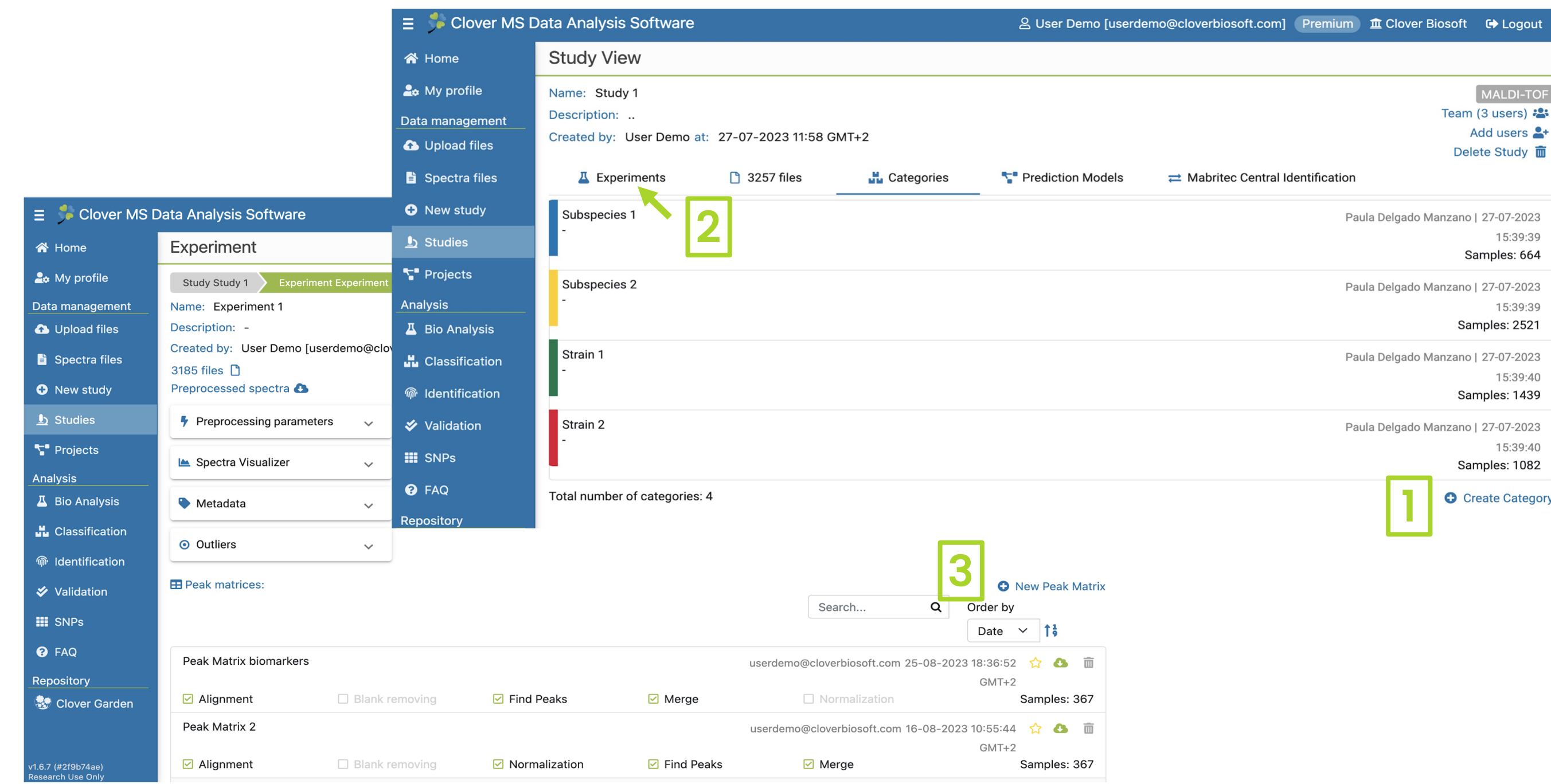


How to perform Hierarchical Clustering

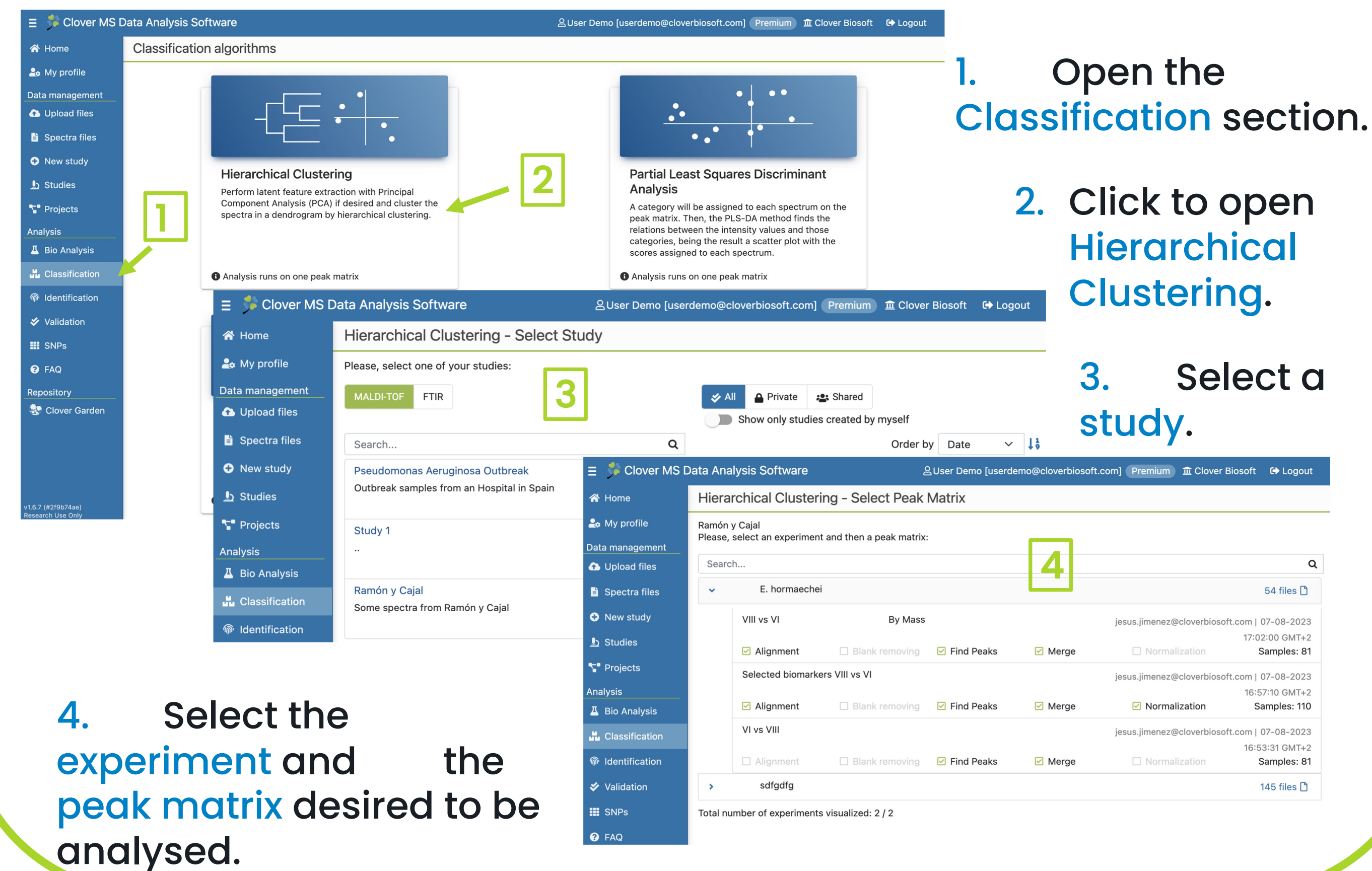
CLOVER MS Data Analysis Software

1 First steps

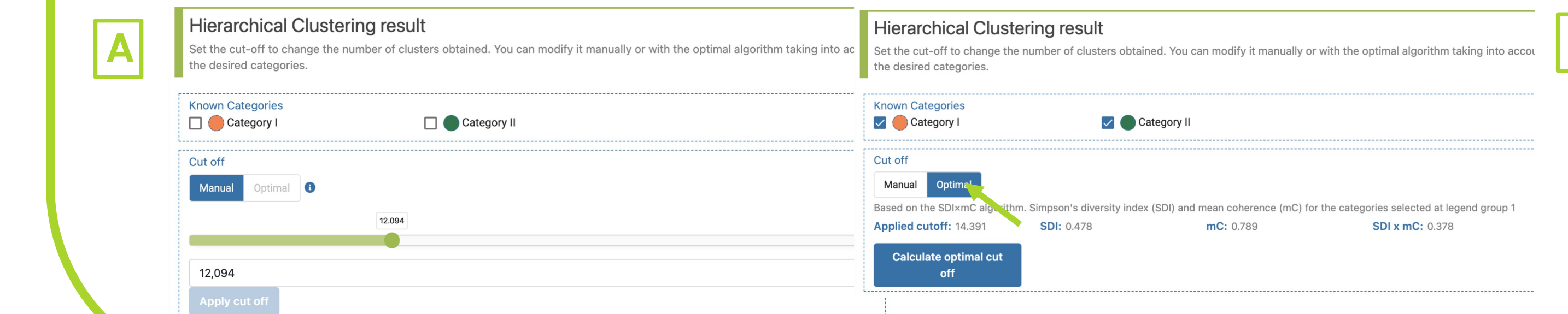
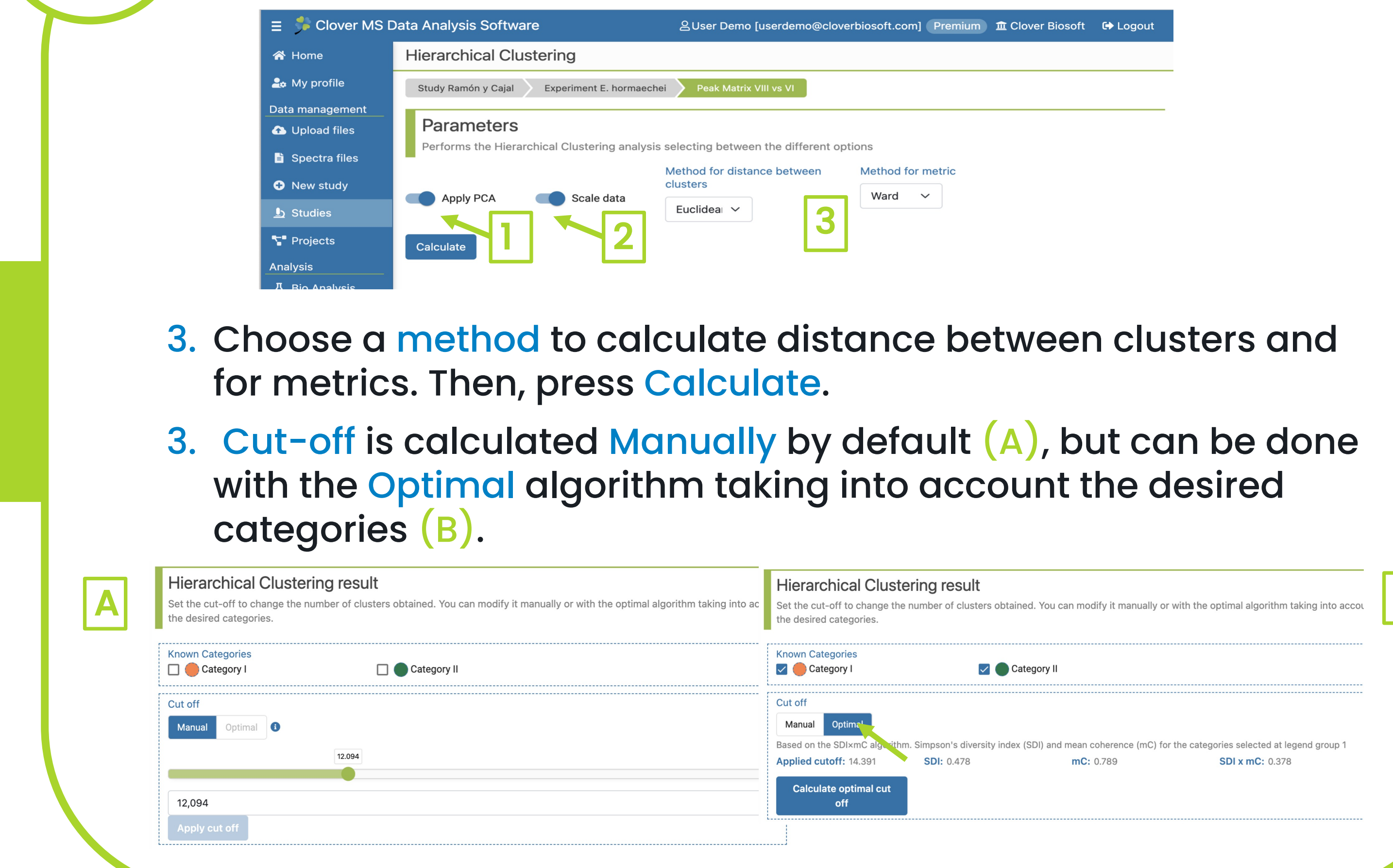
This guide is the same for MALDI and FTIR spectra



2 Hierarchical Clustering Analysis

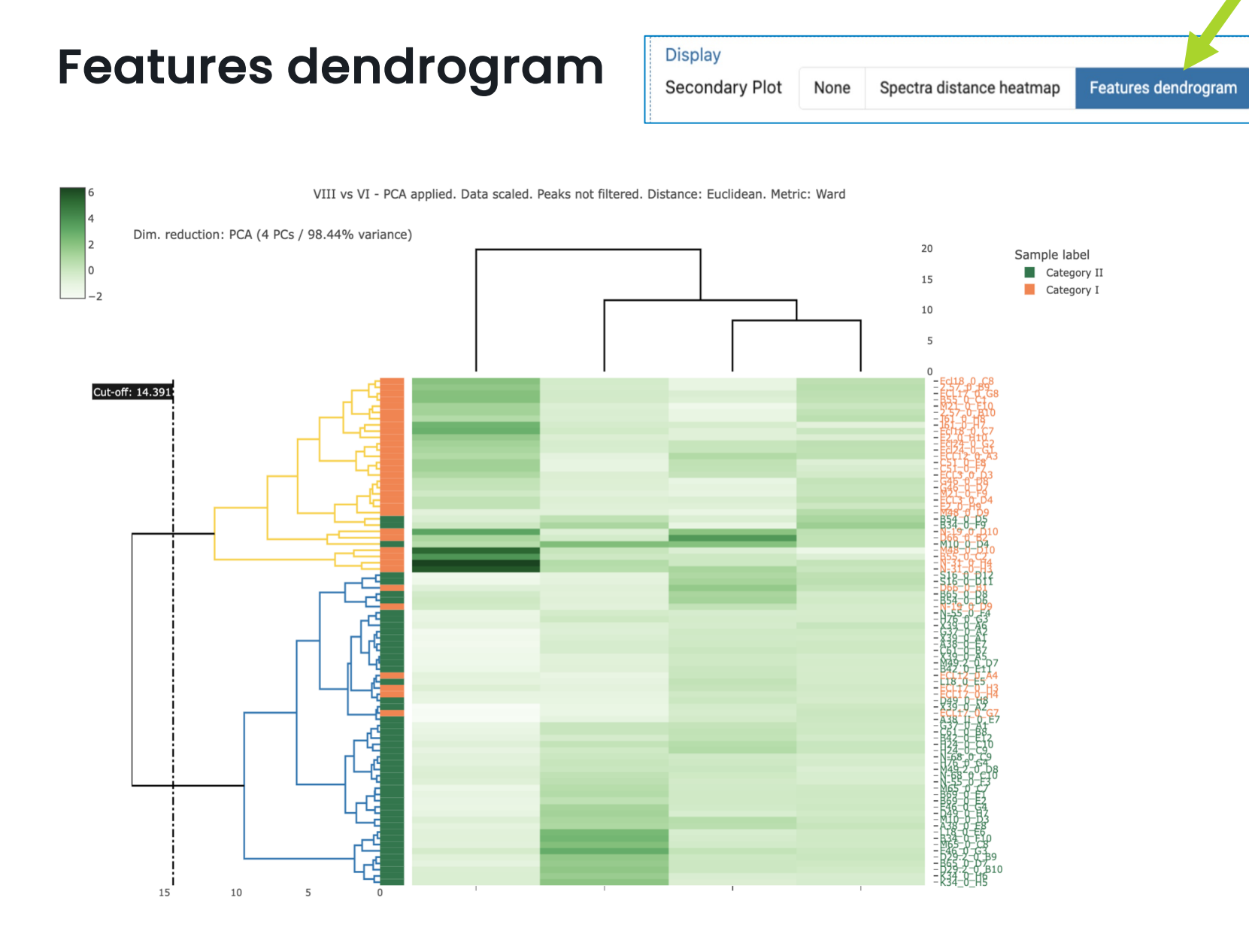
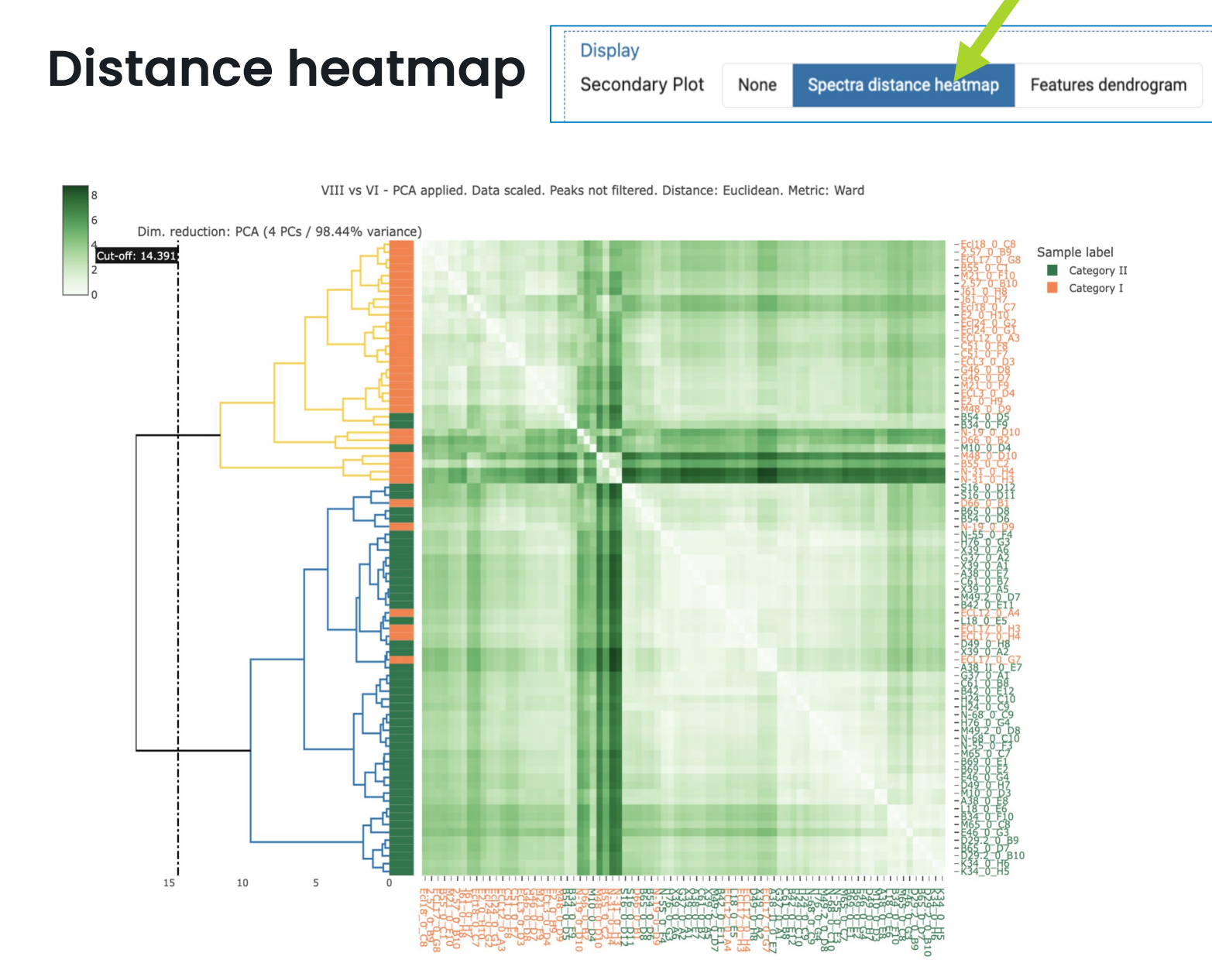
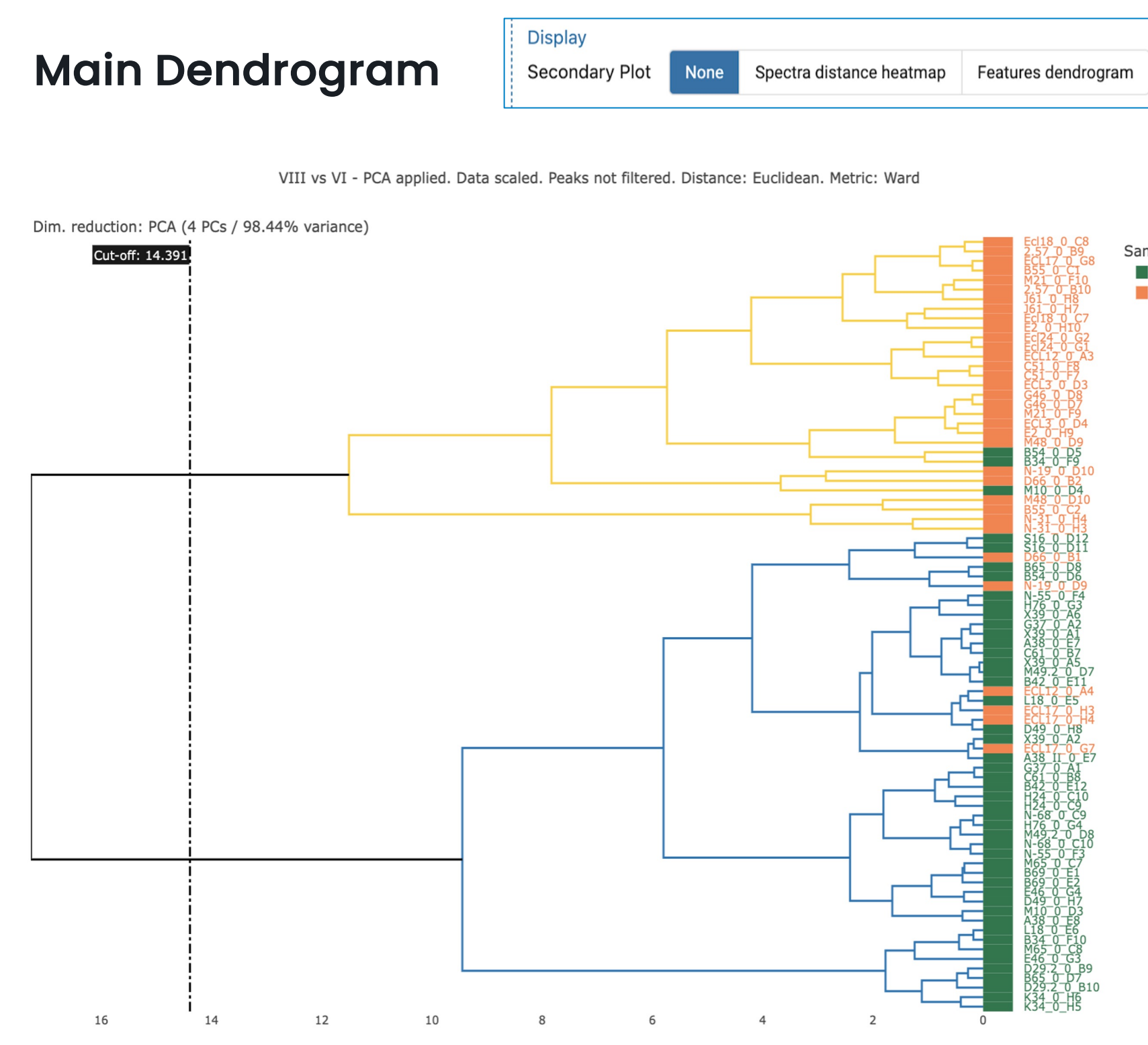


2.1 1. Apply (or not) PCA. 2. Scale (or not) your data.



3 HC Results

Apart from the main dendrogram, there are two secondary plots which become visible by selecting in **Display** section above:



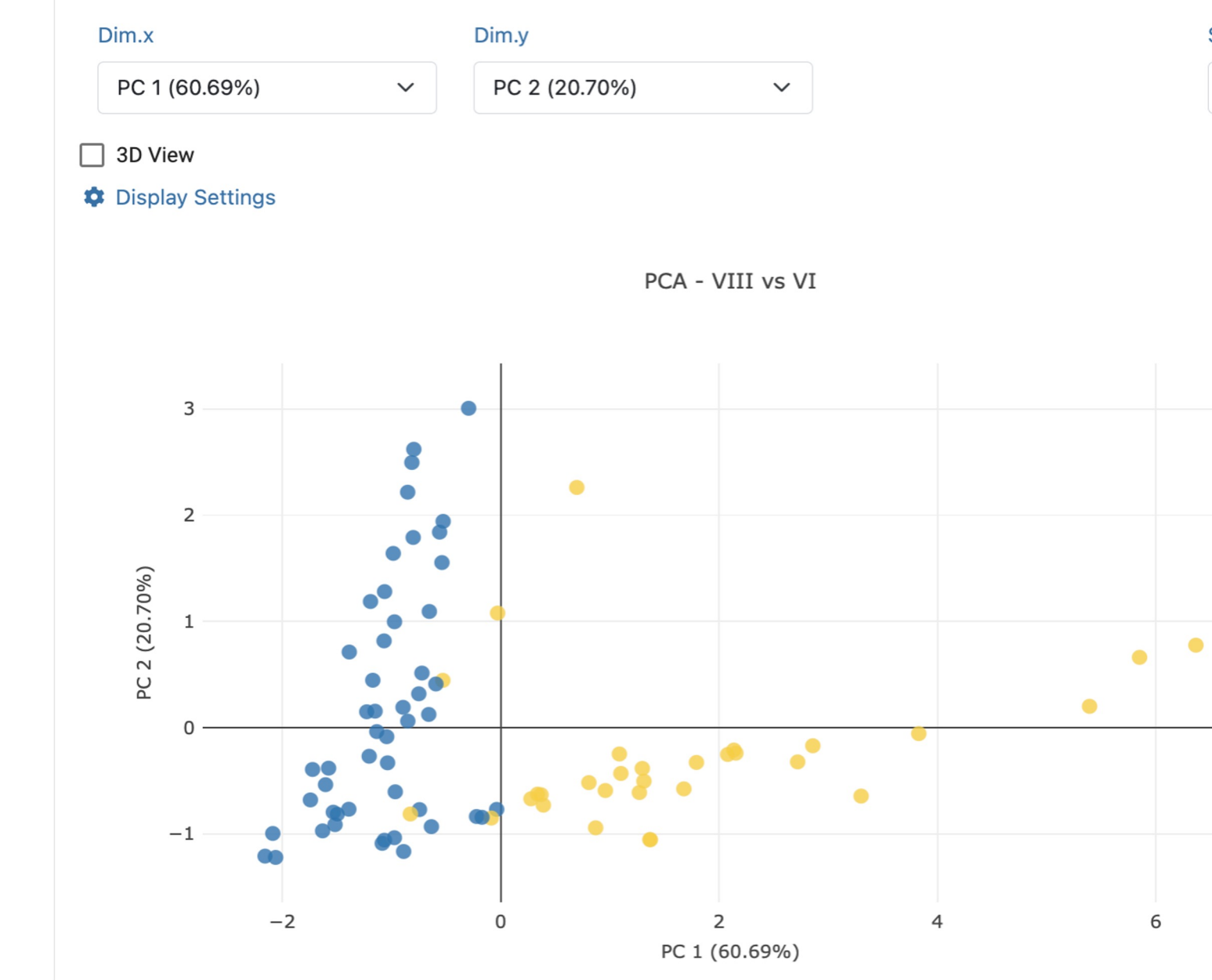
Category Summary

Category name ^ 1	Number of samples ↓	Cluster 1 ↓	Cluster 2 ↓
Category I	34	17.65 %	82.35 %
Category II	47	93.62 %	6.38 %

This table shows the percentages of samples that belong to the same category in each cluster created by Hierarchical Clustering.

- Colours only appear if categories are selected on the previous step.
- Cut-off can be changed and optimized according to the number of main categories.

PCA
If PCA was selected on step 2.1, it will display at the end of the results page.



- The algorithm, by default, chooses the minimum number of components to perform the PCA, having at least 95% of the variance.
- If PCA was selected, dendrograms are realized with PCA components rather than peaks.



cloverbiosoft.com
clovermsdataanalysis.com

Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone:
email:

+34 958 991 543
info@cloverbiosoft.com

support@cloverbiosoft.com

Clover MSDAS, HC Quick Start Guide



How to perform Hierarchical Clustering

CLOVER MS Data Analysis Software

1

First steps



This guide is the same for MALDI and FTIR spectra

Study View

Name: Study 1
Description: ..
Created by: User Demo at: 27-07-2023 11:58 GMT+2

Experiments | 3257 files | Categories | Prediction Models | Mabritec Central Identification

Subspecies	Created by	Date	Samples
Subspecies 1	Paula Delgado Manzano	27-07-2023 15:39:39	664
Subspecies 2	Paula Delgado Manzano	27-07-2023 15:39:39	2521
Strain 1	Paula Delgado Manzano	27-07-2023 15:39:40	1439
Strain 2	Paula Delgado Manzano	27-07-2023 15:39:40	1082

Total number of categories: 4

Experiment

Name: Experiment 1
Description: -
Created by: User Demo [userdemo@cloverbiosoft.com]
3185 files
Preprocessed spectra

Preprocessing parameters
Spectra Visualizer
Metadata
Outliers

Peak matrices:

Peak Matrix	Created by	Date	Samples
Peak Matrix biomarkers	userdemo@cloverbiosoft.com	25-08-2023 18:36:52 GMT+2	367
Peak Matrix 2	userdemo@cloverbiosoft.com	16-08-2023 10:55:44 GMT+2	367

Create Categories

Choose an Experiment

Create Peak Matrix



cloverbiosoft.com
clovermsdataanalysis.com



Clover Bioanalytical Software

Copyright© 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com

support@cloverbiosoft.com

Clover MSDAS, HC Quick Start Guide

2

Hierarchical Clustering Analysis

1

2

Hierarchical Clustering
Perform latent feature extraction with Principal Component Analysis (PCA) if desired and cluster the spectra in a dendrogram by hierarchical clustering.

Analysis runs on one peak matrix

1. Open **Classification** section.

2. Click to open **Hierarchical Clustering**.

3

Please, select one of your studies:

MALDI-TOF FTIR

Search...

Order by Date

Pseudomonas Aeruginosa Outbreak
Outbreak samples from an Hospital in Spain

Study 1
..

Ramón y Cajal
Some spectra from Ramón y Cajal

3. Select a **study**.

4

Please, select an experiment and then a peak matrix:

Search...

E. hormaechei 54 files

Peak Matrix	By Mass	Created
VIII vs VI	Alignment <input checked="" type="checkbox"/> Blank removing <input type="checkbox"/> Find Peaks <input checked="" type="checkbox"/> Merge <input checked="" type="checkbox"/> Normalization <input type="checkbox"/>	jesus.jimenez@cloverbiosoft.com 07-08-2023 17:02:00 GMT+2 Samples: 81
Selected biomarkers VIII vs VI	Alignment <input checked="" type="checkbox"/> Blank removing <input type="checkbox"/> Find Peaks <input checked="" type="checkbox"/> Merge <input checked="" type="checkbox"/> Normalization <input checked="" type="checkbox"/>	jesus.jimenez@cloverbiosoft.com 07-08-2023 16:57:10 GMT+2 Samples: 110
VI vs VIII	Alignment <input type="checkbox"/> Blank removing <input type="checkbox"/> Find Peaks <input checked="" type="checkbox"/> Merge <input checked="" type="checkbox"/> Normalization <input type="checkbox"/>	jesus.jimenez@cloverbiosoft.com 07-08-2023 16:53:31 GMT+2 Samples: 81

sdfgdfg 145 files

Total number of experiments visualized: 2 / 2

4. Select the **experiment** and **peak matrix** desired to be analysed.



cloverbiosoft.com
clovermsdataanalysis.com



Clover Bioanalytical Software

Copyright© 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com

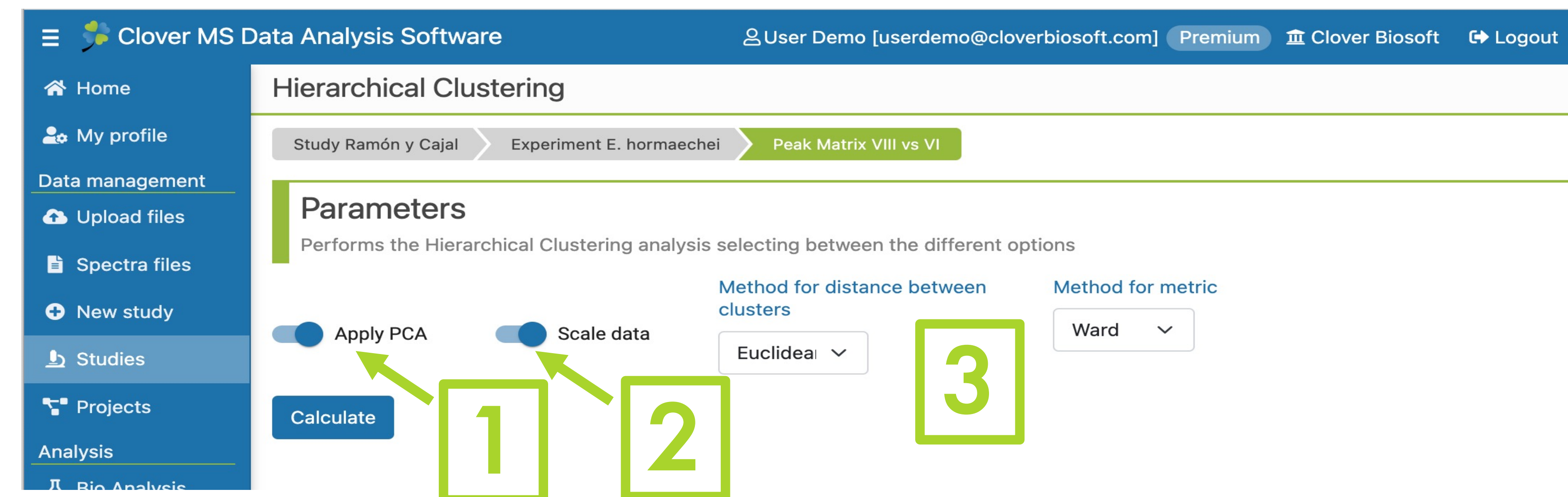
support@cloverbiosoft.com

Clover MSDAS, HC Quick Start Guide

2.1

1. Apply (or not) PCA.

2. Scale (or not) your data.



3. Choose a **method** to calculate distance between clusters and for metrics. Then press **Calculate**.

4. **Cut-off** is calculated **Manually** by default (A), but can be done with the **Optimal** algorithm taking into account the desired categories (B).

A Hierarchical Clustering result

Set the cut-off to change the number of clusters obtained. You can modify it manually or with the optimal algorithm taking into account the desired categories.

Known Categories
 Category I Category II

Cut off
Manual Optimal

12,094

Apply cut off

B Hierarchical Clustering result

Set the cut-off to change the number of clusters obtained. You can modify it manually or with the optimal algorithm taking into account the desired categories.

Known Categories
 Category I Category II

Cut off
Manual Optimal

Based on the SDIxmC algorithm. Simpson's diversity index (SDI) and mean coherence (mC) for the categories selected at legend group 1

Applied cutoff: 14.391 SDI: 0.478 mC: 0.789 SDI x mC: 0.378

Calculate optimal cut off



If any configuration is changed, it is needed to press **Calculate** again to see the new Dendrogram.



cloverbiosoft.com
clovermsdataanalysis.com



Clover Bioanalytical Software

Copyright© 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone:
email:

+34 958 991 543
info@cloverbiosoft.com

support@cloverbiosoft.com

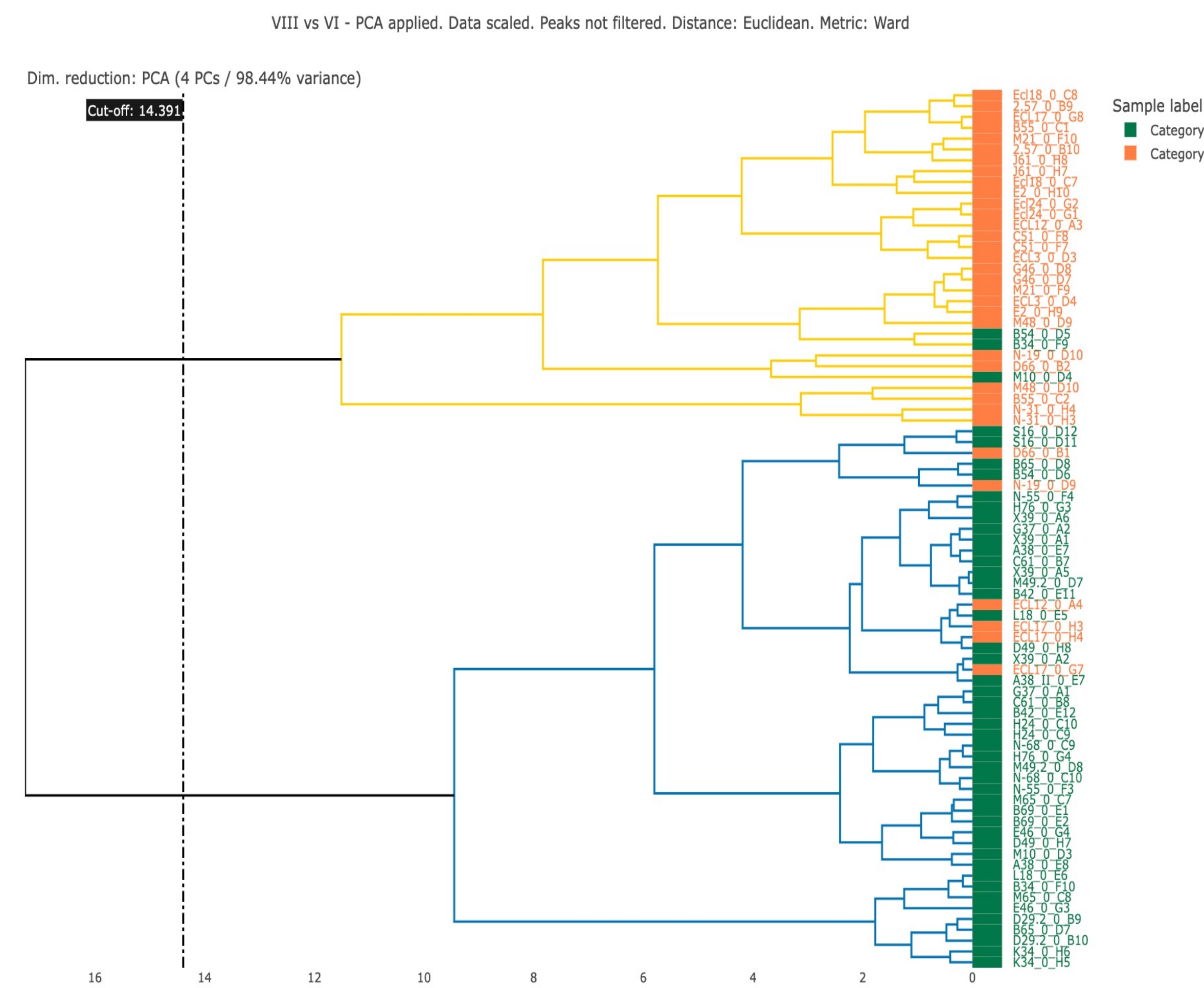
Clover MSDAS, HC Quick Start Guide

3 HC Results

Apart from the main dendrogram there are two secondary plots which become visible by selecting in **Display** section above:

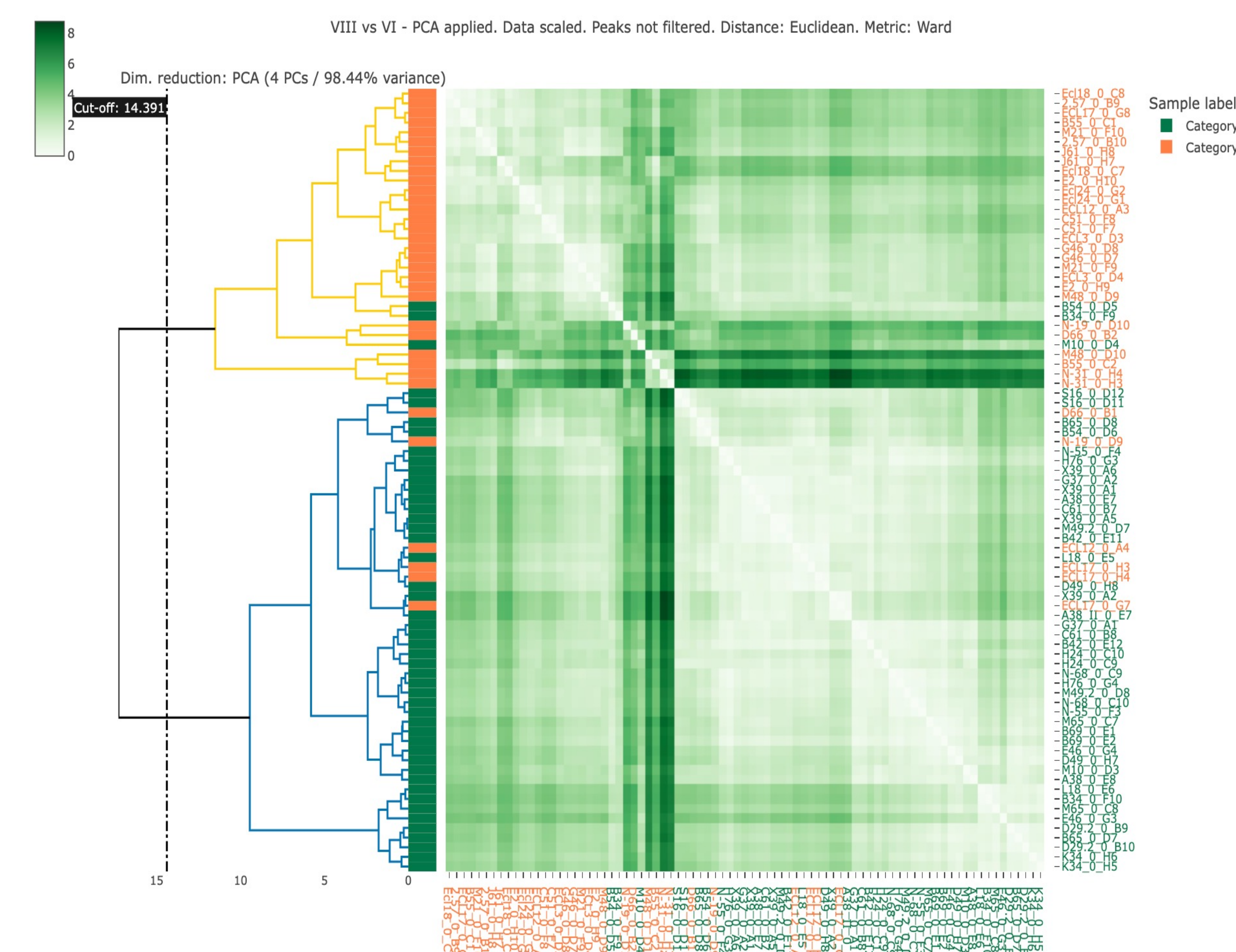
1. Main Dendrogram

Display
Secondary Plot: None Spectra distance heatmap Features dendrogram



1.1. Distance heatmap

Display
Secondary Plot: None Spectra distance heatmap Features dendrogram



1.2. Features dendrogram

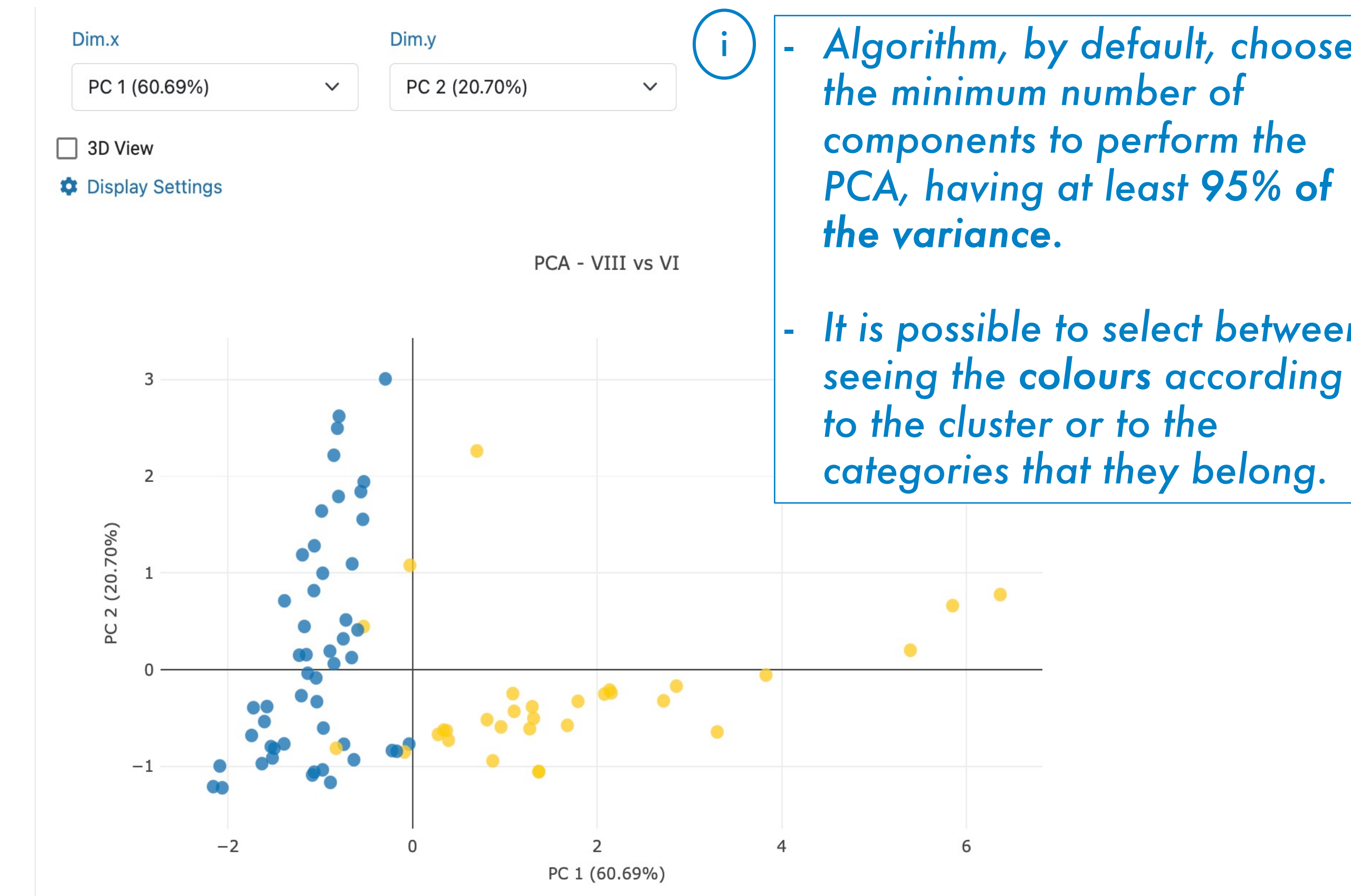
Display
Secondary Plot: None Spectra distance heatmap Features dendrogram

Features dendrogram is only useful if PCA is not applied, as it shows how the features affect the dendrogram and after performing PCA, features will be grouped on components.

- Colours only appear if categories are selected on the previous step.
- Cut-off can be changed and optimized according to the number of main categories.
- If PCA was selected dendrograms are realized with PCA components rather than peaks.

2. PCA

If PCA was selected on step 2.1, it will display at the end of the results page.



- Algorithm, by default, chooses the minimum number of components to perform the PCA, having at least 95% of the variance.
- It is possible to select between seeing the colours according to the cluster or to the categories that they belong.

3. Category Summary

Category name ^ 1	Number of samples ↓	Cluster 1 ↕	Cluster 2 ↕
Category I	34	17.65 %	82.35 %
Category II	47	93.62 %	6.38 %

This table shows the percentage of samples that belong to the same category in each of the clusters created by the Hierarchical Clustering. For example, there are 47 samples in Category II, of which 93.62% belongs to Cluster 1, a high purity level.

