# Tutorial for importing data into CaseSolver

This is a graphical step-by-step tutorial for importing the RealCase1 data into CaseSolver.

Extract the folder named "Cases" from the zip file to an easy assessible destination on your system.

After installation (see INSTALLATION text file), open CaseSolver by typing the following in R:
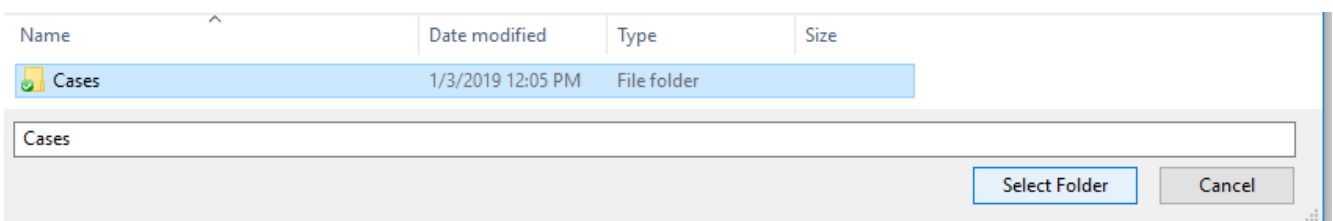library(casesolver);gui()

Then this window will show up (when "Case directory" is not correctly specified):
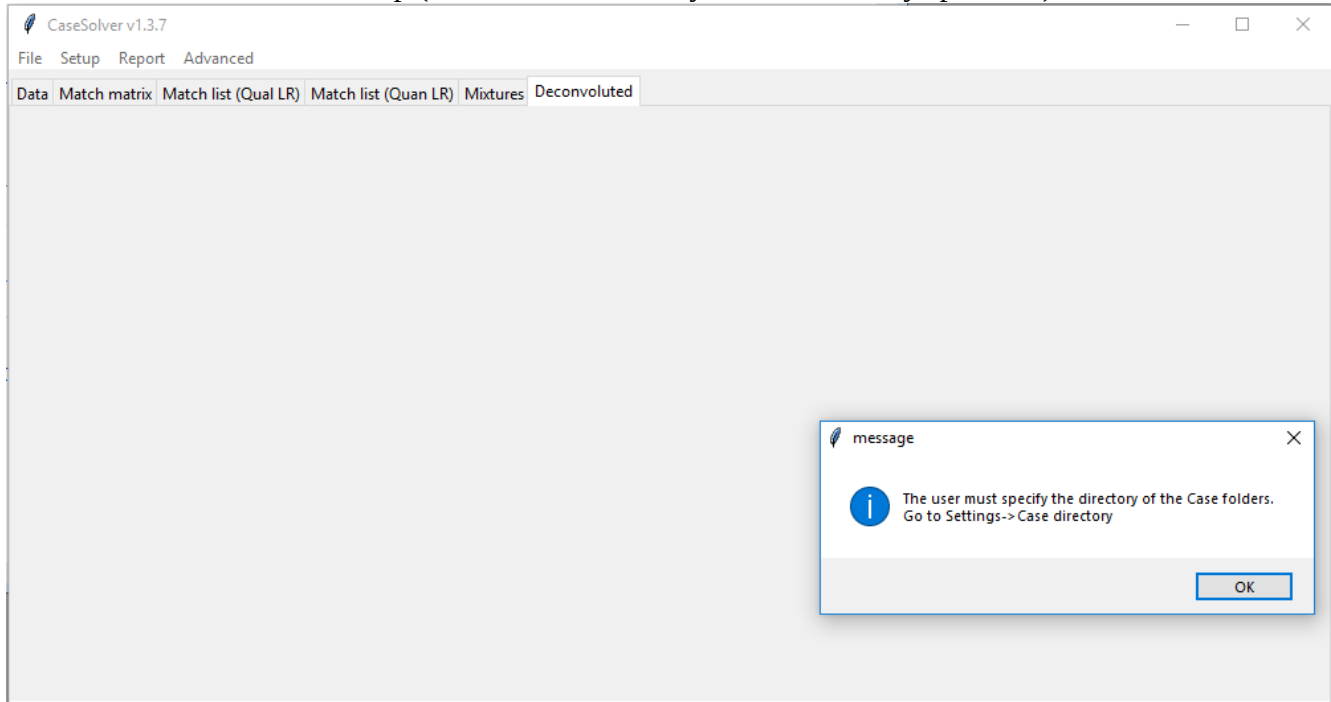
As the message suggests, we first need to specify the directory with the "case-folders".
Click "OK" and and go to "Setup" and click "Set case directory":
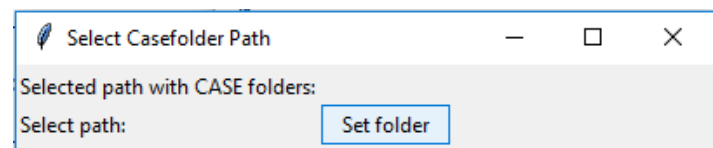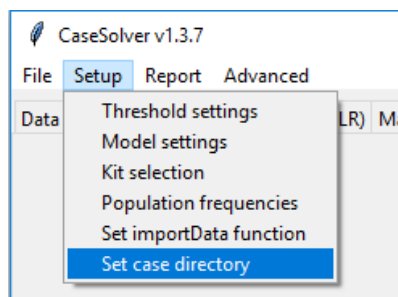
Then click on "Set folder" and select the folder named "Cases" (extracted from the zip file) by clicking "Select folder":

This folder contains the case-folder "RealCase1" which again contains the evidence profiles (the file "evids") and the reference profiles (the file "refs"). The purpose with this structure is that you can have hundreds of cases (with names given by the folder name) with data that instantly can be imported into CaseSolver. The software quits and restarts when a new directory is selected:
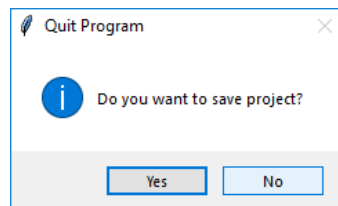

Select "No" for the question "Do you want to save project?". The GUI should now be changed with the possibility to select cases in the drop-down meny:
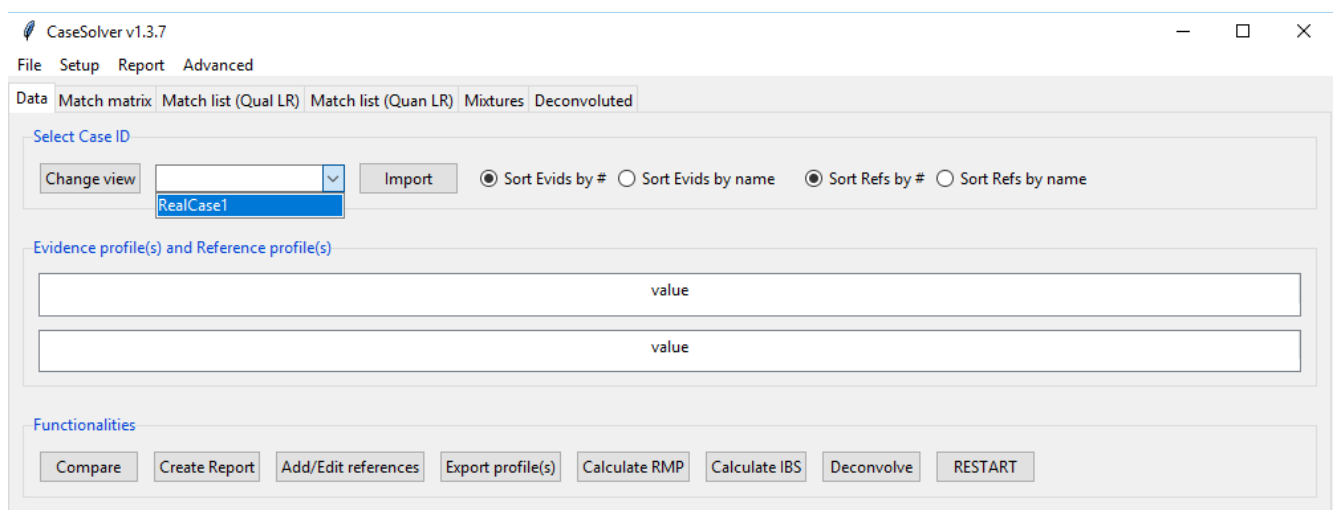
The name of the cases in the drop-down menu is taken from the folder names. If the drop-down menu does not contain your case names you have misspecified the "Casefolder path".
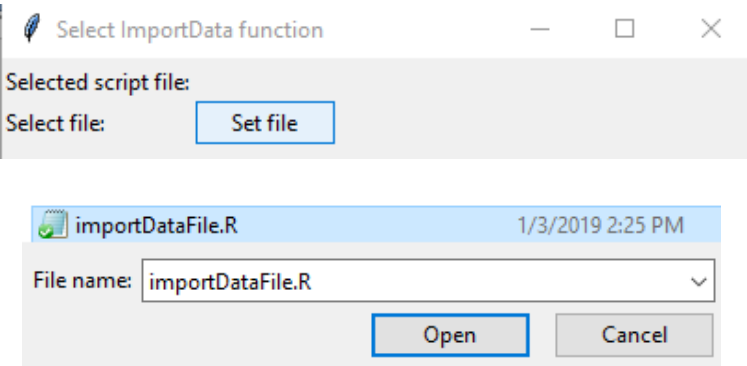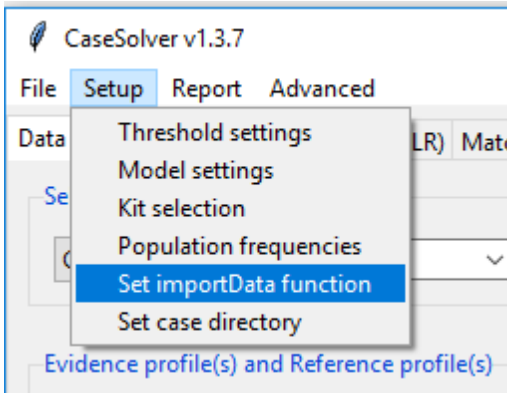
Select the case "RealCase1" and click "Import". The following message will occur:
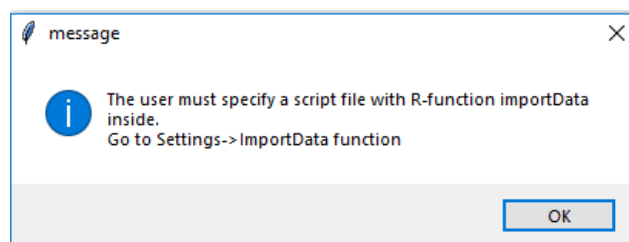
This message means that a R-file containing the R-function importData hasn't been specified. The purpose of this function is to translate the evidence/reference data in a casefolder to a format which can be recognised by CaseSolver; hence the output from this function must be given in a particular format. Importantly, the input of this function is only a text string variable which will be the name of one of the files in the selected case folder (for instance "evids.csv" if the case "RealCase1" is selected).

To select this file, select "Setup"-> "Set importData function" and select "Set file" and select the file with name "importDataFile" in the "Cases" folder. The name of the file is not important as long as it

contains the R-function importData.

Select the case "RealCase1" in the drop-down list and click the "Import" button the get the following data imported:

If no data is shown, something wrong happended in the importData function.

Set settings for analysing data in CaseSolver

It now remains for the user to specify other settings in CaseSolver for further analysing of the data. These are following settings: 1) "Population frequencies"
2) "Kit selection"
3) "Model settings"
4) "Threshold settings"
5) Advanced

Setup	Report	Advanced
Threshold settings		
Model settings		
Kit selection		
Population frequencies		
Set importData function		
Set case directory		
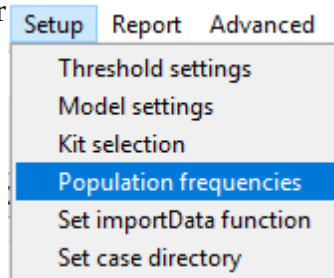
1) Population frequencies
The population frequency file must be in the same format as used for LRmix Studio and EuroForMix. Importing frequencies into CaseSolver is necessary for most of the functionalities.

To import the frequencies go to "Setup" → "Population frequencies" and click "Select file". Select the file with name "freqs.csv".
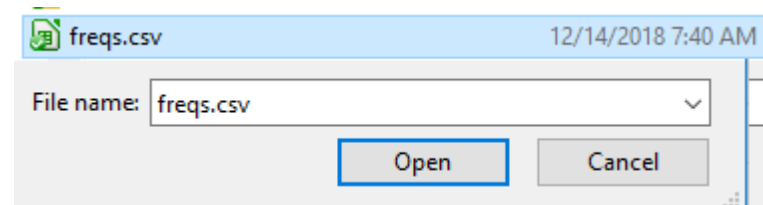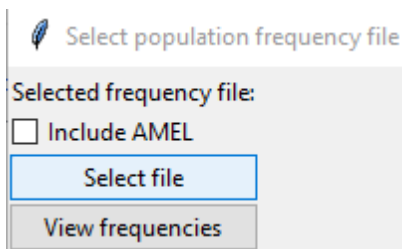
The user may select to include the AMEL marker to the allele frequencies as part of the statistical model (X=.75, Y=.25 given as allele frequencies). We did not use this in the paper.

To check that the allele frequencies are imported, go to "Setup" → "Population frequencies", check that the path to the file name is there and click "View frequencies" to check that the data are correct:
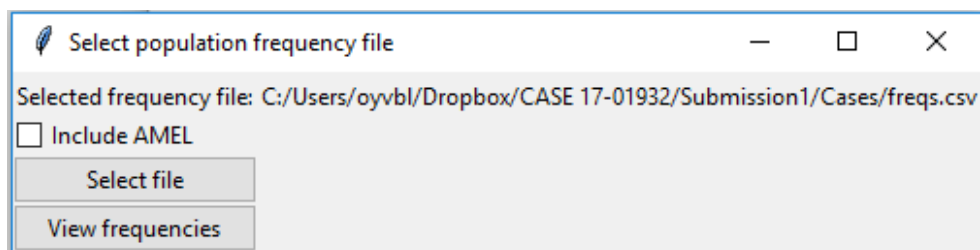
2) Kit selection

Necessary for:
- "visualizing epg by (left) double-clicking profiles" in the Data panel.
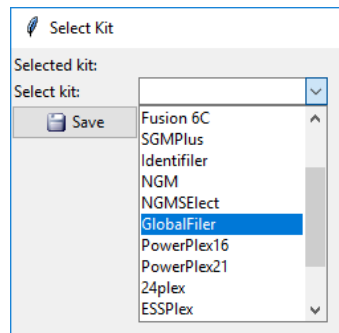*- using the degradation model when using the quantitative model for LR*

To select the kit used for your data go to "Setup" → "Kit selection " and click on "GlobalFiler" in the dropdown menu:

CaseSolver uses the euroformix function "getKit" to get this list. When kit is selected click "Save". The name of the current selected kit in CaseSolver is indicated under under "Selected kit" under "Setup" → "Kit selection ":
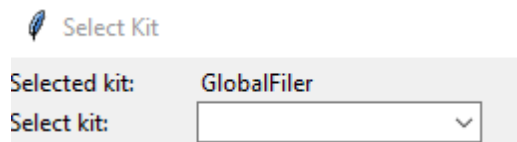
Visualising evid profiles (with added ref profiles) are now possible by double clicking profiles in the Data panel tables. By (left) double-clicking on evid sample 20.01 and then (left) double-clicking on ref sample REF1 gives (selecting another evid sample will reset the selected ref samples):
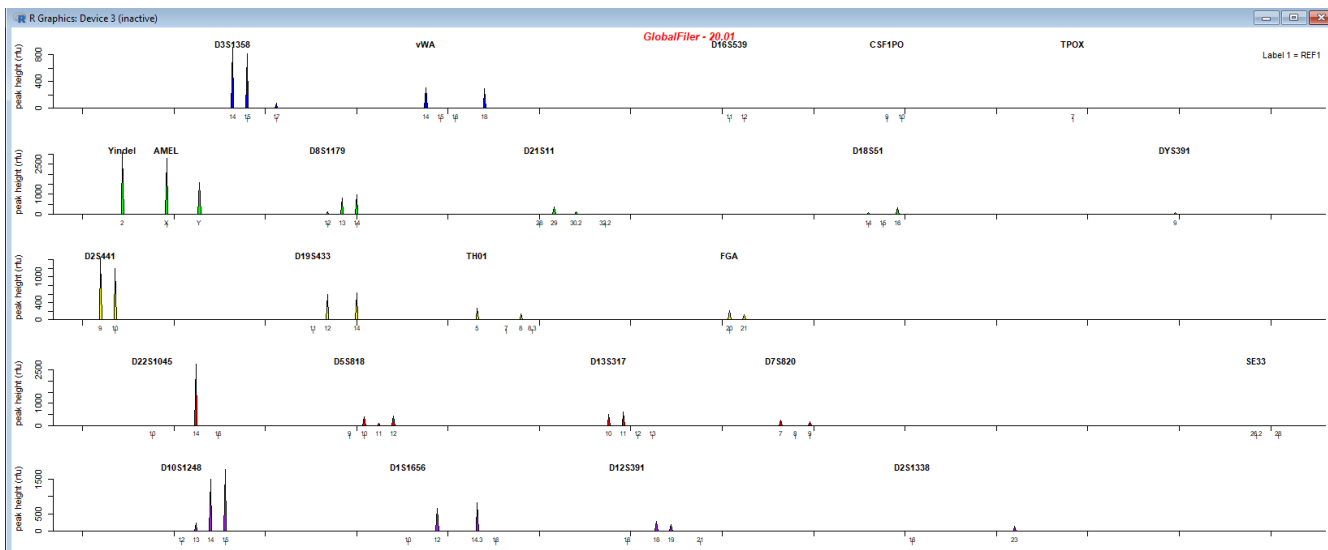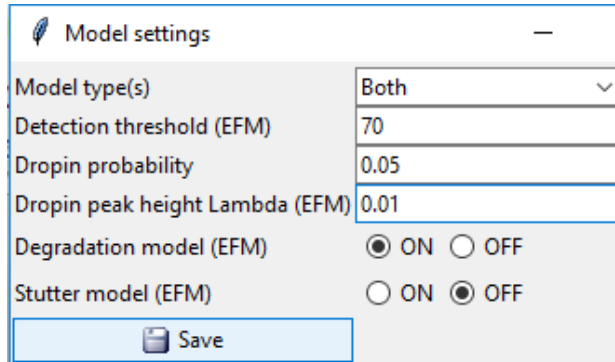

The epg will be shown in the R GUI (as for the euroformix software).

<u>3) Model settings</u>
Necessary for specifying statistical models used to calculate LR for mixture comparisons.

Go to "Setup" → "Model settings" to select the model specification for calculating LR. Click "Save" to save selected settings. In the paper we used the following:
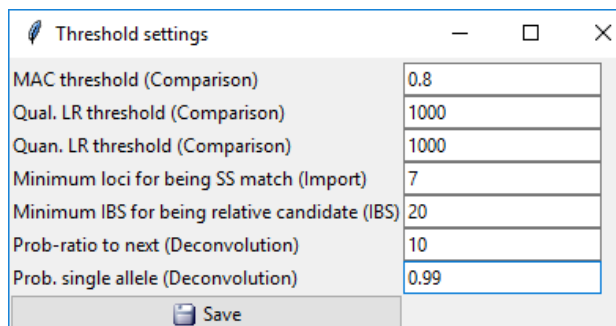

Here "Model type(s)" indicate what search strategy that should be used in the mixture comparison. "Both"=First LRmix and then EuroForMix, "Qualitative=Only LRmix" or "Quantiative=Only EFM".

<u>4) Threshold settings</u>
Necessary for:
- *Thresholds for obtaining candidate lists when comparing mixtures.*
- *Single source profiles comparisons.*
- *Deduction of deconvoluted profiles (degree of certainty)*

Go to "Setup" → "Threshold settings" to select the different thresholds used in CaseSolver. Click "Save" to save selected settings. In the paper we used the following (but we dynamically changed the 3 first settings):
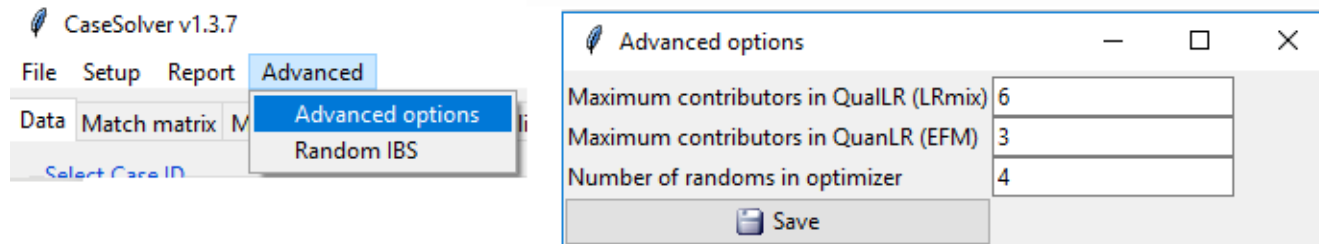
5) Advanced option
This setting is related to computation speed vs robustness for the mixture comparison.

The two first settings regards the upper limit of the number of contributors to be assigned to an evidence profile in the mixture comparison. LRmix is fast compared to EFM and hence a higher limit is given by default. In the paper we dynamically changed the "Maximum contributors in QuanLR (EFM)" to 4.

"Number of randoms in optimizer" is same as "Select number of random startpoints" for the "Optimization setting" in the EuroForMix software; Ensuring that a global maximum (MLE) has been obtained.

Analysing data in CaseSolver

See other teaching material/tutorial at http://euroformix.com/casesolver for how this is carried out.

Simulating the 1 mill. large database as done in paper

Create a copy of the allele frequency file "freqs.csv" and place it under the folder "RealCase1" (among the evid/ref files). CaseSolver will then recognise the file and send it as input in the "importData" function and then active last block in the code where reference profiles are generated. Importantly, the number of references to generate is given by variable "M" which is set to 1e6 by default. Other database sizes can be tried out by changing this value.

Useful note when importing large files into R:
The function fread from the R-package data.table is fast for reading large text files.