A beginner's guide to CaseSolver

A program based on EuroForMix
for analysing case data

Version 1.7 (compiled April 2020)

by Øyvind Bleka, Oslo University Hospital

About

- A GUI R-program for comparing and visualizing DNA profiles

- Most important functionality is profile comparisons:
  - References to evidence profiles
  - Single source profiles to other evidence profiles
  - Can utilize peak heights (EuroForMix)

- Mixture comparison follows a step-wise strategy as first given in

Forensic Science International: Genetics Supplement Series

Volume 6, December 2017, Pages e404-e406

dnamatch2: An open source software to carry out large scale database searches of mixtures using qualitative and quantitative models

Ø. Bleka [a], M. Bouzga [a], A. Fonneløp [a], P. Gill [a, b]

- Suitable for cases with large amount of profiles

- Detailed mixture analyzes using EuroForMix
  - Advanced deconvolution module
  - Show ”Model fitted Peak heights”
  - Automatic LR calculations
  - Hypothesis-tailored LR calculations

- Customize friendly:
  - Design your own data import function
  - Design your own strategy/report
  - Kits from euroformix supported

- Supports visualization of MPS profiles (STR or SNP)

Publication

ELSEVIER

Forensic Science International: Genetics

Volume 41, July 2019, Pages 83-92

CaseSolver: An investigative open source expert system based on EuroForMix

Øyvind Bleka [a], Lourdes Prieto [b], Peter Gill [a, c]

Demonstrates the use of CaseSolver in a very complicated case

- Paper Data available at http://euroformix.com/datasets
- A tutorial is available in the *Cases* folder in the zip-file

Installation/Startup

- Software R installed (newer than version 3.0.1)
- Install required R-packages from CRAN:

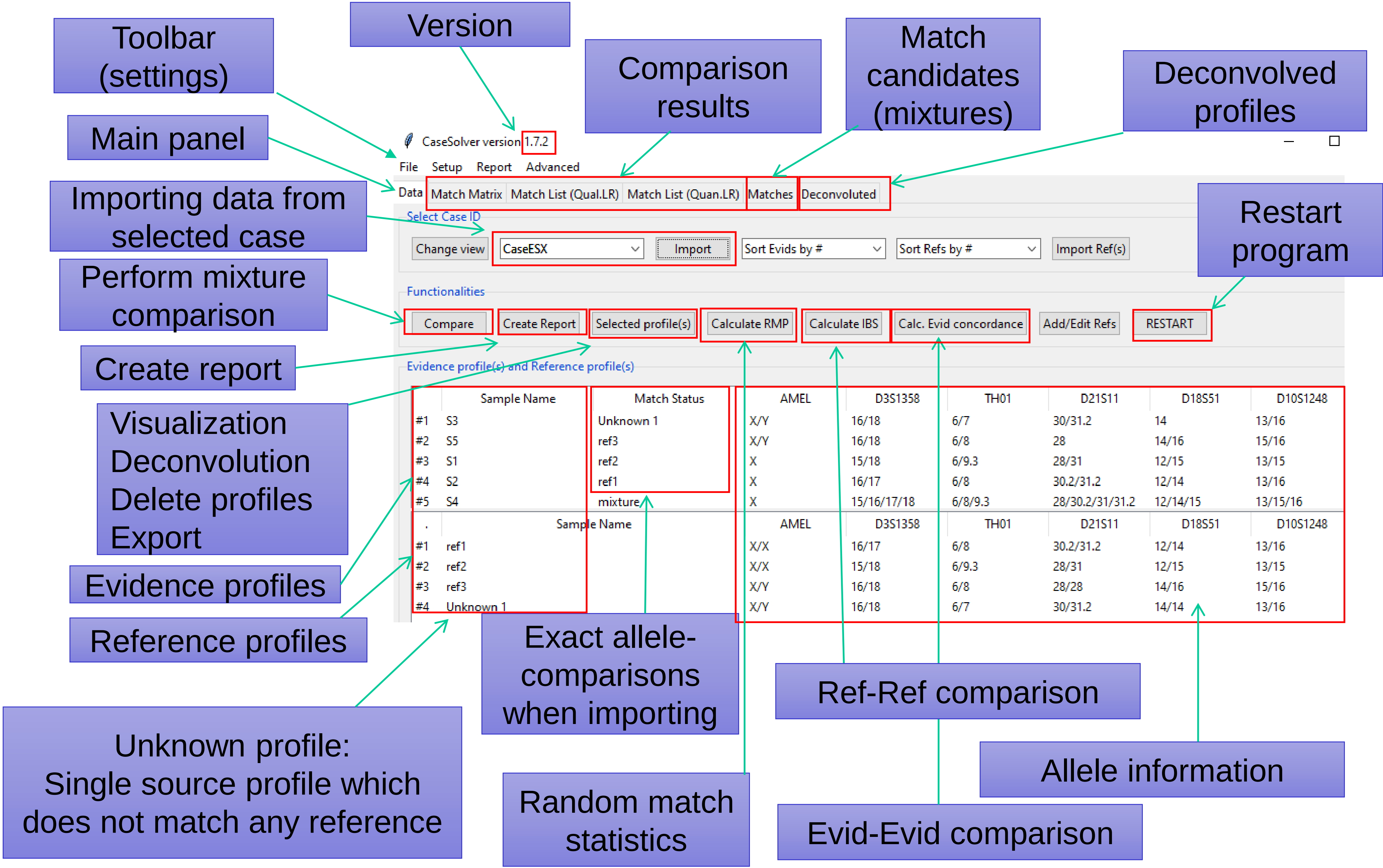
```
install.packages(c("gWidgets2tcltk","forensim","R2HTML","igraph","readxl","plotly"),dependencies=c("Depends", "Imports"))
```

- The R-package euroformix (newer than 2.0.1)
  - See http://euroformix.com for installation
- The R-package casesolver
  - See http://euroformix.com/casesolver for installation

- Start-up:
  - Open R and type (copy-paste command)

```
library(casesolver);gui()
```

Overview of the CaseSolver GUI



	Sample Name	Match Status	AMEL	D3S1358	TH01	D21S11	D18S51	D10S1248
#1	S3	Unknown 1	X/Y	16/18	6/7	30/31.2	14	13/16
#2	S5	ref3	X/Y	16/18	6/8	28	14/16	15/16
#3	S1	ref2	X	15/18	6/9.3	28/31	12/15	13/15
#4	S2	ref1	X	16/17	6/8	30.2/31.2	12/14	13/16
#5	S4	mixture	X	15/16/17/18	6/8/9.3	28/30.2/31/31.2	12/14/15	13/15/16

.	Sample Name	AMEL	D3S1358	TH01	D21S11	D18S51	D10S1248
#1	ref1	X/X	16/17	6/8	30.2/31.2	12/14	13/16
#2	ref2	X/X	15/18	6/9.3	28/31	12/15	13/15
#3	ref3	X/Y	16/18	6/8	28/28	14/16	15/16
#4	Unknown 1	X/Y	16/18	6/7	30/31.2	14/14	13/16

Toolbar (settings)

Version

Comparison results

Match candidates (mixtures)

Deconvolved profiles

Main panel

Importing data from selected case

Restart program

Perform mixture comparison

Create report

Visualization
Deconvolution
Delete profiles
Export

Evidence profiles

Reference profiles

Exact allele-comparisons when importing

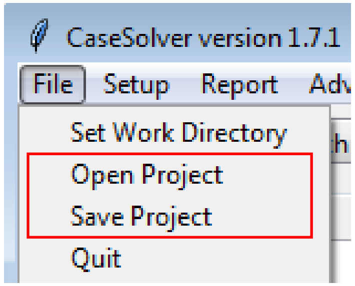
Ref-Ref comparison

Allele information

Unknown profile:
Single source profile which does not match any reference

Random match statistics

Evid-Evid comparison



Software management

Save project → Should be done after following:

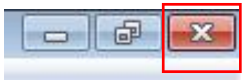
- Time consuming comparisons
- Deconvolution
- Done with analysis

Save project → Save the project file inside the selected case

Open project → Save the project file inside the selected case

Restart program → Always restart program before a new case is imported (remember to save project)

Quit program → Close **R** directly (closing all)



Quit program ← If CaseSolver freezes

Visualization of EPG

Showing replicates together is possible
Peak height info printed in R-Console
Two ways to show the EPGs:

Go to the «Data» panel

Recommended approach

Mark* the profiles you want to visualize (single left mouse click)

Click «Selected profile(s)»

Click «Show data»

Alternative approach

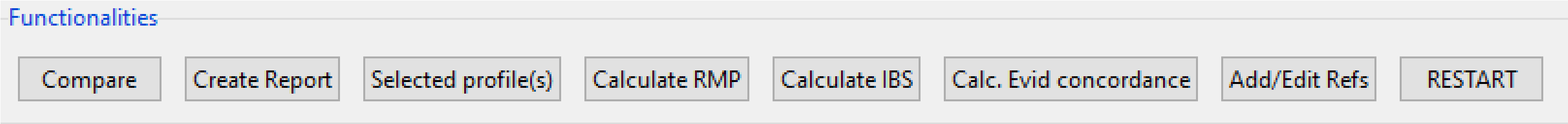
1) One profile: Left double click on a selected evidence
2) Several profiles: Mark* the evidence(s) and hit the ENTER button

Add more reference profiles to EPG:

1) One profile: Left double click on a selected reference
2) Flere profiler: Mark* the reference(s) and hit the ENTER button

*Hold the **ctrl** button for marking several profiles

Functionalities



Compare: Comparing refs to mixtures (allele comparison + LR calculations)

Create Report: Creates a HTML report (set layout under Report in toolbar)

Selected profile(s): Consists of several functions for selected profiles (evid/ref) in data panel:

Show data: Visualize profiles in EPG

Deconvolve: Specify a hypothesis involving the profiles and estimate unknown contributors

Delete from GUI: Remove profiles from tables

Open in EFM: Directly open EFM with selected profiles imported

File export: Create text-files (EFM format) with/without peak heights

Calculate RMP: Random match probability calculated for all references. RMNE calculated for all evidences.

Calculate IBS: Gives number of shared alleles between all reference profiles

Calc Evid concordance: Gives proportion of shared alleles between all evidence profiles

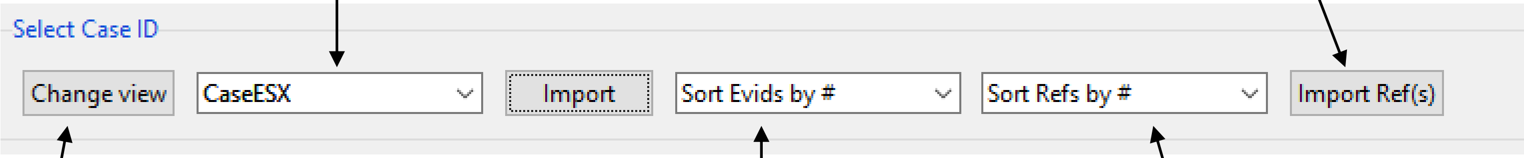
Add/Edit Refs: Add/edit reference profiles

RESTART: Gives fresh restart of program

Functionalities (extra)

It’s possible to edit the case names for:
- Copy-paste
- Scanning

Load additional reference profiles from separate text files (EFM format)



Changes the Table layout of Evid and Ref tables Vertical vs Horizontal

Sort Evid profiles

Sort Ref profiles

Demonstration using CaseSolver with example

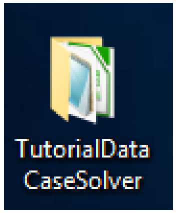
An ESX17 toy example

The tutorial data can be downloaded from
http://euroformix.com/casesolver
(the Tutorial data for CaseSolver zip-file))

Setting up CaseSolver for analyzing tutorial data

See also <u>tutorialvideo</u> for an introduction for how to set up CaseSolver the first time

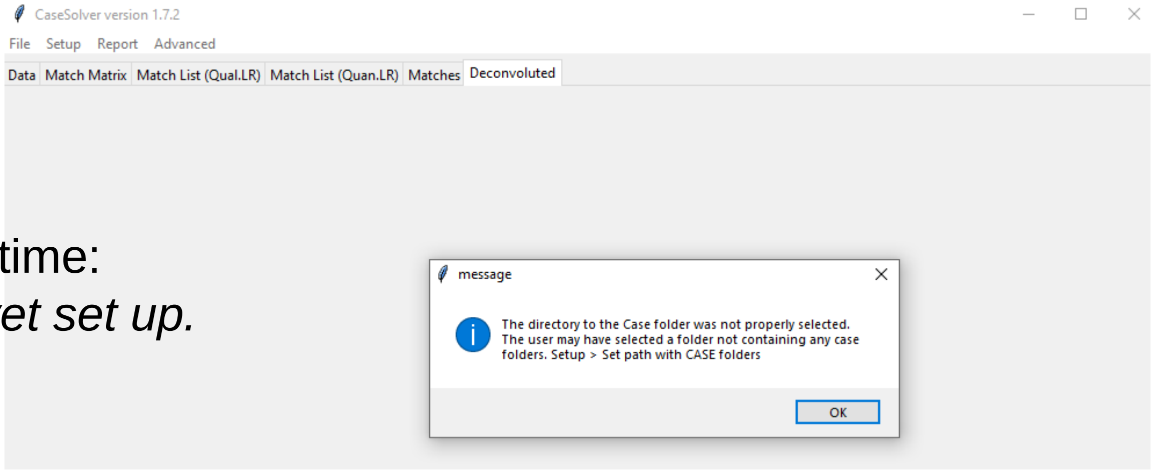
1) Extract folder *TutorialDataCaseSolver* from zip-file to the desktop (or another preferred place).



2) Open R and copy-paste this command to open CaseSolver:

```
library(casesolver);gui()
```

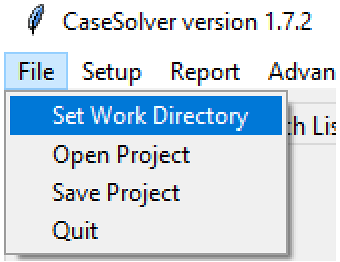
Following message will occur when casesolver is opened first time:
This means the program is not yet set up.



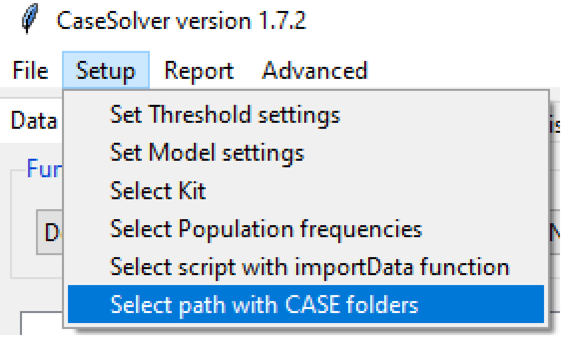
The directory to the Case folder was not properly selected. The user may have selected a folder not containing any case folders. Setup > Set path with CASE folders

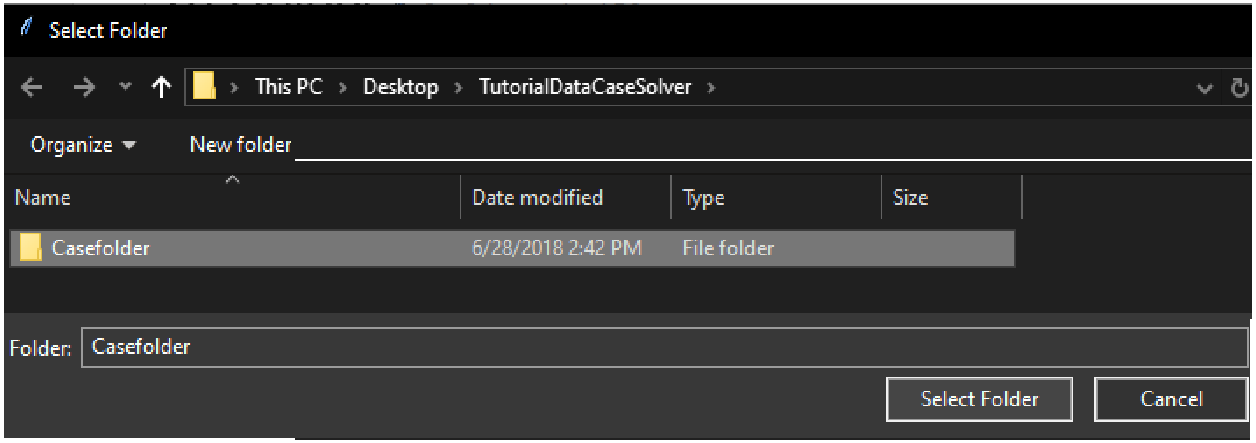
3) Useful to set working directory:
Select the copied folder “TutorialDataCaseSolver” under “File>Set Work Directory” in GUI.

This makes it easier to find the folder and file in the next steps



4) Select path with Case folders

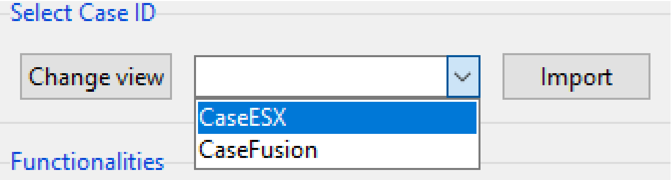




Name	Date modified	Type	Size
Casefolder	6/28/2018 2:42 PM	File folder	

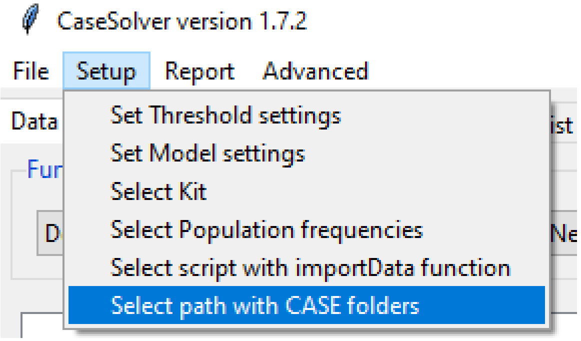
Folder: Casefolder

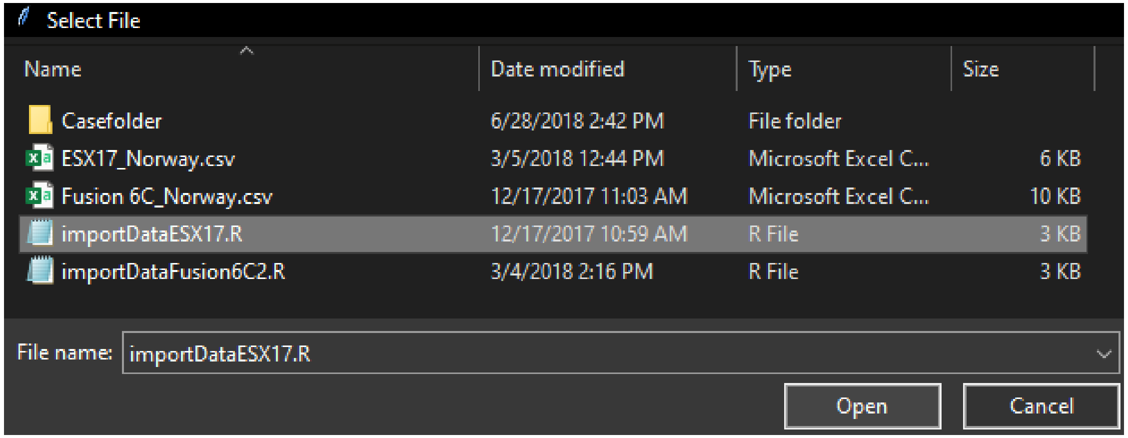
Single click the folder **Casefolder** and click **Select folder**



The two cases should now be found in the dropdown menu
Otherwise the path with CASE folders is wrongly specified

5) Select ImportData function





Name	Date modified	Type	Size
Casefolder	6/28/2018 2:42 PM	File folder	
ESX17_Norway.csv	3/5/2018 12:44 PM	Microsoft Excel C...	6 KB
Fusion 6C_Norway.csv	12/17/2017 11:03 AM	Microsoft Excel C...	10 KB
importDataESX17.R	12/17/2017 10:59 AM	R File	3 KB
importDataFusion6C2.R	3/4/2018 2:16 PM	R File	3 KB

File name: importDataESX17.R

Single click the file **importDataESX17** and click **Open**

This is file is important to tell how data should be imported to CaseSolver For instance what markers that should be considered

Notification

- At this point, **it is possible** to import case data to casesolver

- **However** the usage of casesolver is **limited** to the following:

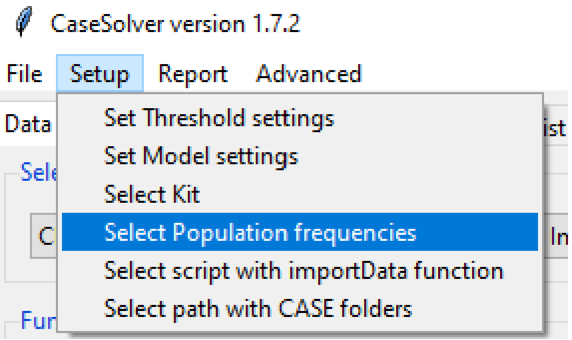
    - LR calculations cannot be carried out
        - Requires **allele frequencies**

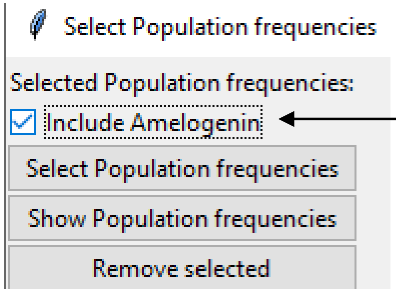
    - Visualization of profiles is limited to barplots (not EPG)
        - **Select kit** for providing EPG plots

    - Deconvolution is not possible
        - Requires **allele frequencies**
        - Degradation model also require **specification of kit**

We encourage users to follow the remaining steps

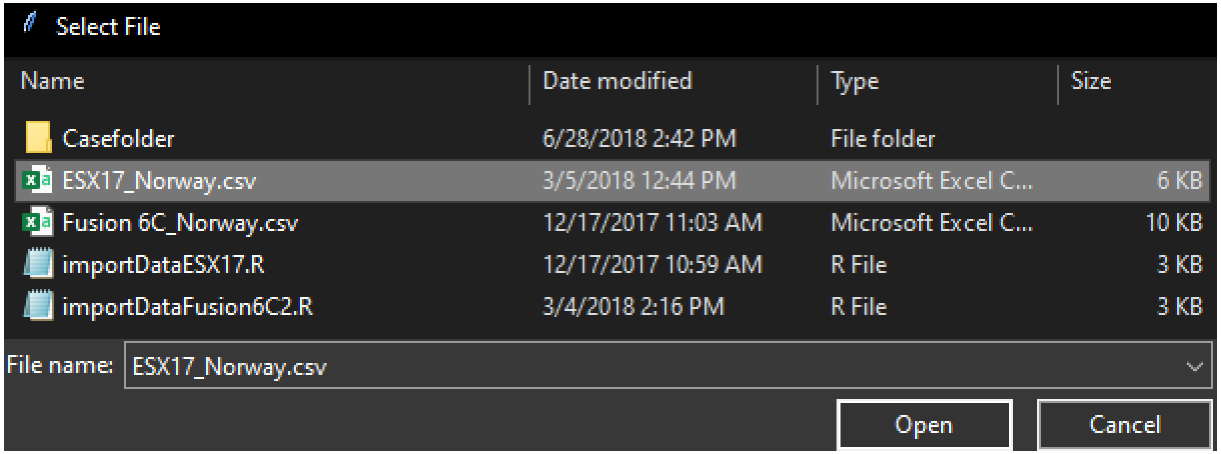
6) Select frequency file





☑ Include Amelogenin

Enables the use of AMEL in LR calculations, deconvolution and contribution visualisation

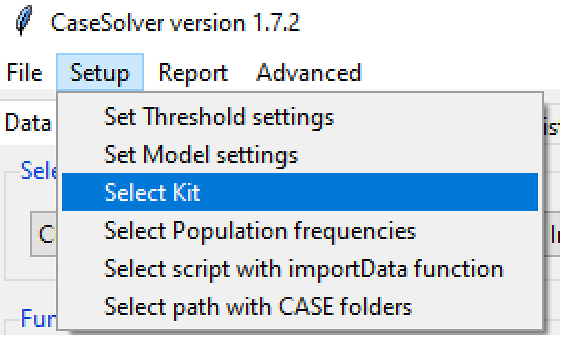


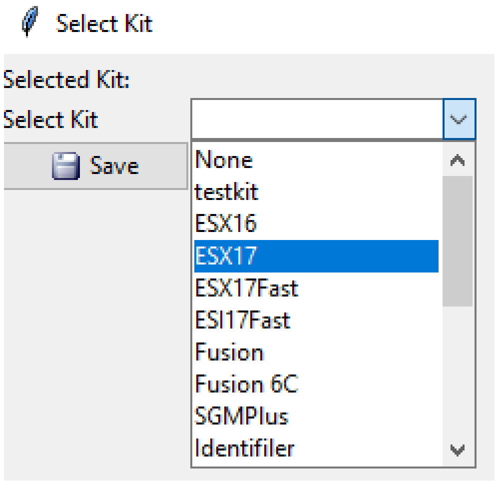
Name	Date modified	Type	Size
Casefolder	6/28/2018 2:42 PM	File folder	
ESX17_Norway.csv	3/5/2018 12:44 PM	Microsoft Excel C...	6 KB
Fusion 6C_Norway.csv	12/17/2017 11:03 AM	Microsoft Excel C...	10 KB
importDataESX17.R	12/17/2017 10:59 AM	R File	3 KB
importDataFusion6C2.R	3/4/2018 2:16 PM	R File	3 KB

File name: ESX17_Norway.csv

1) Tick “Include Amelogenin”
2) Single click “Select Population frequencies”
3) Single click the file **ESX17_Norway** and click **Open**

7) Select Kit

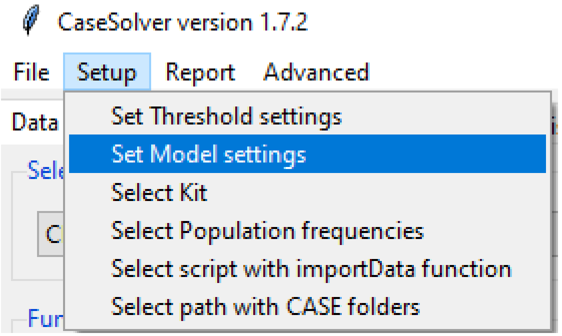


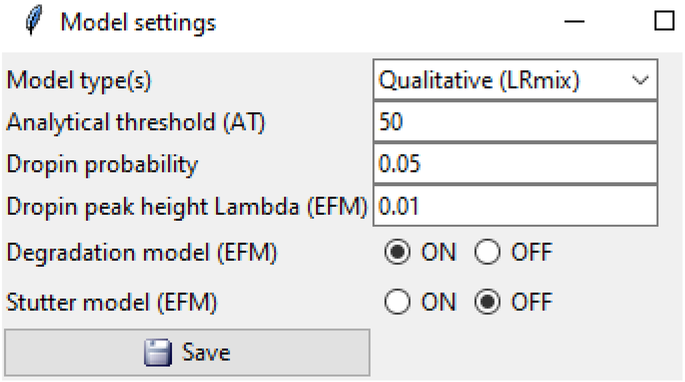


1) Select the **ESX17** kit from the dropdown menu
2) Click **Save**

Notification: Selecting kit is important for:
- Visualizing EPG
- Be able to use degradation model

8) Select Model settings




Model settings

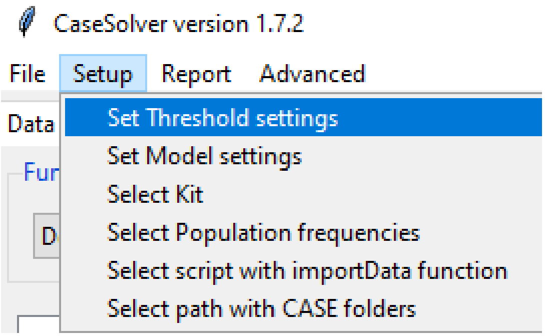
Model type(s)	Qualitative (LRmix)
Analytical threshold (AT)	50
Dropin probability	0.05
Dropin peak height Lambda (EFM)	0.01
Degradation model (EFM)	● ON ○ OFF
Stutter model (EFM)	○ ON ● OFF

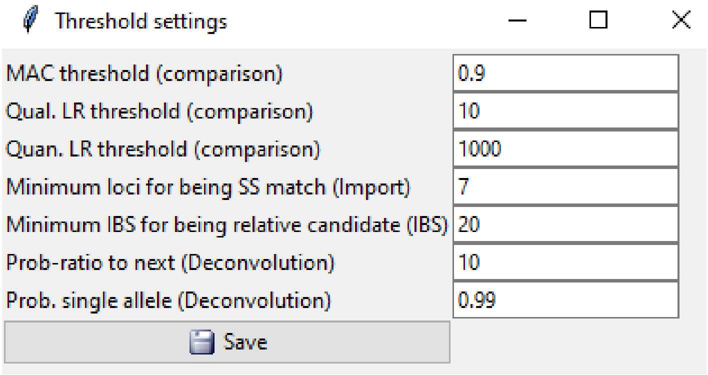
Use settings shown and click **Save**

Notification:

- The **Model settings** only affects LR calculations and deconvolution
- The Quantitative (EuroForMix) model can still be run even if not selected here
- Using Qualitative (LRmix) model is useful as a pre-step in comparison (fast)

9) Select Threshold settings





MAC threshold (comparison)	0.9
Qual. LR threshold (comparison)	10
Quan. LR threshold (comparison)	1000
Minimum loci for being SS match (Import)	7
Minimum IBS for being relative candidate (IBS)	20
Prob-ratio to next (Deconvolution)	10
Prob. single allele (Deconvolution)	0.99

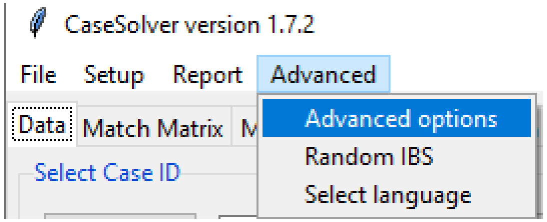
Use settings shown and click **Save**

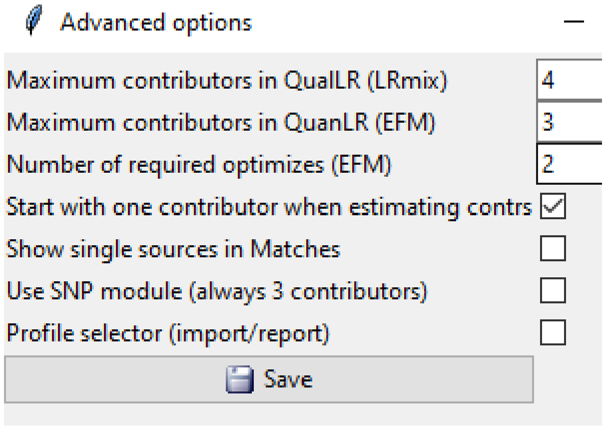
Notification:

The **Threshold settings** are the **rules** which CaseSolver uses for following:

- Providing list of candidate matches (Mixture comparison)
- Providing estimated unknown profile (Deconvolution)
- Sufficient number of matching markers in partial single source profiles (Import)
- The required number of alleles overlapping two references (IBS)

10) Set Advanced options (optional)





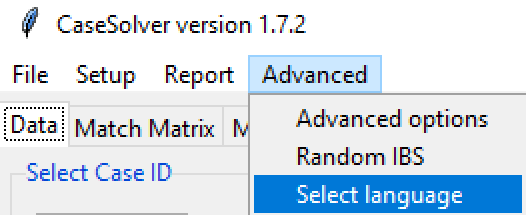
Maximum contributors in QualLR (LRmix) 4
Maximum contributors in QuanLR (EFM) 3
Number of required optimizes (EFM) 2
Start with one contributor when estimating contrs ☑
Show single sources in Matches ☐
Use SNP module (always 3 contributors) ☐
Profile selector (import/report) ☐

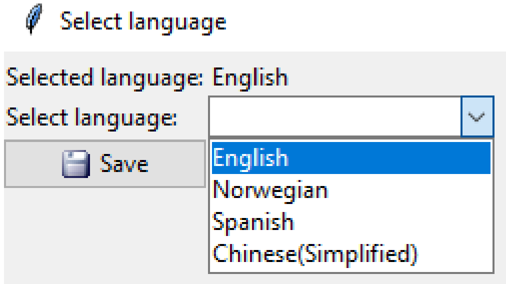
Use following settings and click **Save**

Notifications:

- The **maximum** (number of) **contributor** specification is only used for the *automatic* mixture comparisons
- Tick **Show single sources in Matches** to show all samples in **Matches** list
- Tick **Profile selector** to manually select which profiles from a case that should be imported or reported

11) Set Language (extra)

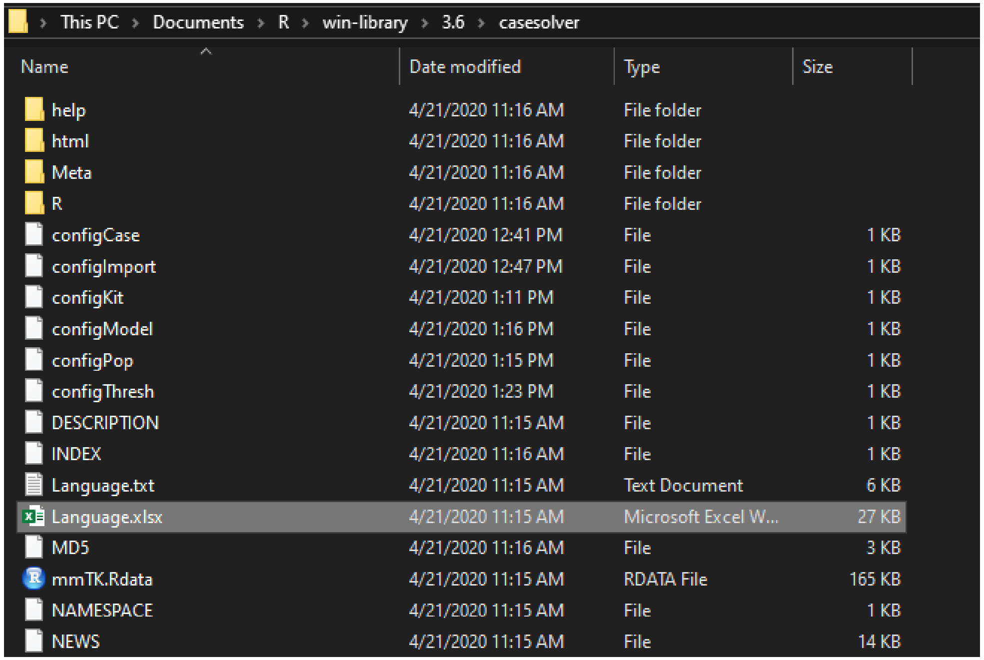




Selected language: English

Imports text phrases from the **Language.xlsx** file which is in the casesolver R-installation folder

Modify the **Language.xlsx** file to **add** new language or edit existing

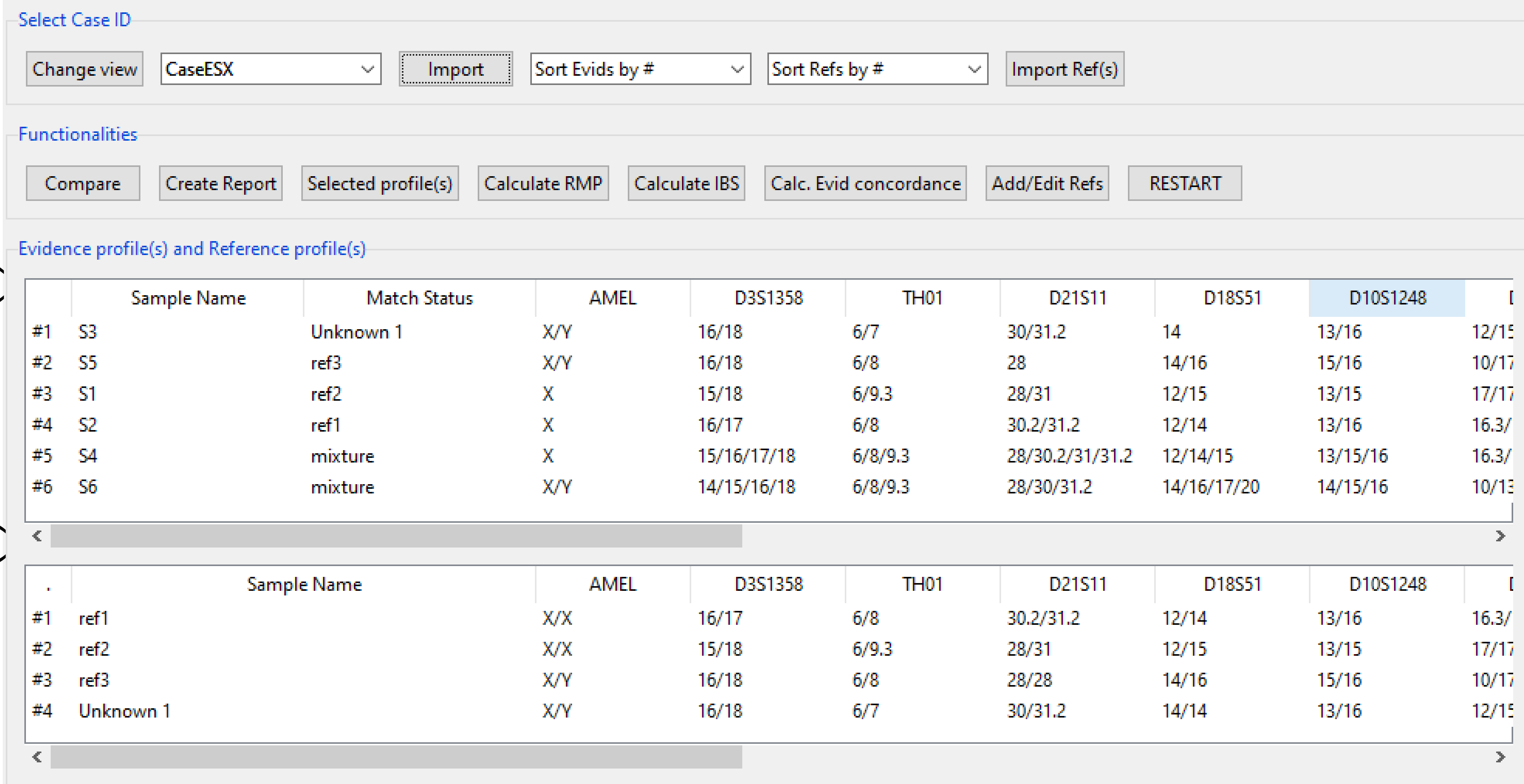


This PC › Documents › R › win-library › 3.6 › casesolver

Name	Date modified	Type	Size
help	4/21/2020 11:16 AM	File folder	
html	4/21/2020 11:16 AM	File folder	
Meta	4/21/2020 11:16 AM	File folder	
R	4/21/2020 11:16 AM	File folder	
configCase	4/21/2020 12:41 PM	File	1 KB
configImport	4/21/2020 12:47 PM	File	1 KB
configKit	4/21/2020 1:11 PM	File	1 KB
configModel	4/21/2020 1:16 PM	File	1 KB
configPop	4/21/2020 1:15 PM	File	1 KB
configThresh	4/21/2020 1:23 PM	File	1 KB
DESCRIPTION	4/21/2020 11:15 AM	File	1 KB
INDEX	4/21/2020 11:16 AM	File	1 KB
Language.txt	4/21/2020 11:15 AM	Text Document	6 KB
Language.xlsx	4/21/2020 11:15 AM	Microsoft Excel W...	27 KB
MD5	4/21/2020 11:16 AM	File	3 KB
mmTK.Rdata	4/21/2020 11:15 AM	RDATA File	165 KB
NAMESPACE	4/21/2020 11:15 AM	File	1 KB
NEWS	4/21/2020 11:15 AM	File	14 KB

12) Importing data from a case

Select case **CaseESX** from drop-down and click **import**



CaseESX

Evidence(s)

	Sample Name	Match Status	AMEL	D3S1358	TH01	D21S11	D18S51	D10S1248	D
#1	S3	Unknown 1	X/Y	16/18	6/7	30/31.2	14	13/16	12/1
#2	S5	ref3	X/Y	16/18	6/8	28	14/16	15/16	10/1
#3	S1	ref2	X	15/18	6/9.3	28/31	12/15	13/15	17/1
#4	S2	ref1	X	16/17	6/8	30.2/31.2	12/14	13/16	16.3/
#5	S4	mixture	X	15/16/17/18	6/8/9.3	28/30.2/31/31.2	12/14/15	13/15/16	16.3/
#6	S6	mixture	X/Y	14/15/16/18	6/8/9.3	28/30/31.2	14/16/17/20	14/15/16	10/1

Reference(s)

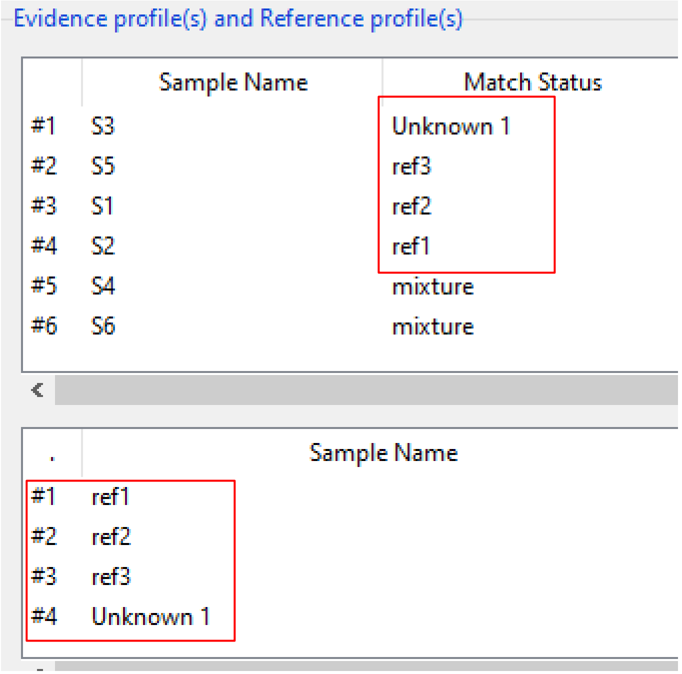
.	Sample Name	AMEL	D3S1358	TH01	D21S11	D18S51	D10S1248	D
#1	ref1	X/X	16/17	6/8	30.2/31.2	12/14	13/16	16.3/
#2	ref2	X/X	15/18	6/9.3	28/31	12/15	13/15	17/1
#3	ref3	X/Y	16/18	6/8	28/28	14/16	15/16	10/1
#4	Unknown 1	X/Y	16/18	6/7	30/31.2	14/14	13/16	12/1

13) Comparison

Automatic comparison by case-import

<u>When data are **imported:**</u>
- Evidence profiles are categorized as either **single source** or **mixture**

<u>Automatic exact allele comparison:</u>
- Every reference profile (ref) is compared to each **single source** profile



Evidence profile(s) and Reference profile(s)

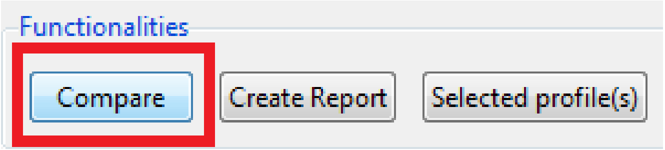
	Sample Name	Match Status
#1	S3	Unknown 1
#2	S5	ref3
#3	S1	ref2
#4	S2	ref1
#5	S4	mixture
#6	S6	mixture

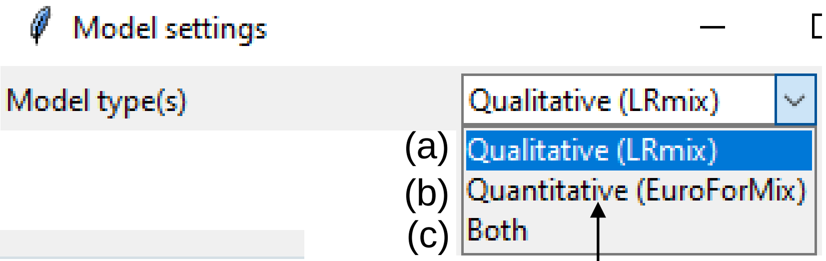
.	Sample Name
#1	ref1
#2	ref2
#3	ref3
#4	Unknown 1

<u>**Matching references** are labeled under MatchStatus:</u>
- Supports partial/missing loci of reference
- Assigned as an **Unknown** if none are matching
  - Unknowns are added to Reference list
  - Supports any number of unknowns (tagged with ID)

Mixture comparison

Click «Compare» under the «Data» panel to compare references against mixtures.



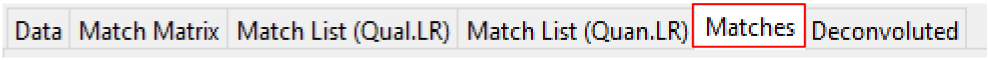


Model type(s) Qualitative (LRmix)

Choose mixture comparison strategy

Performs 2 steps to retrieve candidate matches

These will be shown the Matches table



Step 1: Simple allele comparison (results shown as Match matrix)

Step 2: LR based comparison with 3 possible strategies:

– (a) Using qualitative model only (LRmix)

– (b) Using quantitative model only (EuroForMix)

– (c) Both: First (a) and then (b)

(a) is very fast and useful to reduce false candidates

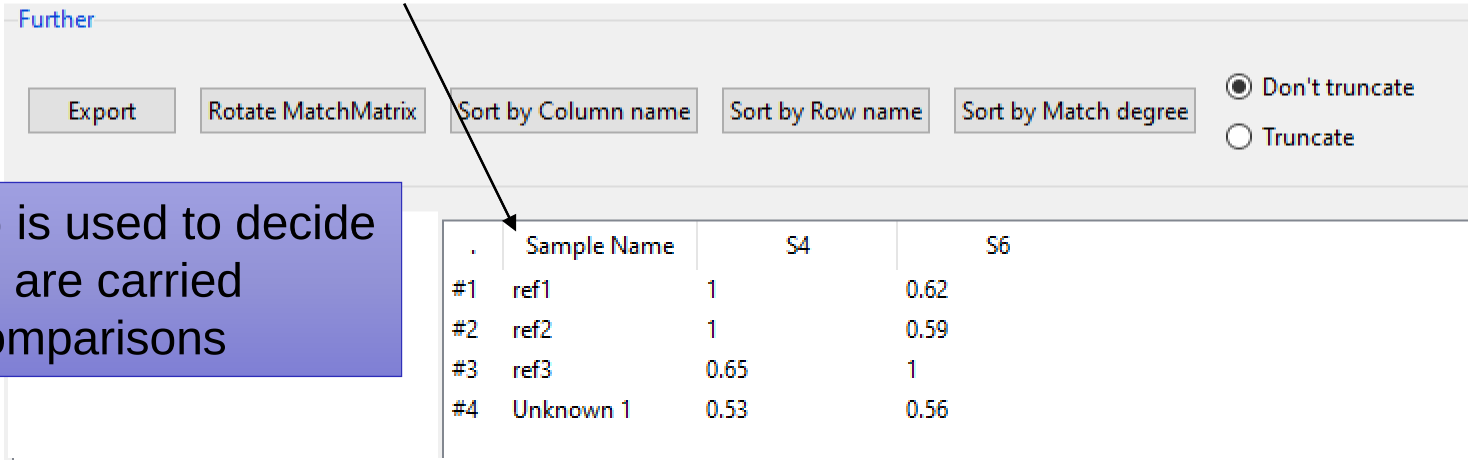
Note: From version 1.7 Evid profiles with *empty** **MatchStatus** will also be compared (even if they are single source profile)

*empty if reference profile (for instance unknown) is removed from GUI

Comparison ***step 1)*** MatchMatrix

Obtain a table with simple allele comparisons

The table shows the **proportion** of reference alleles which overlaps with the evidence profile alleles



● Don't truncate
○ Truncate

.	Sample Name	S4	S6
#1	ref1	1	0.62
#2	ref2	1	0.59
#3	ref3	0.65	1
#4	Unknown 1	0.53	0.56

A threshold (MAC*) is used to decide which comparisons are carried further to **Step 2** comparisons

Click«Truncate» to only see comparisons which satisfy this MAC* threshold.

Value «1» means 100% concordance

*MAC = matching allele counting

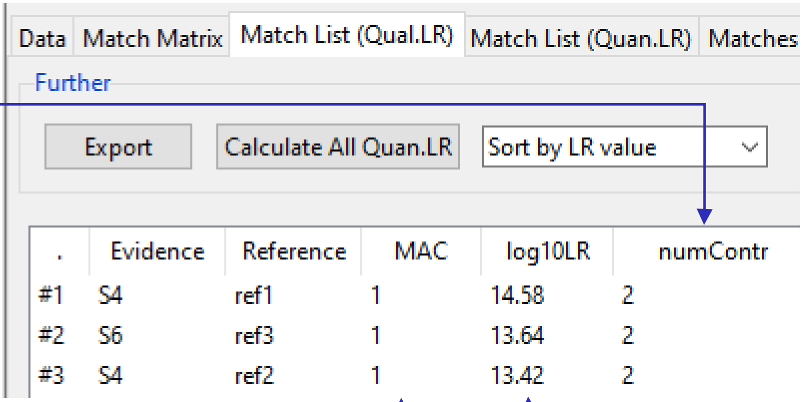
Comparison ***step 2*** (a/c) - qualitative

The program first estimates the number of contributors for all evidence profiles

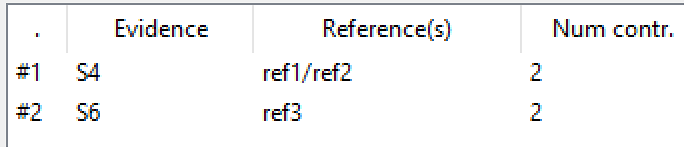
Then it calculates the LR for all comparisons satisfying the *MAC threshold*

Look at the LR values in the table under «Match List (Qual.LR)»

Comparisons with LR above *LR threshold* gives *candidate matches* in the «**Matches**» panel



.	Evidence	Reference	MAC	log10LR	numContr
#1	S4	ref1	1	14.58	2
#2	S6	ref3	1	13.64	2
#3	S4	ref2	1	13.42	2



.	Evidence	Reference(s)	Num contr.
#1	S4	ref1/ref2	2
#2	S6	ref3	2

The log10LR value is important

A large positive number indicates contribution

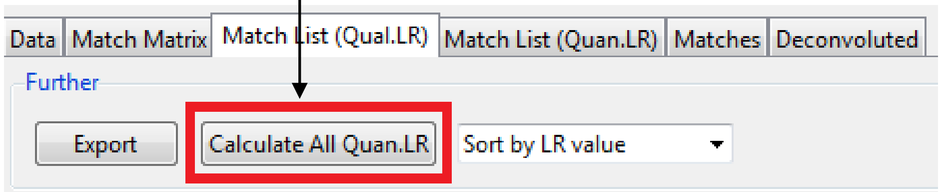
A large negative number indicates non-contributor

A number close to zero indicates neither (neutral)

Comparison ***step 2*** (b/c) - quantitative

The qualitative model does not take into account peak height information

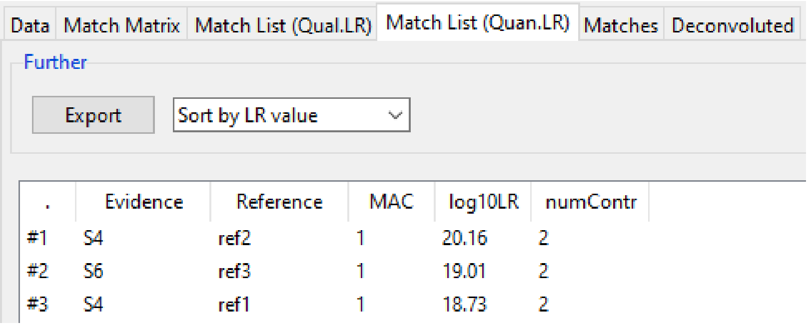
Calculate quan. based LR at any time:
useful as a post-step after strategy (a)



(c) Calculates the quan. based LR for all comparisons satisfying the *qual LR threshold*

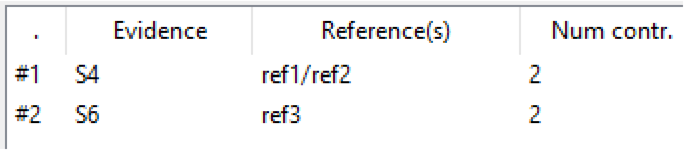
Calculations using EuroForMix (utilizes peak height information)

Look at the LR values in the «Match List (Quan.LR)» table



.	Evidence	Reference	MAC	log10LR	numContr
#1	S4	ref2	1	20.16	2
#2	S6	ref3	1	19.01	2
#3	S4	ref1	1	18.73	2

Comparing with quan. LR threshold to give (potentially) updated list of *candidate matches* in «**Matches**» panel

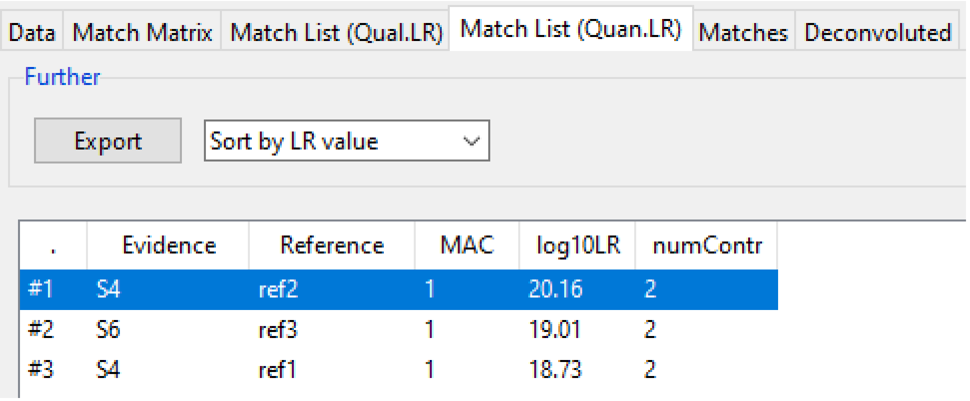


.	Evidence	Reference(s)	Num contr.
#1	S4	ref1/ref2	2
#2	S6	ref3	2

Also: Perform tailored quan. LR calculations by double clicking candidates in the **«Match List (Qual.LR)» table**: «Matches» list will be updated accordingly

Check peak height contributions of compared reference

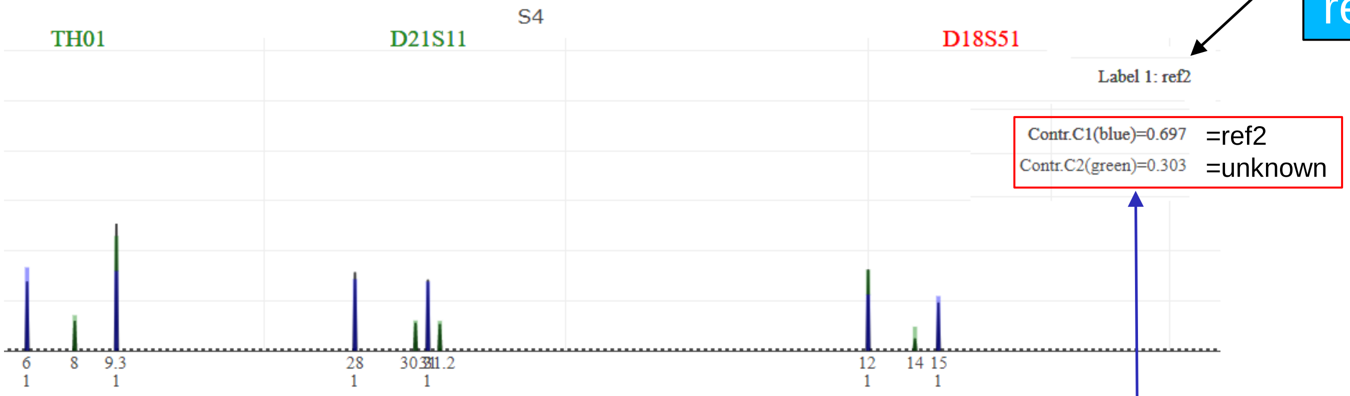
Left double-click a row in the table under **«MatchList (quanLR)»**



.	Evidence	Reference	MAC	log10LR	numContr
#1	S4	ref2	1	20.16	2
#2	S6	ref3	1	19.01	2
#3	S4	ref1	1	18.73	2

Inspecting the peak heights with fitted model (mixture proportions etc)

Conditioning on the compared reference only



=ref2

=unknown

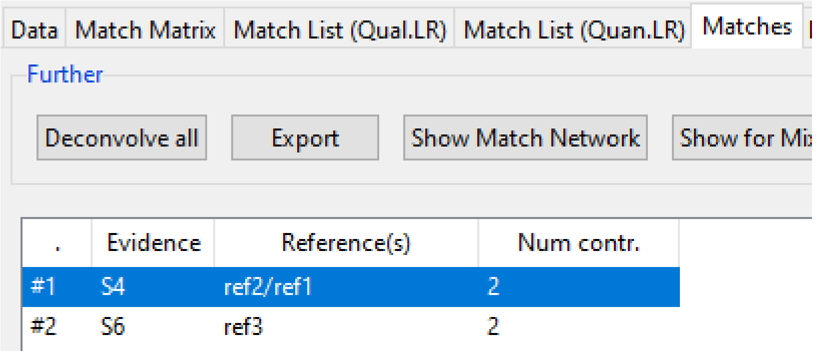
Check whether the bars fits the heights

Estimated mixture proportions

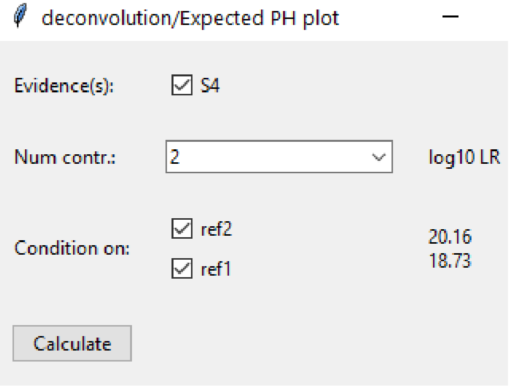
Check peak height contributions with (several) contributors simultaneously

Left double-click a row in the table under «**Matches**»

Specify the conditional profiles and the number of contributors



.	Evidence	Reference(s)	Num contr.
#1	S4	ref2/ref1	2
#2	S6	ref3	2



Evidence(s): ☑ S4

Num contr.: 2

log10 LR

Condition on: ☑ ref2 20.16
☑ ref1 18.73

Conditioning on several references simultaneously

Check whether the candidates simultaneously fit the evidence profile

Contr.C1(blue)=0.694 =ref2

Contr.C2(green)=0.306 =ref1

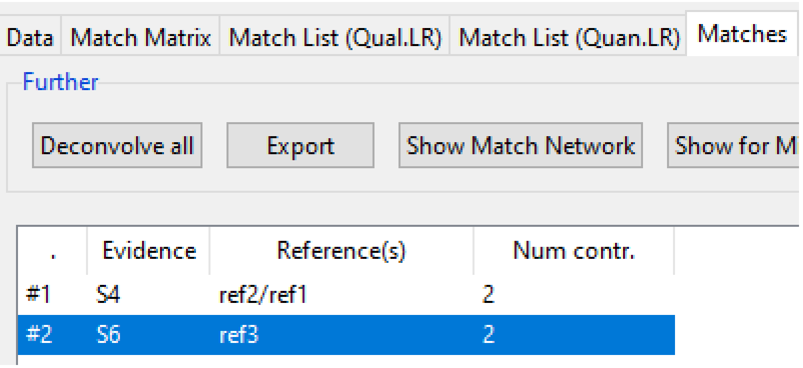
Check whether the bars fits the heights

Estimated mixture proportions

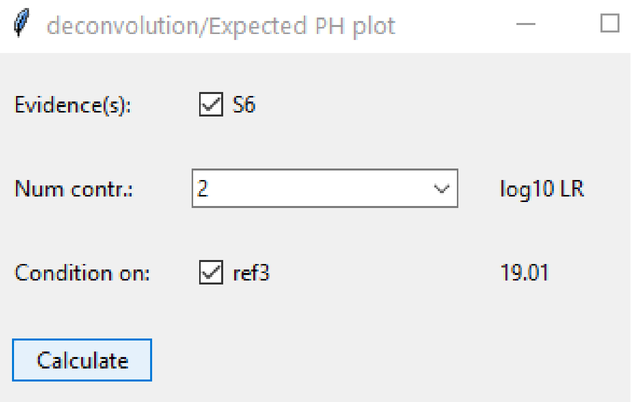
14) Deconvolution

Left double-click a row in the table under «**Matches**»

Specify the conditional profiles and the number of contributors



.	Evidence	Reference(s)	Num contr.
#1	S4	ref2/ref1	2
#2	S6	ref3	2

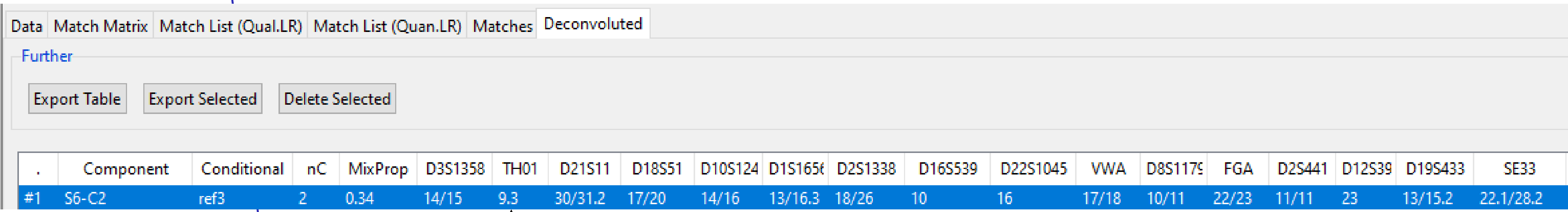


Evidence(s): ☑ S6

Num contr.: 2

log10 LR 19.01

Condition on: ☑ ref3



.	Component	Conditional	nC	MixProp	D3S1358	TH01	D21S11	D18S51	D10S124	D1S165	D2S1338	D16S539	D22S1045	VWA	D8S117	FGA	D2S441	D12S39	D19S433	SE33
#1	S6-C2	ref3	2	0.34	14/15	9.3	30/31.2	17/20	14/16	13/16.3	18/26	10	16	17/18	10/11	22/23	11/11	23	13/15.2	22.1/28.2

Check estimated genotypes of unknown(s) under «**Deconvoluted**» panel

Left double-click a candidate to extract it to the «Reference profile» table in the «**Data**» panel

Supports deconvolution of **one allele**
CaseSolver typically treat the genotype as a homozygous

Useful regarding deconvolution

Perform **compare** again when new reference profiles are added after deconvolution

Deconvolutions can be carried out directly under the *«Selected profile(s)»* functionality in the **«Data»** panel

Deconvolution can be used to estimate the genotype of missing markers of reference profiles

Deconvolution can utilize replicates

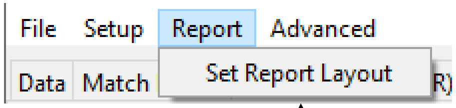
15) Create report

Can be done at any time using CaseSolver, however comparisons must first be carried out to show match results of mixtures

Click the «Create report» button under «Data» panel to produce HTML report

Choose name of report

A preview of report is automatically shown in default browser



Choose which elements the report should include