XCMS Workshop
Alignments with OBI-Warp
Annotation using CAMERA

Ralf Tautenhahn

The Scripps Research Institute

5/31/2009

OBI-Warp

John T. Prince and Edward M. Marcotte
**Chromatographic Alignment of ESI-LC-MS Proteomics Data Sets by
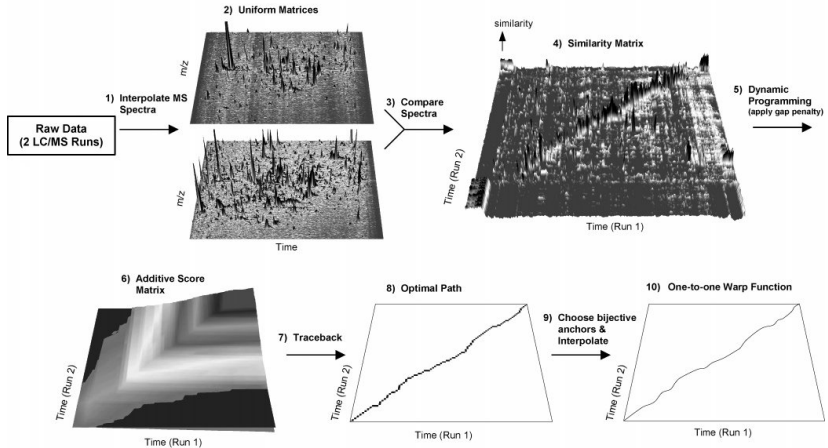Ordered Bijective Interpolated Warping**
Analytical Chemistry, **2060** 78 (17), 6140-6152

`obi-warp.sourceforge.net`

- Retention time correction based on spectra similarity
- Doesn't rely on detected feature
- No initial grouping needed
- Potentially useful for samples with high retention time deviations

- Recently integrated into XCMS
- CAVE: still experimental!

OBI-Warp

I) Alignment by OBI-Warp

Retention time correction using OBI-Warp

```
library(xcms)
cdfpath <- system.file("cdf", package = "faahKO")
cdffiles <- list.files(cdfpath, recursive = TRUE, full.names = TRUE)
xset <- xcmsSet(cdffiles)
```

OBI-Warp   LOESS

```
xrc  <- retcor(xset,          xg <- group(xset)
     method="obiwarp")        xrc <- retcor(xg)
```

Additional parameters see `?retcor.obiwarp`

Retention time correction using OBI-Warp

```
library(xcms)
cdfpath <- system.file("cdf", package = "faahKO")
cdffiles <- list.files(cdfpath, recursive = TRUE, full.names = TRUE)
xset <- xcmsSet(cdffiles)
```

OBI-Warp   LOESS

```
xrc  <- retcor(xset,            xg <- group(xset)
     method="obiwarp")          xrc <- retcor(xg)
```

final grouping

```
x <- group(xrc)
```

Additional parameters see `?retcor.obiwarp`

Retention time correction using OBI-Warp

```
library(xcms)
cdfpath <- system.file("cdf", package = "faahKO")
cdffiles <- list.files(cdfpath, recursive = TRUE, full.names = TRUE)
xset <- xcmsSet(cdffiles)
```

OBI-Warp   LOESS

```
xrc  <- retcor(xset,              xg <- group(xset)
     method="obiwarp")            xrc <- retcor(xg)
```

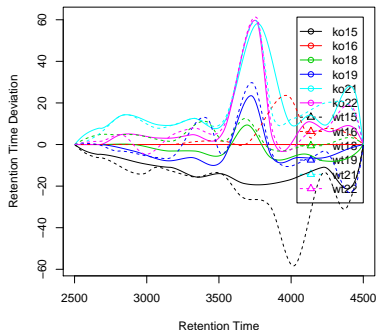final grouping

```
x <- group(xrc)
```

OR

```
x <- group(xrc, method="nearest")
```

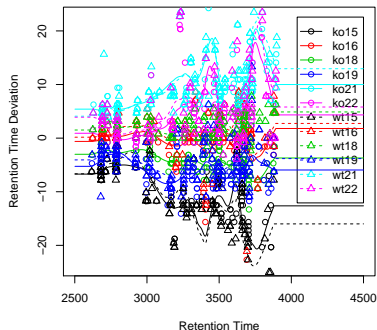Additional parameters see `?retcor.obiwarp`

4

Example: faahKO dataset

OBI-Warp   LOESS

Alignment using the "nearest" algorithm

- alternative to group.density (default)
- inspired by the alignment algorithm of mzMine
- main difference: no binning necessary
- usage

      xsg <- group(xs,method="nearest", mzCheck=0.2,
            rtCheck=15, kNN=10)

- *mzCheck* maximum tolerated distance for *m/z*
- *rtCheck* maximum tolerated distance for RT.
- *kNN* number of nearest neighbors to check

?group.nearest
To use this method, the R-package *RANN* has to be installed.
http://rforge.net/RANN/

Annotation of LC/MS Features

m/z	Retention time	Intensity
135.05	280.43	97554
153.06	280.43	4207
175.04	280.43	7468
197.02	280.76	1015
117.04	278.07	995
145.03	278.07	20534
149.06	278.07	3561
177.05	278.07	31096
195.06	278.07	1925
217.05	277.74	1704
233.01	278.07	3541
427.07	278.07	1897
621.13	278.07	435

.
.
.

Annotation of LC/MS Features

m/z	Retention time	Intensity	Compound No.
135.05	280.43	97554	1
153.06	280.43	4207	1
175.04	280.43	7468	1
197.02	280.76	1015	1
117.04	278.07	995	2
145.03	278.07	20534	2
149.06	278.07	3561	2
177.05	278.07	31096	2
195.06	278.07	1925	2
217.05	277.74	1704	2
233.01	278.07	3541	2
427.07	278.07	1897	2
621.13	278.07	435	2

.
.
.

- Grouping of all features belonging to the same compound

Annotation of LC/MS Features

m/z	Retention time	Intensity	Ion	Compound No.
135.05	280.43	97554		1
153.06	280.43	4207	$[M+H]^+$	1
175.04	280.43	7468	$[M+Na]^+$	1
197.02	280.76	1015		1
117.04	278.07	995		2
145.03	278.07	20534		2
149.06	278.07	3561		2
177.05	278.07	31096		2
195.06	278.07	1925	$[M+H]^+$	2
217.05	277.74	1704	$[M+Na]^+$	2
233.01	278.07	3541	$[M+K]^+$	2
427.07	278.07	1897	$[2M+K]^+$	2
621.13	278.07	435	$[3M+K]^+$	2

$\vdots$

- Grouping of all features belonging to the same compound

- Detect charge state

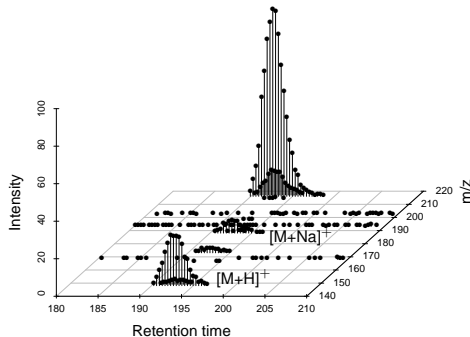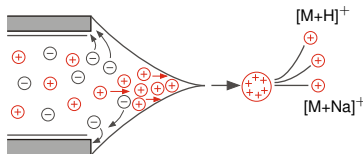- Assignment of ion species

Annotation of LC/MS Features

m/z	Retention time	Intensity	Ion	Molecular mass	Compound No.
135.05	280.43	97554			1
153.06	280.43	4207	[M+H]$^+$		1
175.04	280.43	7468	[M+Na]$^+$	$\left.\begin{array}{c}\\\\\end{array}\right\} M = 152.05$	1
197.02	280.76	1015			1
117.04	278.07	995			2
145.03	278.07	20534			2
149.06	278.07	3561			2
177.05	278.07	31096			2
195.06	278.07	1925	[M+H]$^+$		2
217.05	277.74	1704	[M+Na]$^+$		2
233.01	278.07	3541	[M+K]$^+$	$\left.\begin{array}{c}\\\\\\\\\end{array}\right\} M = 194.05$	2
427.07	278.07	1897	[2M+K]$^+$		2
621.13	278.07	435	[3M+K]$^+$		2

$\vdots$

- Grouping of all features belonging to the same compound

- Detect charge state

- Assignment of ion species

- Calculate molecular mass

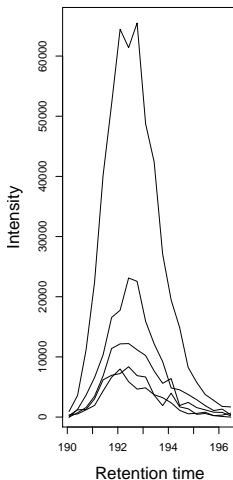$\rightarrow$ Database search, molecular formula calculation

Grouping all features of the same compound (1)

- Electrospray ionization
  - Molecule M: Formation of different ion species e.g. $[M+H]^+$, $[M+Na]^+$
  - in the same ratio
- Features belonging to the same compound
  - Same retention time
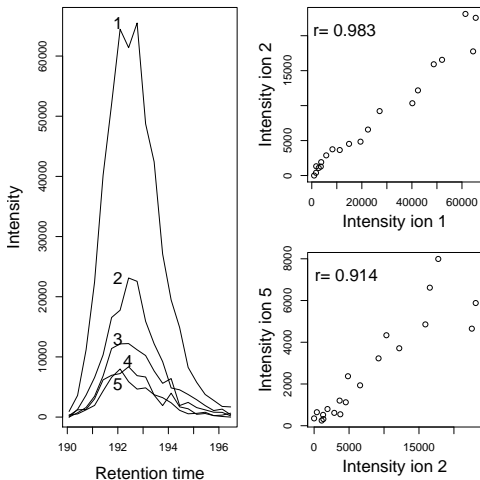 - Feature intensities have a linear relationship

Grouping all features of the same compound (1)

- Electrospray ionization
  - Molecule M: Formation of different ion species e.g. $[M+H]^+$, $[M+Na]^+$
  - in the same ratio
- Features belonging to the same compound
  - Same retention time
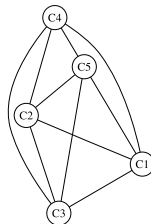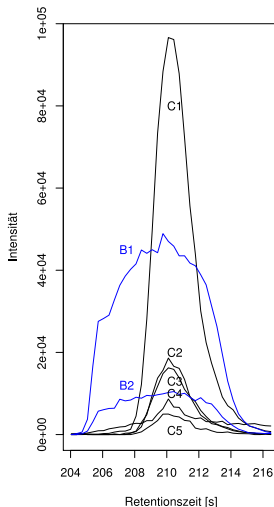 - Feature intensities have a linear relationship

Grouping all features of the same compound (1)

- Electrospray ionization
  - Molecule M: Formation of different ion species e.g. $[M+H]^+$, $[M+Na]^+$
  - in the same ratio
- Features belonging to the same compound
  - Same retention time
  - Feature intensities have a linear relationship

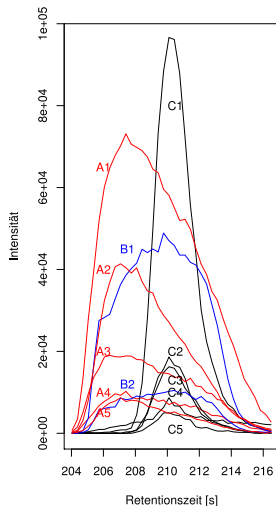# Grouping all features of the same compound (2)

- Compute correlation coefficients
- Create graph
  - Node: Feature
  - Edge: $r > r_{Thr}$

Grouping all features of the same compound (2)

- Compute correlation coefficients
- Create graph
  - Node: Feature
  - Edge: $r > r_{Thr}$

Grouping all features of the same compound (2)

- Compute correlation coefficients
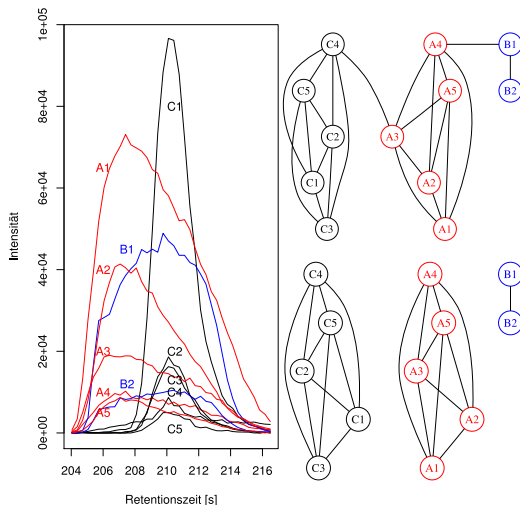- Create graph
  - Node: Feature
  - Edge: $r > r_{Thr}$
- *HCS*-Clustering
  [Hartuv and Shamir 2000]
- highly connected subgraphs

Assignment of ion species (1)
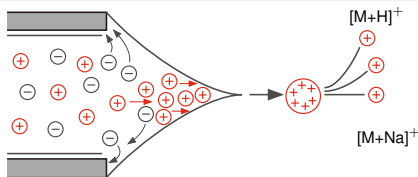
- Ion formation depends on the compound

Assignment of ion species (1)

- Ion formation depends on the compound
  - quasi-molecular ion, e.g. $[M+H]^+$
  - cluster ions, e.g. $[2M+Na]^+$
  - fragment ions, e.g. $[M+H-R]^+$
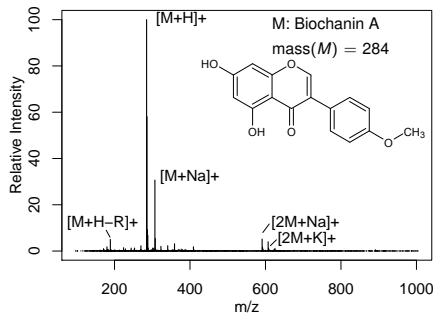
Assignment of ion species (1)

- Ion formation depends on the compound

    - quasi-molecular ion, e.g. $[M+H]^+$
    - cluster ions, e.g. $[2M+Na]^+$
    - fragment ions, e.g. $[M+H-R]^+$

Observed *m/z* value *s* for a single ion species:

$$s = \frac{n \cdot \text{mass}(M) + \text{mass}(I) - \text{mass}(R)}{z}$$

Assignment of ion species (1)

- Ion formation depends on the compound

  - quasi-molecular ion, e.g. $[M+H]^+$
  - cluster ions, e.g. $[2M+Na]^+$
  - fragment ions, e.g. $[M+H-R]^+$

Observed m/z value s for a single ion species:

$$s = \frac{n \cdot \mathrm{mass}(M) + \mathrm{mass}(I) - \mathrm{mass}(R)}{z}$$

Example

$[M+Na]^+$, mass(M)=284, n=1,z=1, mass(I)=23

$$\begin{aligned} s &= \frac{1 \cdot 284 + 23}{1} \\ &= 307 \ m/z \end{aligned}$$
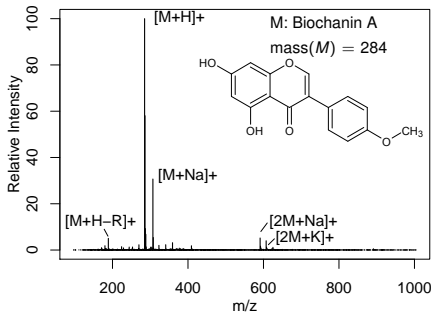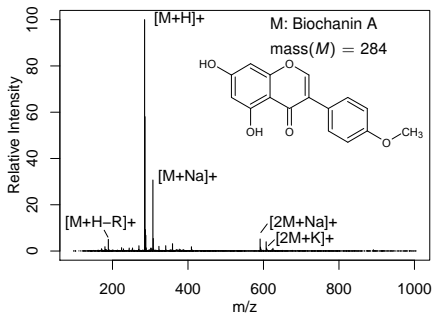
10

Assignment of ion species (1)

- Ion formation depends on the compound
    - quasi-molecular ion, e.g. $[M+H]^+$
    - cluster ions, e.g. $[2M+Na]^+$
    - fragment ions, e.g. $[M+H-R]^+$
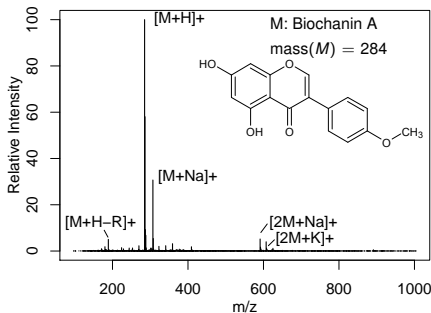
Observed *m/z* value *s* for a single ion species:

$$s = \frac{n \cdot \text{mass}(M) + \text{mass}(I) - \text{mass}(R)}{z}$$

$$\text{mass}(M) = \frac{z \cdot s - \text{mass}(I) + \text{mass}(R)}{n}$$

$$n, z > 0$$

Example

$[M+Na]^+$, mass(M)=284, n=1,z=1, mass(I)=23

$$s = \frac{1 \cdot 284 + 23}{1}$$
$$= 307 \ m/z$$

10

Assignment of ion species (2)

rule table (z, n, I, R)	
rule j	mass(I_j)
$[M+H]^+$	1.01
$[M+2H]^{2+}$	2.01
$[M+Na]^+$	22.99
$[M+K]^+$	38.96
$[2M+Na]^+$	22.99
$[2M+K]^+$	38.96
. . .	

feature group
s_i [m/z]
270.05
285.07
307.06
591.12
607.09
. . .

[Tautenhahn et al. 2007, Springer LNCS]

Assignment of ion species (2)

- Creation of mass-hypotheses

$$M_{i,j} = \frac{z_j \cdot s_i - l_j + R_j}{n_j}$$

rule table (z, n, l, R)	
rule j	mass(l_j)
$[M+H]^+$	1.01
$[M+2H]^{2+}$	2.01
$[M+Na]^+$	22.99
$[M+K]^+$	38.96
$[2M+Na]^+$	22.99
$[2M+K]^+$	38.96
. . .	

feature group
s_i [m/z]
270.05
285.07
307.06
591.12
607.09
. . .

[Tautenhahn et al. 2007, Springer LNCS]

11

Assignment of ion species (2)

- Creation of mass-hypotheses

$$M_{i,j} = \frac{z_j \cdot s_i - I_j + R_j}{n_j}$$

rule table (z, n, I, R)	
rule j	mass(I_j)
$[M+H]^+$	1.01
$[M+2H]^{2+}$	2.01
$[M+Na]^+$	22.99
$[M+K]^+$	38.96
$[2M+Na]^+$	22.99
$[2M+K]^+$	38.96
. . .	

feature group
s_i [m/z]
270.05
285.07
307.06
591.12
607.09
. . .

$$M_{i,j} =$$

$$\begin{pmatrix} 269 & 538.1 & 247.1 & 231.1 & 123.5 & 115.5 \\ 284.1 & 568.1 & 262.1 & 246.1 & 131 & 123.1 \\ 306.1 & 612.1 & 284.1 & 268.1 & 142 & 134.1 \\ 590.1 & 1180.2 & 568.1 & 552.2 & 284.1 & 276.1 \\ 606.1 & 1212.2 & 584.1 & 568.1 & 292.1 & 284.1 \end{pmatrix}$$

[Tautenhahn et al. 2007, Springer LNCS]

Assignment of ion species (2)

- Creation of mass-hypotheses

$$M_{i,j} = \frac{z_j \cdot s_i - I_j + R_j}{n_j}$$

$$\text{e.g. } M_{5,6} = \frac{1 \cdot 607.09 - 38.96 + 0}{2}$$

$$= 284.1$$

rule table (z, n, I, R)	
rule j	mass(I_j)
$[M+H]^+$	1.01
$[M+2H]^{2+}$	2.01
$[M+Na]^+$	22.99
$[M+K]^+$	38.96
$[2M+Na]^+$	22.99
$[2M+K]^+$	38.96
. . .	

feature group
s_i [m/z]
270.05
285.07
307.06
591.12
607.09
. . .

$$M_{i,j} =$$

$$\begin{pmatrix} 269 & 538.1 & 247.1 & 231.1 & 123.5 & 115.5 \\ 284.1 & 568.1 & 262.1 & 246.1 & 131 & 123.1 \\ 306.1 & 612.1 & 284.1 & 268.1 & 142 & 134.1 \\ 590.1 & 1180.2 & 568.1 & 552.2 & 284.1 & 276.1 \\ 606.1 & 1212.2 & 584.1 & 568.1 & 292.1 & 284.1 \end{pmatrix}$$

[Tautenhahn et al. 2007, Springer LNCS]

Assignment of ion species (2)

- Creation of mass-hypotheses

$$M_{i,j} = \frac{z_j \cdot s_i - I_j + R_j}{n_j}$$

e.g. $M_{5,6} = \dfrac{1 \cdot 607.09 - 38.96 + 0}{2}$

$= 284.1$

rule table (z, n, I, R)	
rule j	mass(I_j)
$[M+H]^+$	1.01
$[M+2H]^{2+}$	2.01
$[M+Na]^+$	22.99
$[M+K]^+$	38.96
$[2M+Na]^+$	22.99
$[2M+K]^+$	38.96
. . .	

feature group		
s_i [m/z]	Ion	Mass
270.05		
285.07	$[M+H]^+$	
307.06	$[M+Na]^+$	M=284.1
591.12	$[2M+Na]^+$	
607.09	$[2M+K]^+$	
. . .		

$M_{i,j} =$

$$\begin{pmatrix} 269 & 538.1 & 247.1 & 231.1 & 123.5 & 115.5 \\ 284.1 & 568.1 & 262.1 & 246.1 & 131 & 123.1 \\ 306.1 & 612.1 & 284.1 & 268.1 & 142 & 134.1 \\ 590.1 & 1180.2 & 568.1 & 552.2 & 284.1 & 276.1 \\ 606.1 & 1212.2 & 584.1 & 568.1 & 292.1 & 284.1 \end{pmatrix}$$

[Tautenhahn et al. 2007, Springer LNCS]

Usage

- http://bioconductor.org/packages/devel/bioc/html/CAMERA.html
- or (within R) type:
  ```
  source("http://bioconductor.org/biocLite.R")
  biocLite("CAMERA")
  ```

- example using the faahKO dataset
  ```
  library(CAMERA)
  xs <- xcmsSet(...)
  ## group, retcor, etc.

  an <- annotate(xs)
  peaklist <- getPeaklist(an)
  write.csv(peaklist,file='xsannotated.csv')
  ```

Interpretation of the Results

id	mz	rt	isotopes	adduct	pc
65	176.04	280.09			
76	136.05	280.43	[14][M+1]1+		5
77	135.05	280.43	[14][M]1+		5
74	153.06	280.43		[M+H]+ 152.05437	5
75	175.04	280.43		[M+Na]+ 152.05437	5
73	197.02	280.76		[M+2Na-H]+ 152.05437	5
78	377.74	286.15			
79	732.5	286.49			
83	488.32	286.82		[M+Na]+ 465.33205	7
82	466.34	286.82		[M+H]+ 465.33205	7
...					