681525873933	0.887370946510306	0.938784402410529	0.964427259764198	0.795804628866346	0.942665695074166	0.372650711191787
SIL-GM1_4	0.67619559238544	0.522094759038555	0.565681525873933	1	0.623613666413636	0.596214722134496	0.579406120830404	0.586537590671043	0.524240237640015	0.858973709846568
SIL-GM1_5	0.947630256344774	0.922305552735914	0.887370946510306	0.623613666413636	1	0.965402503162488	0.916532704271786	0.893153901886449	0.931567228926557	0.455814210052779
SIL-GM1_6	0.921339737711138	0.971321200436765	0.938784402410529	0.596214722134496	0.965402503162488	1	0.981186629460332	0.892259913275145	0.927292945535872	0.386086231958885
SIL-GM1_7	0.874036973202246	0.981865919458206	0.964427259764198	0.579406120830404	0.916532704271786	0.981186629460332	1	0.843449767345282	0.908052077491045	0.366553805240431
SIL-GM1_8	0.862911728339762	0.839643910603088	0.795804628866346	0.586537590671043	0.893153901886449	0.892259913275145	0.843449767345282	1	0.848779049096775	0.405240526556645
SIL-GM1_9	0.899033321541915	0.947870548447443	0.942665695074166	0.524240237640015	0.931567228926557	0.927292945535872	0.908052077491045	0.848779049096775	1	0.36511649630234

i Types of results available for this type of analysis: numerical results



3rd International Electronic Conference on Metabolomics
15-30 November 2018



WebSpecmine: Data Analysis

Regression Analysis

Correlation Analysis



i Types of results available for this type of analysis: plot results



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Metabolite Identification

LC-MS Data

The metabolite identification is performed using the MAIT package.

Peaks are first annotated, by using the default MAIT table for adducts in positive polarization.

Next, statistically significant features are detected, followed by the identification of biontransformations between features, as well as looking for adducts.

Finally, the metabolite identification for the significant features is performed, by using the Human Metabolome Database (HMDB), version 2009/07. The peak tolerance value is set to 0.005.

ANALYSIS OPTIONS

Give a name to the analysis:

Column in the metadata that can help to identify the metabolites

type

Identify metabolites

Analysis options



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Metabolite Identification

LC-MS Data

Search:

Name	ENTRY	Query Mass	Database Mass (neutral mass)	Retention Time	Isotope	Adduct	spectra	Biofluid	p.adj
Pyridinoline	HMDB0000851	429.2	428.190704	64.78	[27][M]+		27	Blood; Urine	0.188547409625503
Biotripyrrin-a	HMDB0003323	466.2	465.189972	61.63	[34][M]+		32	Not Available	0.429014064049224
Biotripyrrin-b	HMDB0003324	466.2	465.189972	61.63	[34][M]+		32	Not Available	0.429014064049224
6-Hydroxymelatonin	HMDB0004081	249.1239	248.116089	61.13	[3][M]+		41	Blood	0.85908976745116
Imipramine	HMDB0001848	281.2	280.193939	61.13			41	Not Available	0.879721470481121
Valproic acid glucuronide	HMDB0000901	321.1537	320.147125	61.12			41	Urine	0.797418768561799
Octanoylglucuronide	HMDB0010347	321.1537	320.147125	61.12			41	Not Available	0.797418768561799
N-Acetylaspartylglutamic acid	HMDB0001067	305.1	304.090668	48.78	[10][M]+		43	Blood; Cerebrospinal Fluid; Urine	0.220571118000746
12-oxo-20-dihydroxy-leukotriene B4	HMDB0012551	366.2048	365.196411	64.12			45	Not Available	0.879195548135019
PE(P-16:0e/0:0)	HMDB0011152	438.3	437.290619	67.6	[30][M]+		49	Not Available	0.0105769947872573
N-Acetylaspartylglutamic acid	HMDB0001067	305.1	304.090668	58.4	[11][M]+		50	Blood; Cerebrospinal Fluid; Urine	0.89573574415104

Showing 1 to 49 of 49 entries

[Download HTML](#) [Download CSV](#) [Download EXCEL](#) [Save HTML](#) [Save CSV](#) [Save EXCEL](#)

 Results available for this type of analysis



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Metabolite Identification

NMR Data

Give a name to the analysis:

ppm tolerance:

Number of top metabolites matched to show in the results:

Construction of clusters parameters:

Choose the correlation method to use in the formation of clusters:
 Pearson Spearman

Minimum correlation threshold to use in the formation of clusters:
 Value given
 Calculate optimum value (leads to the maximum number of clusters)

Give the minimum correlation value

Minimum number of peaks in each cluster

Filtering of reference metabolites:

Frequency (MHz)
 400 500 600

Nucleus
 ^1H ^{13}C

Use solvent feature to filter reference metabolites
 Use pH feature to filter reference metabolites
 Use temperature feature to filter reference metabolites

 Analysis options



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Metabolite Identification

NMR Data

Metabolite	Metabolite.Name	Reference.Peaks.Matched	Cluster.Peaks.Matched	Cluster	Jaccard.Index
HMDB0001547	Corticosterone	0.931; 0.983; 0.991; 1.01; 1.06; 1.071; 1.092; 1.121; 1.147; 1.402; 1.416; 1.433; 1.462; 1.572; 1.58; 1.606; 1.79; 1.811; 1.994; 2.002; 2.025; 2.05; 2.091; 2.125; 2.178; 2.184; 2.202; 2.235; 2.263; 2.329; 2.36; 2.37; 2.382; 2.41; 2.449; 2.471; 2.502; 3.348; 3.35; 3.358; 4.182; 4.191; 4.206; 4.415; 5.68; 5.683	0.93; 0.96; 0.99; 1.02; 1.06; 1.08; 1.11; 1.15; 1.17; 1.4; 1.43; 1.46; 1.49; 1.55; 1.57; 1.62; 1.82; 1.84; 1.99; 2.01; 2.05; 2.08; 2.11; 2.14; 2.17; 2.2; 2.23; 2.26; 2.29; 2.32; 2.34; 2.38; 2.41; 2.44; 2.47; 2.5; 2.53; 3.32; 3.35; 3.38; 4.17; 4.21; 4.23; 4.4; 5.67; 5.7	1	0.123
HMDB0000016	Deoxycorticosterone	0.695; 0.976; 0.984; 1.003; 1.082; 1.085; 1.092; 1.175; 1.18; 1.192; 1.223; 1.34; 1.348; 1.371; 1.404; 1.431; 1.461; 1.562; 1.57; 1.59; 1.627; 1.66; 1.794; 1.856; 1.937; 1.967; 2.011; 2.023; 2.053; 2.218; 2.223; 2.246; 2.268; 2.298; 2.315; 2.357; 2.384; 2.413; 2.442; 2.479; 2.503; 3.317; 3.319; 4.189; 4.204; 5.736	0.69; 0.96; 0.99; 1.02; 1.06; 1.08; 1.11; 1.15; 1.17; 1.21; 1.24; 1.34; 1.37; 1.4; 1.43; 1.46; 1.49; 1.55; 1.57; 1.62; 1.65; 1.67; 1.82; 1.84; 1.96; 1.99; 2.01; 2.05; 2.08; 2.2; 2.23; 2.26; 2.29; 2.32; 2.34; 2.38; 2.41; 2.44; 2.47; 2.5; 2.53; 3.3; 3.32; 4.17; 4.21; 5.75	1	0.123
HMDB0001830	Progesterone	0.67; 0.982; 0.991; 1.01; 1.163; 1.168; 1.18; 1.211; 1.25; 1.437; 1.449; 1.48; 1.547; 1.556; 1.576; 1.628; 1.632; 1.645; 1.851; 2.025; 2.03; 2.051; 2.083; 2.13; 2.144; 2.175; 2.202; 2.295; 2.301; 2.312; 2.357; 2.38; 2.41; 2.445; 2.522; 2.545; 2.567; 5.735	0.66; 0.96; 0.99; 1.02; 1.15; 1.17; 1.21; 1.24; 1.27; 1.43; 1.46; 1.49; 1.52; 1.55; 1.57; 1.62; 1.65; 1.67; 1.84; 2.01; 2.05; 2.08; 2.11; 2.14; 2.17; 2.2; 2.23; 2.29; 2.32; 2.34; 2.38; 2.41; 2.44; 2.47; 2.5; 2.53; 2.56; 5.75	1	0.103
HMDB0000234	Testosterone	0.794; 0.921; 0.932; 0.962; 0.991; 1.047; 1.058; 1.08; 1.197; 1.285; 1.3; 1.316; 1.414; 1.424; 1.44; 1.461; 1.49; 1.561; 1.57; 1.599; 1.62; 1.654; 1.834; 1.846; 2.007; 2.015; 2.02; 2.053; 2.081; 2.133; 2.286; 2.293; 2.297; 2.351; 2.357; 2.383; 2.418; 3.649; 5.731	0.79; 0.93; 0.96; 0.99; 1.02; 1.06; 1.08; 1.11; 1.17; 1.27; 1.3; 1.34; 1.4; 1.43; 1.46; 1.49; 1.52; 1.55; 1.57; 1.62; 1.65; 1.67; 1.82; 1.84; 1.99; 2.01; 2.05; 2.08; 2.11; 2.14; 2.26; 2.29; 2.32; 2.34; 2.38; 2.41; 2.44; 3.62; 5.75	1	0.103
HMDB0000054	Bilirubin	1.945; 1.961; 1.982; 2.022; 2.073; 2.164; 2.436; 2.455; 2.475; 2.515; 2.52; 2.532; 4.006; 5.285; 5.292; 5.314; 5.597; 5.601; 5.619; 5.645; 6.093; 6.164; 6.208; 6.533; 6.562; 6.577; 6.606; 6.754; 6.784; 6.797; 6.827	1.96; 1.99; 2.01; 2.05; 2.08; 2.14; 2.41; 2.44; 2.47; 2.5; 2.53; 2.56; 4.02; 5.27; 5.31; 5.34; 5.57; 5.61; 5.64; 5.67; 6.11; 6.17; 6.21; 6.51; 6.58; 6.6; 6.62; 6.75; 6.78; 6.81; 6.85	1	0.089
HMDB0000488	4E,15Z-Bilirubin IXa	1.945; 1.961; 1.982; 2.022; 2.073; 2.164; 2.436; 2.455; 2.475; 2.515; 2.52; 2.532; 4.006; 5.285; 5.292; 5.314; 5.597; 5.601; 5.619; 5.645; 6.093; 6.164; 6.208; 6.533; 6.562; 6.577; 6.606; 6.754; 6.784; 6.797; 6.827	1.96; 1.99; 2.01; 2.05; 2.08; 2.14; 2.41; 2.44; 2.47; 2.5; 2.53; 2.56; 4.02; 5.27; 5.31; 5.34; 5.57; 5.61; 5.64; 5.67; 6.11; 6.17; 6.21; 6.51; 6.58; 6.6; 6.62; 6.75; 6.78; 6.81; 6.85	1	0.089

Showing 1 to 10 of 10 entries

[Download HTML](#) [Download CSV](#) [Download EXCEL](#) [Save HTML](#) [Save CSV](#) [Save EXCEL](#)

 Results available for this type of analysis



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Pathway Analysis

1. Choose the group of organisms where the organism wanted is:

Mammals Birds Reptiles Amphibians Fishes

Insects Nematodes Mollusks Cnidarians

Eudicots Monocots Green Algae Red Algae

Fungi Protists

Bacteria Archaea

2. Choose the organism:

Choose organism, whose pathways will be used:

Click in a group of organisms in the box at the left

3. Further options and Submit:

Analysis Name:

pathway_analysis

Choose the metabolite Analysis

metabolite_identification_nmr

Submit

 Analysis options



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Data Analysis

Pathway Analysis

Pathways Results Metabolites Analyzed Information Table Results available to download

Search:

The 'Ratio' column consists on the ratio between the number of compounds given in the analysis that are in the pathway and the number of total compounds in the pathway

Compounds' Names	KEGG Codes	Ratio
Glycine, serine and threonine metabolism	cpd:C00719; cpd:C00581; cpd:C01026; cpd:C00037; cpd:C00188; cpd:C00740; cpd:C00300; cpd:C00022	0.235
Phenylalanine, tyrosine and tryptophan biosynthesis	L-Tyrosine cpd:C00082	0.25
Citrate cycle (TCA cycle)	Oxoglutaric acid; Succinic acid; cis-Aconitic acid; Citric acid; Pyruvic acid; Fumaric acid cpd:C00026; cpd:C00042; cpd:C00417; cpd:C00158; cpd:C00022; cpd:C00122	0.3
D-Glutamine and D-glutamate metabolism	D-Glutamine; Oxoglutaric acid cpd:C00819; cpd:C00026	0.333
Valine, leucine and isoleucine biosynthesis	L-Threonine; L-Isoleucine; L-Valine cpd:C00188; cpd:C00407; cpd:C00183	0.375
Neomycin, kanamycin and gentamicin biosynthesis	D-Glucose cpd:C00031	0.5

Citrate cycle (TCA cycle)

i Results available for this type of analysis



3rd International Electronic Conference on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Analysis of MetaboLights Studies

..|| METABOLIGHTS PROJECTS



List of Metabolights Studies Available

Click in a row to see more detailed information on the study selected.

Search:

	Study Title	Data type(s)
MTBLS10	Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in <i>Nicotiana attenuata</i> leaves during simulated herbivory	lcms-spectra
MTBLS100	The Human Saliva Metabolome	nmr-spectra
MTBLS102	Comparative analysis of the adaptation of <i>Staphylococcus aureus</i> to internalization by different types of human non-professional phagocytic host cells (NMR assay)	nmr-spectra
MTBLS103	Metabolomics reveals impaired maturation of HDL particles in adolescents with hyperinsulinaemic androgen excess	lcms-spectra
MTBLS104	NMR based metabolomics study of Y2 receptor activation by neuropeptide Y in the SK-N-BE2 human neuroblastoma cell line	nmr-spectra
MTBLS116	Application of 1H-NMR Metabolomic Profiling for Reef-Building Corals.	nmr-spectra
MTBLS118	A Metabolic Profiling Strategy for the Dissection of Plant Defense against Fungal Pathogens (GC-MS assay)	gcms-spectra

 You can see information on some of the MetaboLights studies



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Analysis of MetaboLights Studies

MTBLS10

Development and validation of a liquid chromatography-electrospray ionization-time-of-flight mass spectrometry method for induced changes in *Nicotiana attenuata* leaves during simulated herbivory

Data Type

lcms-spectra

Organism

Nicotiana attenuata

Contacts

Emmanuel Gaquerel:
egaquerel@ice.mpg.de

Steffen Newmann:
sneumann@ipb-halle.de

Kuhl Carsten:
kuhl@ipb-halle.de

Detailed Information

Assays Information

Metabolights Study Link

Detailed Information

Description

Metabolite profiling of herbivory-induced changes in *Nicotiana attenuata* was used to construct a correlation-weighted network to visualize intra- and inter-compound spectra relationships from individual mass features, and to annotate 12 major network hubs, including 4 O-acyl sugars and 6 17-hydroxygeranylinalool diterpene glycosides, metabolites with recently uncovered anti-herbivore functions.

Sample Collection

An isogenic line of *Nicotiana attenuata* obtained after 30 generations of inbreeding from field-collected seeds was used. All plants were grown in the glasshouse in 1 L individual pots at 26 - 28 °C under 16 h of light supplied by Philips Sun-T Agro 400- or 600-W sodium lights (Philips, Turnhout, Belgium). Metabolic changes induced during *Manduca sexta* feeding were reproduced by producing with a fabric pattern wheel three rows of puncture onto each side of the midvein of five fully expanded leaves per plant (5 biological replicates) and directly applying 1:1 diluted *M. sexta* oral secretions (OS). Treated leaves from the same plant were harvested and flash frozen 0, 1, 2, 4, 14 and 24 h after elicitation.

Sample Extraction

100 mg of ground leaf tissue was weighed and transferred to a Fast Prep tube containing 0.9 g of Fast Prep matrix (BIO 101, Vista, USA). 1 ml extraction buffer per 100 mg tissue [50 mM acetate buffer, pH 4.8, containing 40 % methanol spiked with reserpine (600 ng/ml), atropine (200 ng/ml)] was added and samples were homogenized. After centrifugation (13200 rpm, 20 min, 4 °C) the supernatant was collected in a fresh 1.5 ml Eppendorf tube, centrifuged again and 100 µl of the supernatant was transferred to a HPLC vial.

MS Chromatography

Two microliters of the leaf extract were separated using a HPLC 1100 Series system (Agilent, Palo Alto, USA). The column used was a 150 mm x 2 mm i.d., 3 µm Phenomenex Gemini NX RP-18 column with a

 You can see detailed information on the protocol and metadata information on each assay



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

WebSpecmine: Analysis of MetaboLights Studies

Import Assay 1 from study MTBLS100

Create New Project for assay
 Save assay in existing project as a new data folder

Project Name:

Name of the folder to create:

The metadata file will be identified by the assay number it corresponds to.

 You can download the data and metadata of an assay into your private account and analyse it

Conclusions

✓ We were able to create an easy-to-use and freely available website with many advantages:

- Wide variety of techniques and data formats supported
- Wide variety of pre-processing methods
- Wide variety of analysis methods
- Flexible Pipeline
- User Account

✓ However, more analyses could be added, to add more biological meaning to data, such as:

- Enrichment Analysis
- Biomarker Analysis



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

For More Detailed Information ...

Website Link: <https://webspecmine.bio.di.uminho.pt/>



We have Tutorials and a complete User Guide at the *?Help* page.

We have a troubleshooting window, from where the users can report any problems and see the problems already encountered, but still being solved.



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites

Acknowledgments



Universidade do Minho
Escola de Engenharia



Christopher Costa,
for being the main
developer of the
specmine R package



3rd International Electronic Conference
on Metabolomics
15-30 November 2018

sponsors:



metabolites