

# Benchmarking Tool in R-shiny

(light version)

The screenshot shows the 'Files Selection' interface of the 'Clustering Methods Cor' application. The interface has a dark sidebar on the left with the following menu items: 'Files Selection', 'Clustering', 'Method Evaluation', and 'Methods Comparison'. The main content area is titled 'Files Selection' and contains two main sections: 'Reference File' and 'File 1'.

**Reference File Section:**

- Buttons: 'Reference File' (highlighted with a red box) and 'Populations Mapping File'.
- Checkbox: 'Use As Test'.
- Dropdown: 'Clusters column'.

**File 1 Section:**

- Button: 'File 1'.
- Dropdown: 'Clusters column'.

**Instructions:**

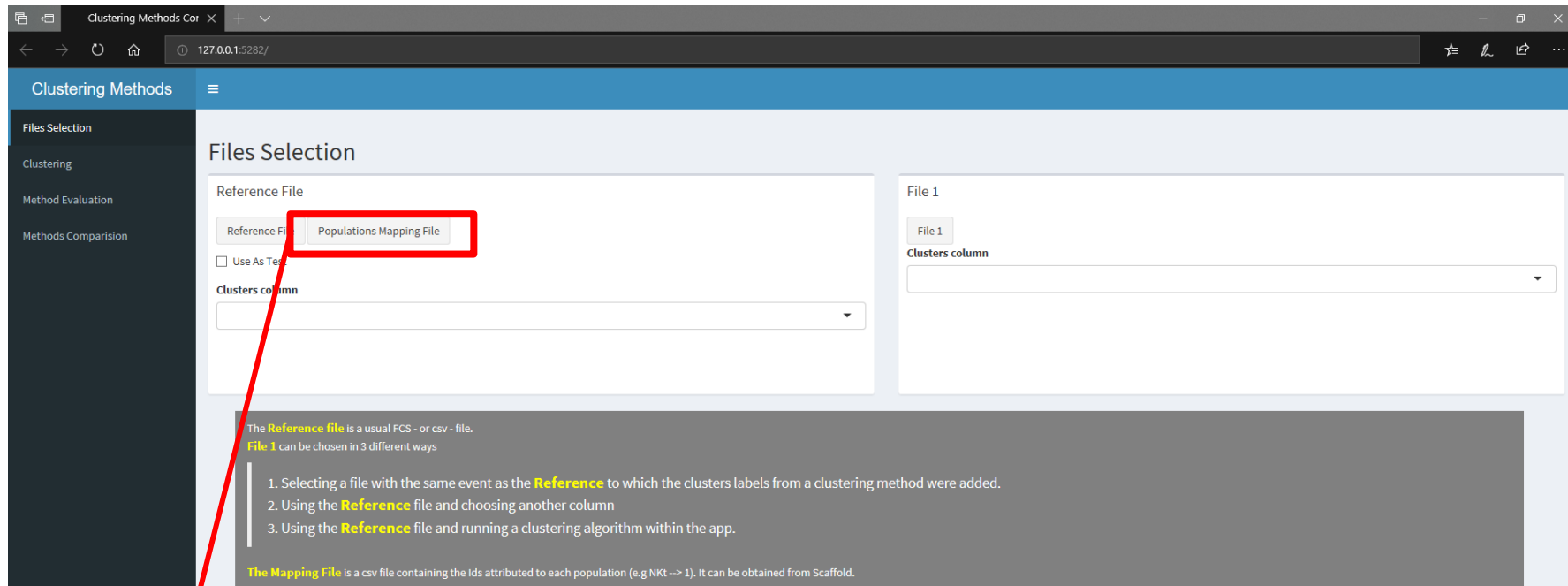
The **Reference file** is a usual FCS - or csv - file.  
**File 1** can be chosen in 3 different ways

1. Selecting a file with the same event as the **Reference** to which the clusters labels from a clustering method were added.
2. Using the **Reference** file and choosing another column
3. Using the **Reference** file and running a clustering algorithm within the app.

**The Mapping File** is a csv file containing the ids attributed to each population (e.g NKT → 1). It can be obtained from Scaffold.

A red arrow points from the 'Reference File' button to a text box below the interface.

Reference File containing the  
populations labels



Clustering Methods Cor

Files Selection

Reference File

Reference File | **Populations Mapping File**

☐ Use As Test

Clusters column

File 1

File 1

Clusters column

The **Reference file** is a usual FCS - or csv - file.  
**File 1** can be chosen in 3 different ways

1. Selecting a file with the same event as the **Reference** to which the clusters labels from a clustering method were added.
2. Using the **Reference** file and choosing another column
3. Using the **Reference** file and running a clustering algorithm within the app.

The **Mapping File** is a csv file containing the Ids attributed to each population (e.g NKt --> 1). It can be obtained from Scaffold.

Population Mapping File (optional) :  $2 \times N$  matrix giving the name of the  $N$  populations of our Input and attributing an Id to them (similar to what can be drawn from Scaffold)

eg: NKt --> 1

NK --> 2

Clustering Methods Cor x + v

127.0.0.1:5282/

Clustering Methods

Files Selection

Clustering

Method Evaluation

Methods Comparison

### Files Selection

Reference File

Reference File Populations Mapping File

☐ Use As Test

Clusters column

File 1

File 1

Clusters column

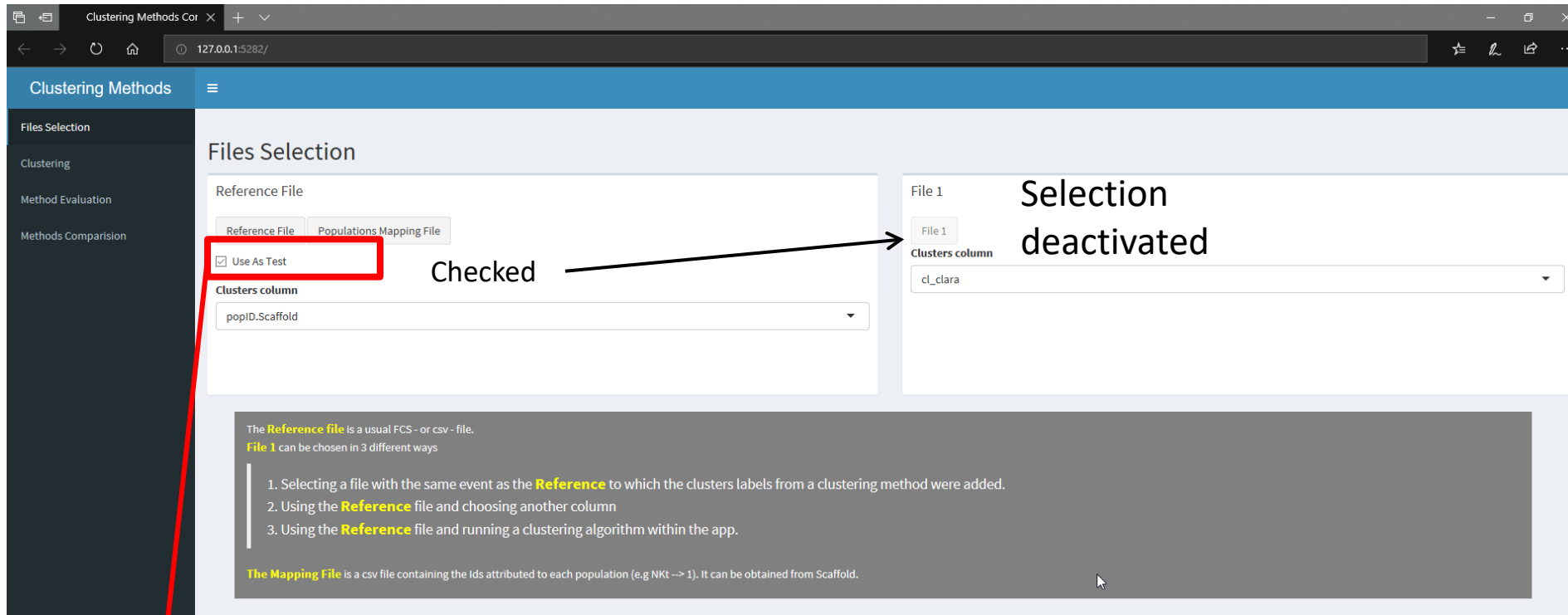
The **Reference file** is a usual FCS - or csv - file.  
**File 1** can be chosen in 3 different ways

1. Selecting a file with the same event as the **Reference** to which the clusters labels from a clustering method were added.
2. Using the **Reference** file and choosing another column
3. Using the **Reference** file and running a clustering algorithm within the app.

The **Mapping File** is a csv file containing the ids attributed to each population (e.g Nkt → 1). It can be obtained from Scaffold.

File 1: Contains the labels generating by a clustering method  
(only 1 column necessary)

OR



Clustering Methods Cor x + v

127.0.0.1:5282/

Clustering Methods

Files Selection

Clustering

Method Evaluation

Methods Comparison

Files Selection

Reference File

Reference File Populations Mapping File

☒ Use As Test

Clusters column

popID.Scaffold

Checked

File 1

Selection deactivated

File 1

Clusters column

cl\_clara

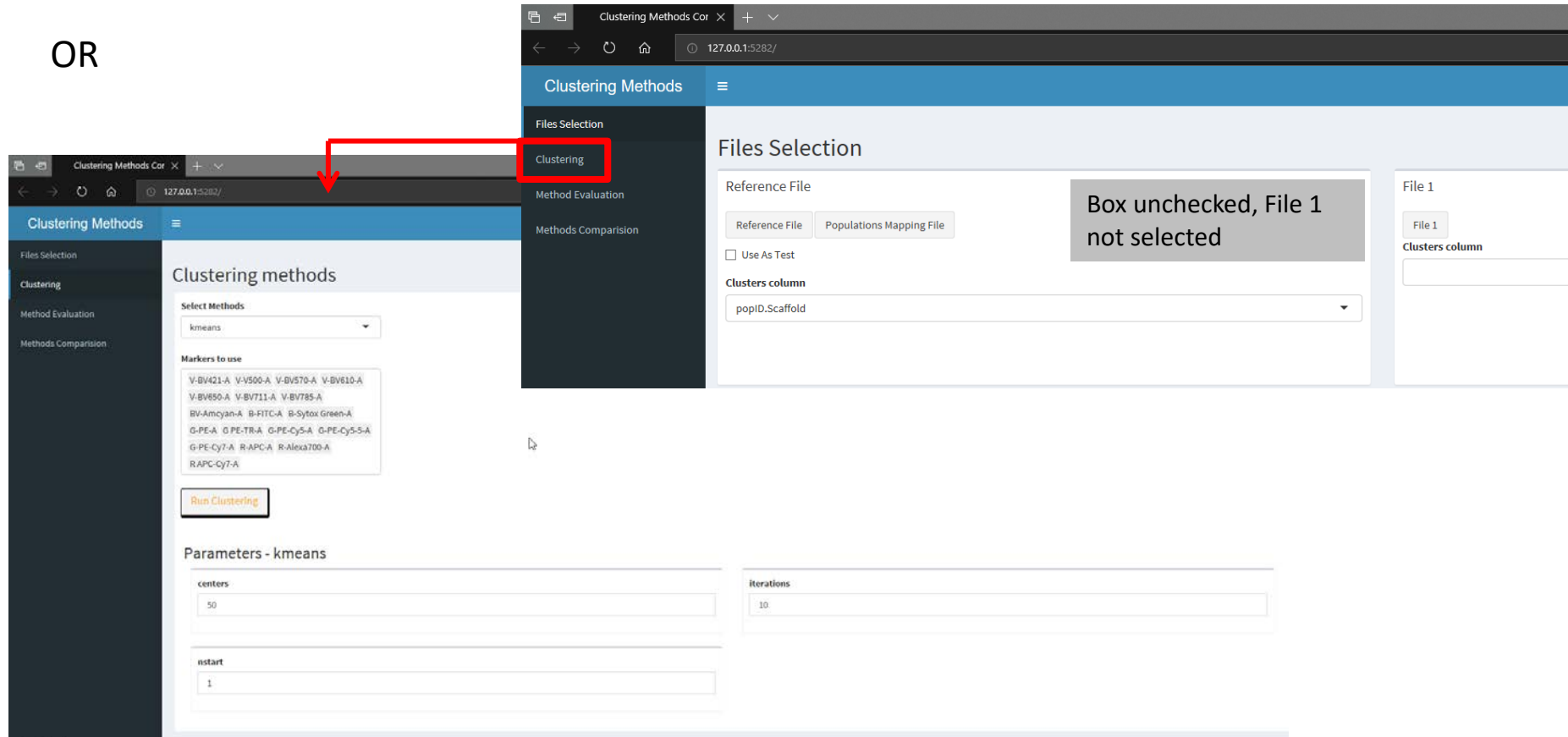
The **Reference file** is a usual FCS - or csv - file.  
**File 1** can be chosen in 3 different ways

1. Selecting a file with the same event as the **Reference** to which the clusters labels from a clustering method were added.
2. Using the **Reference** file and choosing another column
3. Using the **Reference** file and running a clustering algorithm within the app.

The **Mapping File** is a csv file containing the Ids attributed to each population (e.g NKT → 1). It can be obtained from Scaffold.

If the reference file also contains the results from a clustering method, it can be used

OR



Clustering Methods Cor

Clustering Methods

Files Selection

Clustering

Method Evaluation

Methods Comparison

Clustering methods

Select Methods

kmeans

Markers to use

V-BV421-A V-V500-A V-BV570-A V-BV610-A  
V-BV650-A V-BV711-A V-BV785-A  
BV-Amcyan-A B-FITC-A B-Sytox Green-A  
G-PE-A G-PE-TR-A G-PE-Cy5-A G-PE-Cy5.5-A  
G-PE-Cy7-A R-APC-A R-Alexa700-A  
RAPC-Cy7-A

Run Clustering

Parameters - kmeans

centers

50

iterations

10

File 1

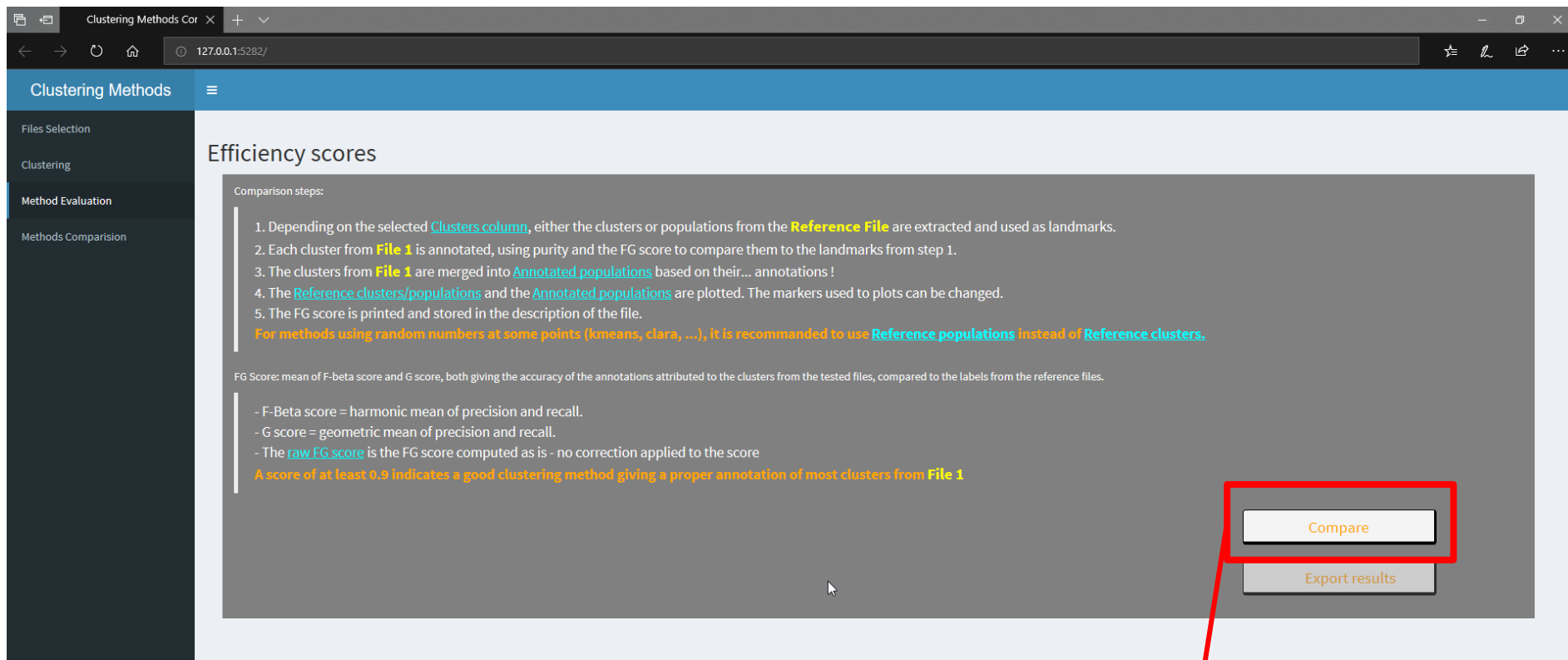
File 1

Clusters column

popID.Scaffold

Box unchecked, File 1 not selected

Running a fast clustering method on the reference File and using the result as the tested Output



Clustering Methods Cor

127.0.0.1:5282/

Clustering Methods

Files Selection

Clustering

Method Evaluation

Methods Comparison

## Efficiency scores

Comparison steps:

1. Depending on the selected **Clusters column**, either the clusters or populations from the **Reference File** are extracted and used as landmarks.
2. Each cluster from **File 1** is annotated, using purity and the FG score to compare them to the landmarks from step 1.
3. The clusters from **File 1** are merged into **Annotated populations** based on their... annotations !
4. The **Reference clusters/populations** and the **Annotated populations** are plotted. The markers used to plots can be changed.
5. The FG score is printed and stored in the description of the file.

**For methods using random numbers at some points (kmeans, clara, ...), it is recommended to use Reference populations instead of Reference clusters.**

FG Score: mean of F-beta score and G score, both giving the accuracy of the annotations attributed to the clusters from the tested files, compared to the labels from the reference files.

- F-Beta score = harmonic mean of precision and recall.
- G score = geometric mean of precision and recall.
- The **raw FG score** is the FG score computed as is - no correction applied to the score

**A score of at least 0.9 indicates a good clustering method giving a proper annotation of most clusters from File 1**

Compare

Export results

Compute the F and G score and their mean value (other data hidden for now)

Export results

Export Reference file and add cluster and annotated groups labels, keywords indicating the mean scores and the columns used  
Export a generated/updated mapping file giving the F score corresponding to each population

FILE: transformed\_estimate\_clara

ACCURACY (FG) = 0.577599999999938

Clustering Method Highly Innacurate

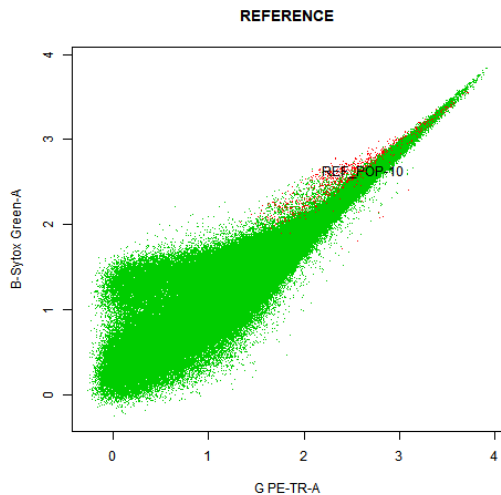
Raw FG score = 0.5776

Theta correction = 6.16499075180869e-14

Fbeta score = 0.5695

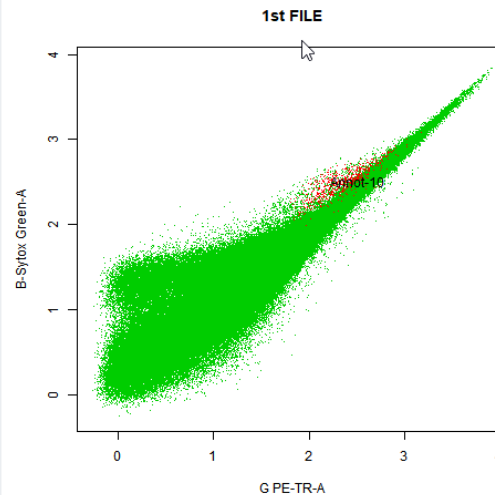
G score = 0.585855555555556

Print mean scores



Reference populations/clusters

REF\_POP-10



Annotated populations

Annot-10

Marker 1

G PE-TR-A

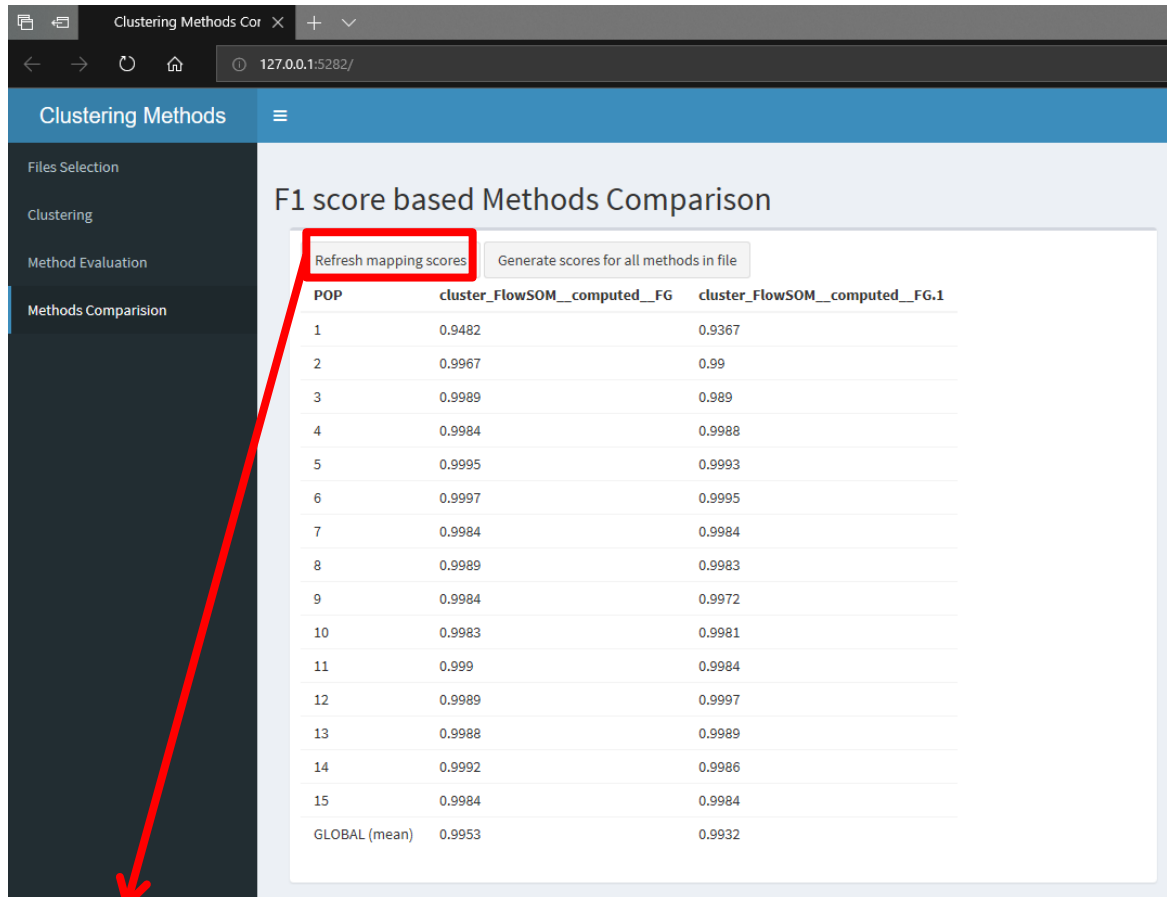
Marker 2

B-Sytox Green-A

Markers to use for the plots

Choose Input populations and Output annotated groups to highlight (in red)





Clustering Methods Cor x + v

127.0.0.1:5282/

Clustering Methods

Files Selection

Clustering

Method Evaluation

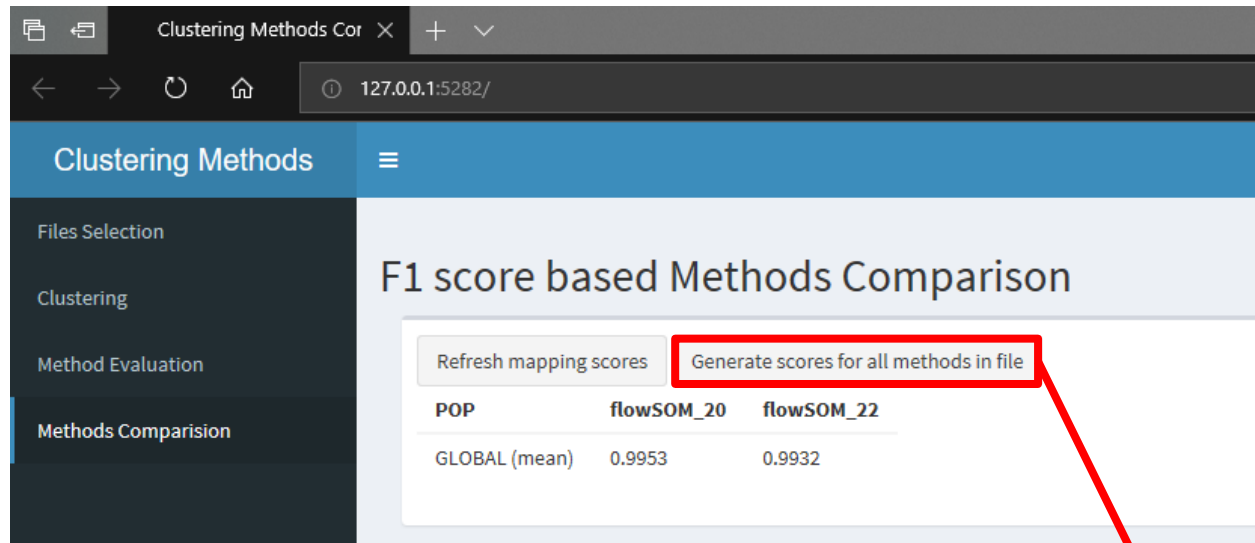
Methods Comparison

## F1 score based Methods Comparison

Refresh mapping scores Generate scores for all methods in file

POP	cluster_FlowSOM_computed_FG	cluster_FlowSOM_computed_FG.1
1	0.9482	0.9367
2	0.9967	0.99
3	0.9989	0.989
4	0.9984	0.9988
5	0.9995	0.9993
6	0.9997	0.9995
7	0.9984	0.9984
8	0.9989	0.9983
9	0.9984	0.9972
10	0.9983	0.9981
11	0.999	0.9984
12	0.9989	0.9997
13	0.9988	0.9989
14	0.9992	0.9986
15	0.9984	0.9984
GLOBAL (mean)	0.9953	0.9932

Print the F score for each population and each method using the mapping file



Clustering Methods Cor X + v

127.0.0.1:5282/

Clustering Methods

Files Selection

Clustering

Method Evaluation

Methods Comparison

## F1 score based Methods Comparison

Refresh mapping scores Generate scores for all methods in file

POP	flowSOM_20	flowSOM_22
GLOBAL (mean)	0.9953	0.9932

Print the mean F score for each method using the keywords stored in the Reference file