Workflow4Metabo

Galaxy and the metabolomics analysis Universe

GCC2013 - Olso 2013 - Gildas Le Corguillé
JOBIM - Toulouse 2013 - Pierre Pericard
### What was our choice until now?

<table>
<thead>
<tr>
<th>ERGONOMICS</th>
<th>PARAMETER COMPLETENESS</th>
<th>MODULARITY</th>
<th>DATA &amp; WORKFLOW SHARING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>USERS</strong></td>
<td>Newbies metabo analysts</td>
<td>Advanced metabo analysts</td>
<td>Additional tools developers</td>
</tr>
</tbody>
</table>

---

```R
> library(xcms)
> setwd("/proj/2424/informatique/ppericard/Galaxy/tmp/mzXML_copper_stress.mz/"
> xset <- xcmsSet()
> xset
An "xcmsSet" object with 16 samples
Time range: 6.8-1528.4 seconds (0.1-25.5 min)
Mass range: 53.0167-699.356 m/z
Peaks: 8941 (about 559 per sample)
Peak Groups: 0
Sample classes: ref, sample
Profile settings: method = bin, step = 0.1
Memory usage: 1.56 MB
> xset <- group(xset)
115 178 240 303 365 428 490 553 615 678
> |
```
What was our choice until now?

- ERGONOMICS
- PARAMETER COMPLETENESS
- MODULARITY
- DATA & WORKFLOW SHARING

USERS
- Newbies metabo analysts
- Advanced metabo analysts
- Additional tools developers
Our Solution

**ERGONOMICS**

**PARAMETER COMPLETENESS**

**MODULARITY**

**DATA & WORKFLOW SHARING**

**USERS**
- Newbies metabo analysts
- Advanced metabo analysts
- Additional tools developers
ERGONOMICS

**ERGONOMICS**

**ERSERS Newbies metabo analysts**

---

ERSERS Newbies metabo analysts

---

**ERGONOMICS**

**ERSERS Newbies metabo analysts**
xcms.xcsmSet (version 20130418)

ms zip file:
1: mzXMLp.ms.zip

MPI-slaves CPU:
9
number of MPI-slaves to use for parallel peak detection

Method:
matchedFilter
Chose the method used for finding peaks

step:
0.01
the peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size defined by the step argument

fwhm:
30
full width at half maximum

Advanced options:
hide

Execute

** The ms_zip input file **

The input file to the metabolomic workflow is a zip file containing all your conditions as sub-directories, with its extension renamed to ".ms.zip". This file can then be loaded in Galaxy using the "Get Data" tool. The file datatype should be "mz_zip".

**What it does?** The class of objects used for preprocessing analyte data from multiple LC/MS files is xcsmSet. It stores peak lists and provides methods for grouping and aligning those peaks.

**Parameters**
The default arguments for xcmsSet should work acceptably in most cases. However, there are a number of parameters that may need to be optimized for a particular
**The ms_zip input file**

The input file to the metabolomic workflow is a zip file containing all your conditions as sub-directories, with its extension renamed to ".ms.zip". This file can then be loaded in Galaxy using the "Get Data" tool. The file datatype should be "ms_zip".

**What it does?** The class of objects used for preprocessing analyte data from multiple LC/MS files is xcmsSet. It stores peak lists and provides methods for grouping and aligning those peaks.

**Parameters**

The default arguments for xcmsSet should work acceptably in most cases. However, there are a number of parameters that may need to be optimized for a particular...
**xcms.xcmsSet (version 20130418)**

**ms zip file:**
- 1: mzXMLp.ms.zip

**MPI-slaves CPU:**
- 9
  number of MPI-slaves to use for parallel peak detection

**Method:**
- matchedFilter
  Chose the method used for finding peaks

**step:**
- 0.01
  the peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size defined by the step argument

**fwhm:**
- 30
  full width at half maximum

**Advanced options:**
- hide

---

**Execute**

---

**The ms_zip input file**

The input file to the metabolomic workflow is a zip file containing all your conditions as sub-directories, with its extension renamed to "ms.zip". This file can then be loaded in Galaxy using the "Get Data" tool. The file datatype should be "mz_zip".

**What it does?** The class of objects used for preprocessing analyte data from multiple LC/MS files is xcmsSet. It stores peak lists and provides methods for grouping and aligning those peaks.

**Parameters**

The default arguments for xcmsSet should work acceptably in most cases. However, there are a number of parameters that may need to be optimized for a particular
PARAMETER COMPLETENESS

**The ms_zip input file**

The input file to the metabolomic workflow is a zip file containing all your conditions as sub-directories, with Galaxy using the "Genome Analysis Tools (GATK)" toolset. What it does? The class contains LC/MS files is xcmsSet. It stores peak lists and provides methods for grouping and aligning those peaks.

Parameters

The default arguments for xcmsSet should work acceptably in most cases. However, there are a number of parameters that may need to be optimized for a particular dataset.

**Method**

matchedFilter

Chose the method used for finding peaks

step:

0.01

the peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size defined by the step argument

fwhm:

30

full width at half max

Advanced options:

hide

**PARAMETER COMPLETENESS**

**USERS** Advanced metabo analysts
xcms.xcmsSet (version 20130418)

**ms.zip file:**
1: mZXMLp.ms.zip

**MPI-slaves CPU:**
9
number of MPI-slaves to use for parallel peak detection

**Method:**
matchedFilter
Chose the method used for finding peaks

**step:**
0.01
the peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size defined by the step argument

**fwhm:**
30
full width at half maximum

**Advanced options:**
hide

**Execute**

**The ms.zip input file**
The input file to the metabolomic workflow is a zip file containing all your conditions as sub-directories, with its extension renamed to "ms.zip". This file can then be loaded in Galaxy using the "Get Data" tool. The file datatype should be "mZ_xml".

What it does? The class of objects used for preprocessing analyte data from multiple LC/MS files is xcmsSet. It stores peak lists and provides methods for grouping and aligning those peaks.

**Parameters**
The default arguments for xcmsSet should work acceptably in most cases. However, there are a number of parameters that may need to be optimized for a particular dataset.

---

**CAMERA annotateDiffreport**
Wrapper function for the xcms dffreport and the annotate function. Returns a dffreport within the annotation results.

**Normalization VdK LOWESS**
Normalize intensities (linear or...
xcms.xcmsSet (version 20130418)

**ms zip file:**

1: mzXMLp.ms.zip

**MPI-slaves CPU:**

9

number of MPI-slaves to use for parallel peak detection

**Method:**

matchedFilter

Chose the method used for finding peaks

**step:**

0.01

the peak detection algorithm creates extracted ion base peak chromatograms (EIBPC) on a fixed step size defined by the step argument

**fwhm:**

30

full width at half maximum

**Advanced options:**

**max:**

5

maximum number of peaks per extracted ion chromatogram

**snthresh:**

10

signal to noise ratio cutoff

**steps:**

2

the peak identification algorithm combines a given number of EIBPCs prior to filtration and peak detection, as defined by the steps argument

Execute
Our Pipeline (so far…)

MODULARITY

USERS
- Advanced metabo analysts
- Additional tools developers

MODULARITY

USERS
Advanced metabo analysts
Additional tools developers

GCC2013 – Oslo 2013
Our Pipeline (so far...)

[Diagram showing a pipeline process with steps such as .ms.zip file, WT, cond1, cond2, xcmsSet, retcor, group, fillPeaks, diffreport, and OR connections.]
Our Pipeline (so far...)

[Diagram of a pipeline process]

14 02/07/2013 GCC2013 – Oslo 2013
Our Pipeline (scalability)
Our Pipeline (scalability)

+ homemade patch on upload.py
ERGONOMICS

MODULARITY

USERS Advanced metabo analysts
Running workflow "Workflow constructed from history 'XCMS screencast'

Step 1: Input dataset

Input Dataset

1: mzXMLp.ms.zip

Step 2: xcms.xcmsSet (version 20130418)

Step 3: xcms.group (version 20130418)

RData file
Output dataset 'output' from step 2

Method

density

bw 30

minfrac 1.0

mzwid 0.25

Advanced options
show

max 5

Step 4: xcms.recor (version 20130418)

Step 5: xcms.group (version 20130418)

Step 6: xcms.fillPeaks (version 20130418)
Running workflow "Workflow constructed from history 'XCMS screencast'"

**Step 1: Input dataset**

- **Input Dataset**: 
  - 1: mzXMLp.ms.zip

**Step 2: xcms.xcmsSet (version 20130418)**

**Step 3: xcms.group (version 20130418)**

- **RData file**
  - Output dataset 'output' from step 2

  - **Method**
    - density
  - bw: 30
  - minfrac: 1.0
  - mzwid: 0.25

  **Advanced options**
  - show

**Step 4: xcms.retcor (version 20130418)**

**Step 5: xcms.group (version 20130418)**

**Step 6: xcms.fillPeaks (version 20130418)**

**History**

- 60: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.data_matrix.tsv_anova_filtered.tab.acp.zip
- 59: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.data_matrix.tsv_anova_filtered.tab.hclust.zip
- 58: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.data_matrix.tsv_anova_filtered.tab.tabular
- 57: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.data_matrix.tsv_anova_pvalue.tabular
- 21: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.Rdata
- 20: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.data_matrix.tsv
- 19: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.data_matrix.tsv
- 18: mzXMLp.group.retcor.group.retcor.group.fillPeaks.annotateDiffreport.tsv
Running workflow "Workflow constructed from history 'XCMS screencast'"

Step 1: Input dataset

Input Dataset
1: mzXMLp.ms.zip

type to filter

Step 2: xcms.xcmsSet (version 20130418)

Step 3: xcms.group (version 20130418)

RData file
Output dataset 'output' from step 2

Method
density
bw
30

minfrac
1.0

mzwid
0.25

Advanced options
show

max
5

Step 4: xcms.retcor (version 20130418)

Step 5: xcms.group (version 20130418)

Step 6: xcms.fillPeaks (version 20130418)
Data & workflows & tools sharing

- **Data**: Easy sharing of Galaxy histories, even between distant instances

- **Workflow**: Advanced users can share pre-configured workflows (e.g. high or medium resolution MS)

- **Tools**: Galaxy tools and pipelines can be shared through Toolsheds
Who’s who

Pierre Pericard
Urszula Czerwinska
Sophie Goulimiquer

Gildas Le Corguillé
Christophe Caron

Estelle Pujos-Guillot
Franck Giacomoni
Marion Landi

Christophe Duperier
Jean-François Martin
Takk

abims.sb-roscoff.fr
galaxy.sb-roscoff.fr