Manual for EuroForMix v1

Author: Øyvind Bleka <Oyvind.Bleka.at.fhi.no>
Date: 12-04-2014

(A) Installation and running program:

1) Run R (>=3.0.1) in Windows, Linux or MAC (http://cran.r-project.org/).
2) Required packages to run GUI:
    a. gWidgetstcltk (depends on digest,tcltk)
    b. gWidgets
3) Other required packages:
    a. cubature
        i. Required for multivariate integration (Integrated LR).
    b. forensim
        i. Required for qualitative Weight-of-Evidence.
4) Installation and run gammadnamix:
    a. install.packages("gammadnamix", repos="http://R-Forge.R-project.org")
    b. library(gammadnamix)
    c. euroformix()

(B) GUI

Sections:

0- Toolbar
1- Importing data
2- Model specification
3- MLE fit: (‘Continuous LR (Maximum Likelihood based)’)
4- Deconvolution (Deconvolution based on the continuous model)
5- Database Search (Database search based on the continuous and qualitative model)
6- Qual.LR (Qualitative model)
7- Generate data (Generation from the continuous model)

0. Toolbar

- File
    - **Set directory**: The user may select the working directory of the R-program.
    - **Open project**: The user may open an earlier project which is saved in a file on the form “projectname.Rdata”.
    - **Save project**: The user may save the existing project into a file with name “projectname”.
        - Extension .Rdata is added automatically to project name.
        - All data imported to the program and resulting calculations are stored into a single project-file which may be open at any time in the program.
        - Saving a project makes:
            - Big reference databases are stored efficiently (the required space for the database is drastically reduced).
            - Time-consuming calculations are restored instantly (only required to be calculated ones).
    - **Quit project**: When pushed, the user get question about saving project before terminating the GUI.
- Frequencies
    - **Set size of frequency database**: User may specify number of samples ‘N’ used to create the population frequencies.
        - When new alleles from imported files are found, these are assigned as freq0.
            - If N=0 (this is default), freq0 is equal minimum observed frequency.
            - If N>0, freq0=‘5/(2N)’.
        - New alleles are updated to the population frequencies when:
            - When a reference database is imported.
            - When interpretations are done.
                - Deconvolution, Weight-of-Evidence and ‘Database search’
            - Frequencies are normalized for each of these two cases.
                - **WARNING**: Normalizing may be done twice if new alleles (not seen in population frequency table or reference database) are observed in the evidence/reference profile.
    - **Set number of wildcards in false positive match**: The user may specify number of wildcards in the random match probability statistics, which are applied when the user has imported and selected an evidence stain together with the population frequencies.

- Optimization

    - **Set number of random startpoints**: The user may set required number of independent random startpoints in the optimizer to ensure that the global maximum is attained for the Maximum Likelihood Estimator (MLE). Default is 3.

    - **Set variance of randomizer**: The user may set the variance parameter used for the random generation of startpoints used in optimizer. Default is 10**.**

- MCMC (Markov Chain Monte Carlo)

    - **Set number of samples**: The user may set the number of samples drawn from the posterior distribution of the parameters. Default is 10000.

    - **Set variance of randomizer**: The user may set the variance parameter scalar used in the 'Markov Chain Monte Carlo (MCMC) random walk Metropolis'. See vignette for details. Default is 10.
        - Note that this value should be tweaked such that acceptance rate of sampler are around 0.2 (to ensure global exploration in the parameter space).

- Integration

    - **Set relative error requirement**: The user may set the required estimated relative error used in the integration function adaptIntegrate {cubature}. See vignette for details. Default is 0.005.

    - **Set maximum of mu-parameter**: The user may set upper limit of mu-parameter (amount of DNA). See vignette for details. Default is 20000.

    - **Set maximum of sigma-parameter**: The user may set upper limit of sigma-parameter (coefficient of variation). See vignette for details. Default is 1.

    - **Set maximum of stutter ratio-parameter**: The user may set upper limit of the stutter ratio parameter (xi). Default is 1.

- Deconvolution

    - **Set required summed probability**: The user may set required summed posterior genotype-probability which the deconvoluted list is ensured to contain. Default is 0.9999.

    - **Set max listsize:** The user may set maximum number of elements in the deconvoluted list. Default is 1000.
        - The greater max listsize, the more time-consuming (and memory consuming) the search-algorithm behind will be.

- Database search

  - **Set maximum view-elements**: The user may set maximum number of individuals to show from the reference-database. Default is 10000.
    - The greater ‘value’, the more time-consuming will it become to show table on screen.
    - Note that the result table from the database search shows only the top ‘value’-ranked elements.

  - **Set drop-in probability for qualitative model**: When searching database with continuous LR model, the qualitative LR model is also considered with a specific drop-in probability parameter given here (default is 0.05).

- Qual LR

  - **Set upper range for sensitivity**: The user may specify the maximum allele dropout-probability in the sensitivity plot (for a qualitative model). Default is 0.6.

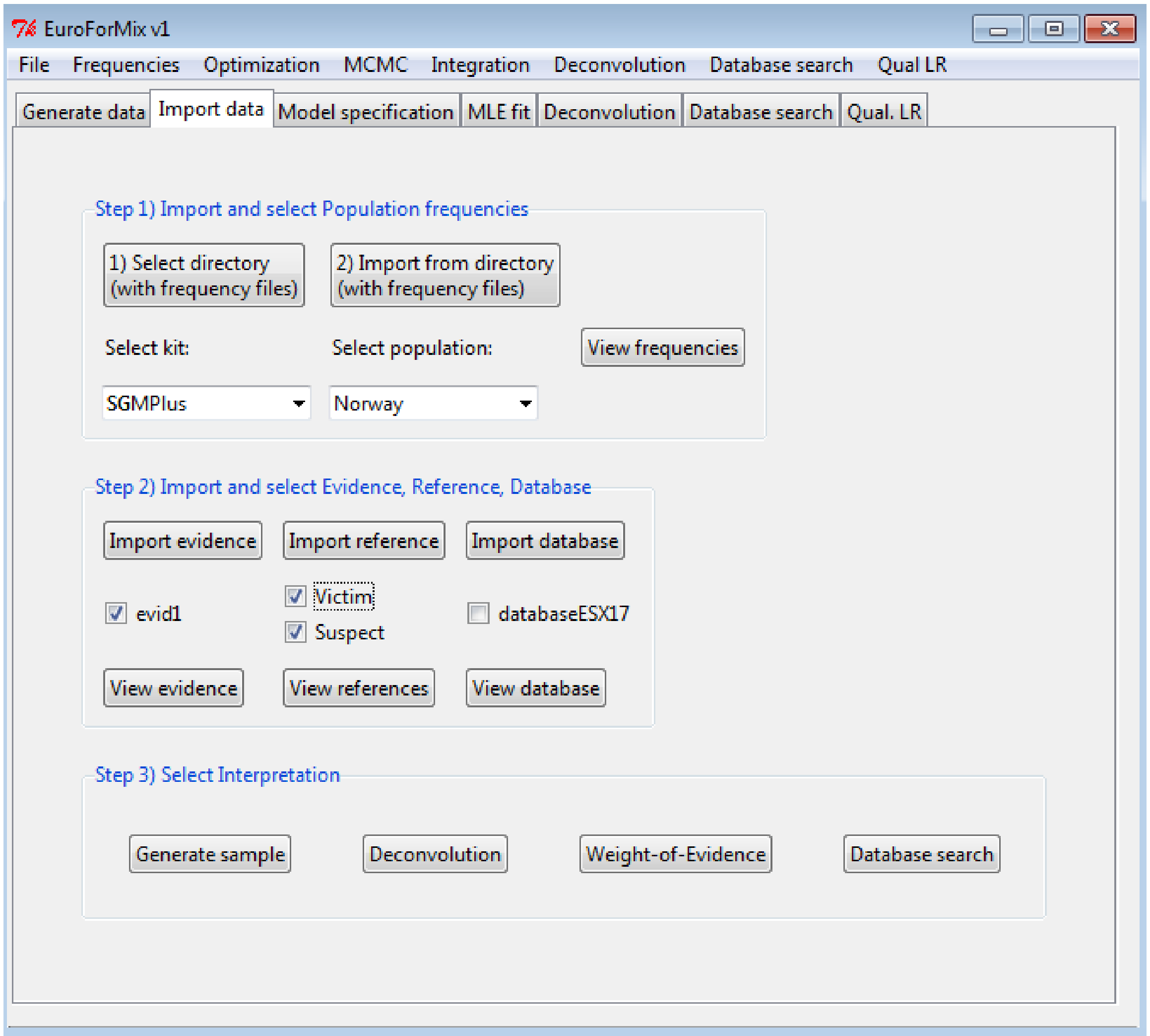
  - **Set nticks for sensitivity**: The user may specify number of grids of the allele dropout-probability in the sensitivity plot (for a qualitative model). Default is 31.

  - **Set required samples in dropout distr.**: The user may specify number of required allele drop-out probability samples used to estimate the quantiles or meadian for the distribution of the ‘*allele drop-out probability given number of observed alleles*’.

  - **Set significance level in dropout distr.**: The user may specify the significance level in the conservative LR calculation (i.e. the quantile for the distribution of the ‘*allele drop-out probability given number of observed alleles*’). Default is 0.05.

  - **Set number of tippets**: The user may specify number of random man tippet samples. Default is 1e6.

1. Importing data



Select kit: SGMPlus

Select population: Norway

☑ evid1

☑ Victim

☑ Suspect

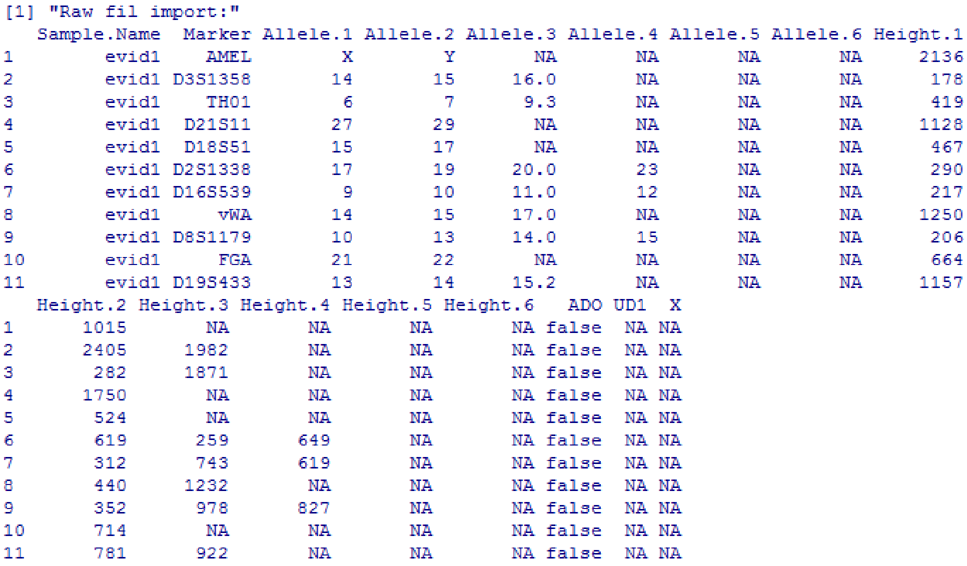
□ databaseESX17

Figure 1: The figure shows the Import data GUI page where the user can import population frequencies, evidence stains, reference profiles and reference databases.

DATA IMPORT:

- **Common** for all files:
    - The extension (denotes file-type) of the file names does not matter. It may also have no extension at all.
    - All imported files must be either comma, semi-colon or tab-separated (‘,’,’;’,’\t’).
    - Required/optional headers (all are capital invariant):
        - “**sample**” is required header for sample(s) name(s).
            - The sample names are NOT capital invariant.
            - If more than one header name contains “**sample**”, it will select the header name which in addition contains “**name**” in the same string.

        - “**marker**” is required header for marker name(s).
            - Marker names are capital invariant.
            - If no header is found, the header containing “**loc**” will be used if found.
        - “**allele**” is required header(s) for allele-information.
            - This may be a vector (“alleleX1”,…,”allelleX10”) of any length denoting allele(s) to a given marker for a given sample. Here X1,…,X10 can be anything.
        - “**height**” optional header(s) for peak height-information.
            - This may be a vector (“heightX1”,…,”heightX10”) of any length denoting peak height to the corresponding allele(s) in “allele”. Here X1,…,X10 can be anything.
    - Note:
        - The imported data will use upper-letter of marker-names found in the file.
        - All imports are printed out in the terminal (see figure 2). From this, the user may check that the data are imported correctly.



```
[1] "Raw fil import:"
   Sample.Name  Marker Allele.1 Allele.2 Allele.3 Allele.4 Allele.5 Allele.6 Height.1
1        evid1    AMEL        X        Y       NA       NA       NA       NA     2136
2        evid1 D3S1358       14       15     16.0       NA       NA       NA      178
3        evid1    TH01        6        7      9.3       NA       NA       NA      419
4        evid1  D21S11       27       29       NA       NA       NA       NA     1128
5        evid1  D18S51       15       17       NA       NA       NA       NA      467
6        evid1 D2S1338       17       19     20.0       23       NA       NA      290
7        evid1 D16S539        9       10     11.0       12       NA       NA      217
8        evid1     vWA       14       15     17.0       NA       NA       NA     1250
9        evid1 D8S1179       10       13     14.0       15       NA       NA      206
10       evid1     FGA       21       22       NA       NA       NA       NA      664
11       evid1 D19S433       13       14     15.2       NA       NA       NA     1157
   Height.2 Height.3 Height.4 Height.5 Height.6   ADO UD1  X
1      1015       NA       NA       NA       NA false  NA NA
2      2405     1982       NA       NA       NA false  NA NA
3       282     1871       NA       NA       NA false  NA NA
4      1750       NA       NA       NA       NA false  NA NA
5       524       NA       NA       NA       NA false  NA NA
6       619      259      649       NA       NA false  NA NA
7       312      743      619       NA       NA false  NA NA
8       440     1232       NA       NA       NA false  NA NA
9       352      978      827       NA       NA false  NA NA
10      714       NA       NA       NA       NA false  NA NA
11      781      922       NA       NA       NA false  NA NA
```

Figure 2: The figure shows the table format in the importing evidence stain file.

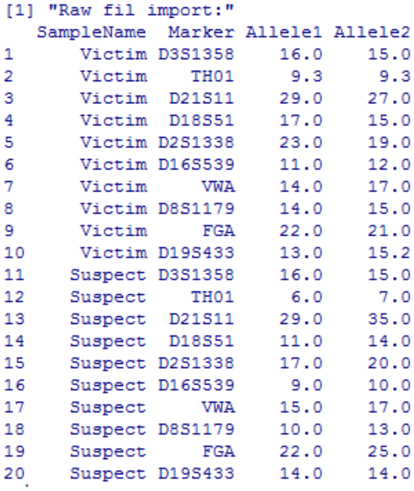
- **Import population frequencies**:

    - Requires an own folder (population-folder) with **only** frequency-files.
    - File-format:
        - Filename:
            - The name of the filenames **needs** to be on the form “kit_population.ext”, where ext can be any extensions (or be missing as well).
            - kit=”kit-name” and population=”population name”
            - The kit-name must be consistent with the short-name of the kit instrument. See ?plotEPG for more details.
        - File:
            - First column needs to be allele-information (header-name may be anything).

      - Other columns are frequency-information (header-name denotes the locus name (loci names are converted to capital letters)).
  - To import frequencies:
    - Push “**1) Select directory**” button to select the population-folder with the population frequency files.
    - Push “**2) Import from directory**” button to import the population frequency files from the selected folder.
      - It is possible to **add new files** into the selected population-folder **at any time** and push the button once again to include new information to the dropdown-list.
  - Selection of kit and population:
    - After importing the frequency-files (after pushed (**2**)), the user may select wanted kit and population from the two drop down lists at any time* (*not after a reference-database file has been imported).
      - This can be useful to see the EPG layout for different selected kits.

- **Import Evidence/Reference** sample (see figure 2 and figure 3):

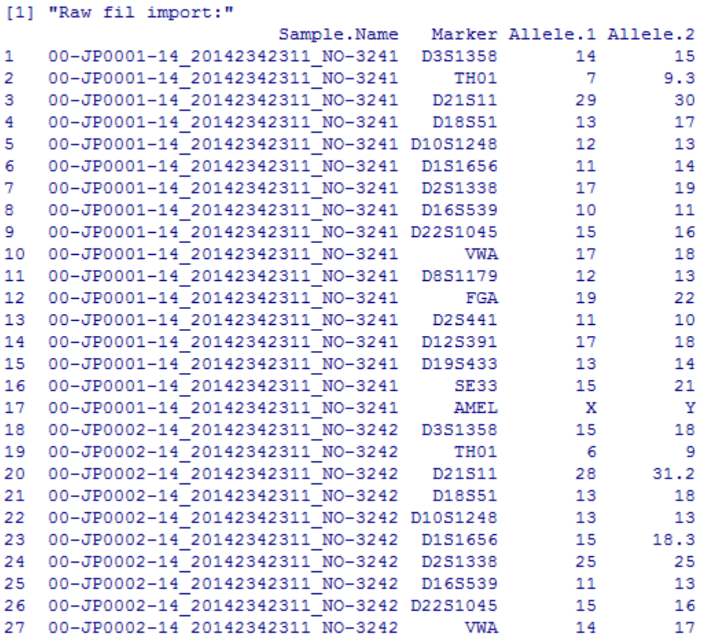
  - **Multiple** evidence or reference profiles are **allowed** in each file.
  - In evidence files:
    - “height” header is required for analysis Deconvolution, Weight-of-Evidence (continuous model) and ‘Database search’. For ‘Qualitative LR’ this is not required.
  - In reference files:
    - “height” header is optional but will not be used further in any analysis.
  - Note:
    - The import function will not check:
      - That the length of allele and heights are equal long for a given locus.
    - Loci without any allele-information (i.e. empty or dropped out), are **NOT** imported.



```
[1] "Raw fil import:"
   SampleName  Marker Allele1 Allele2
1      Victim D3S1358    16.0    15.0
2      Victim    TH01     9.3     9.3
3      Victim  D21S11    29.0    27.0
4      Victim  D18S51    17.0    15.0
5      Victim D2S1338    23.0    19.0
6      Victim D16S539    11.0    12.0
7      Victim     VWA    14.0    17.0
8      Victim D8S1179    14.0    15.0
9      Victim     FGA    22.0    21.0
10     Victim D19S433    13.0    15.2
11    Suspect D3S1358    16.0    15.0
12    Suspect    TH01     6.0     7.0
13    Suspect  D21S11    29.0    35.0
14    Suspect  D18S51    11.0    14.0
15    Suspect D2S1338    17.0    20.0
16    Suspect D16S539     9.0    10.0
17    Suspect     VWA    15.0    17.0
18    Suspect D8S1179    10.0    13.0
19    Suspect     FGA    22.0    25.0
20    Suspect D19S433    14.0    14.0
```

Figure 3: The figure shows the table format in the importing reference file.

- **Import Reference Database (**see figure 4):
    - Exactly same format as reference files.
    - Multiple database file may be imported (**must** be done one-at-the-time).
    - **Requires** that population frequencies are imported and selected.
        - **WARNING**: Population frequencies may not be changed again after database importing!
    - Note:
        - The ranking of databases are done over all selected databases.
        - Same samples within a database needs to be in same block but markers within sample can be different orders.
        - Some samples **may** have more/less markers than others (e.g. SGMplus profiles contra ESX17).
            - **Missing markers** for a sample are given with NA.
        - Only markers shared with selected population frequencies are imported.
            - The imported database files may contain different markers.
        - Homozygote genotype may have an empty allele under ‘Allele 2’.
        - The database file may contain **any** number of individuals.
    - Tips:
        - It is more efficient to import several small databases than one big.
            - Time usage to import a database file with 16 markes:
                - 1e6 profiles takes about 130 seconds
                    - Requires ~1.3GB memory
                - 5e6 profiles takes about 800 seconds.
                    - Requires ~6.1GB memory
        - Save a lot of time and memory by storing a project to file (See File under toolbar). The imported database will be stored very efficiently.

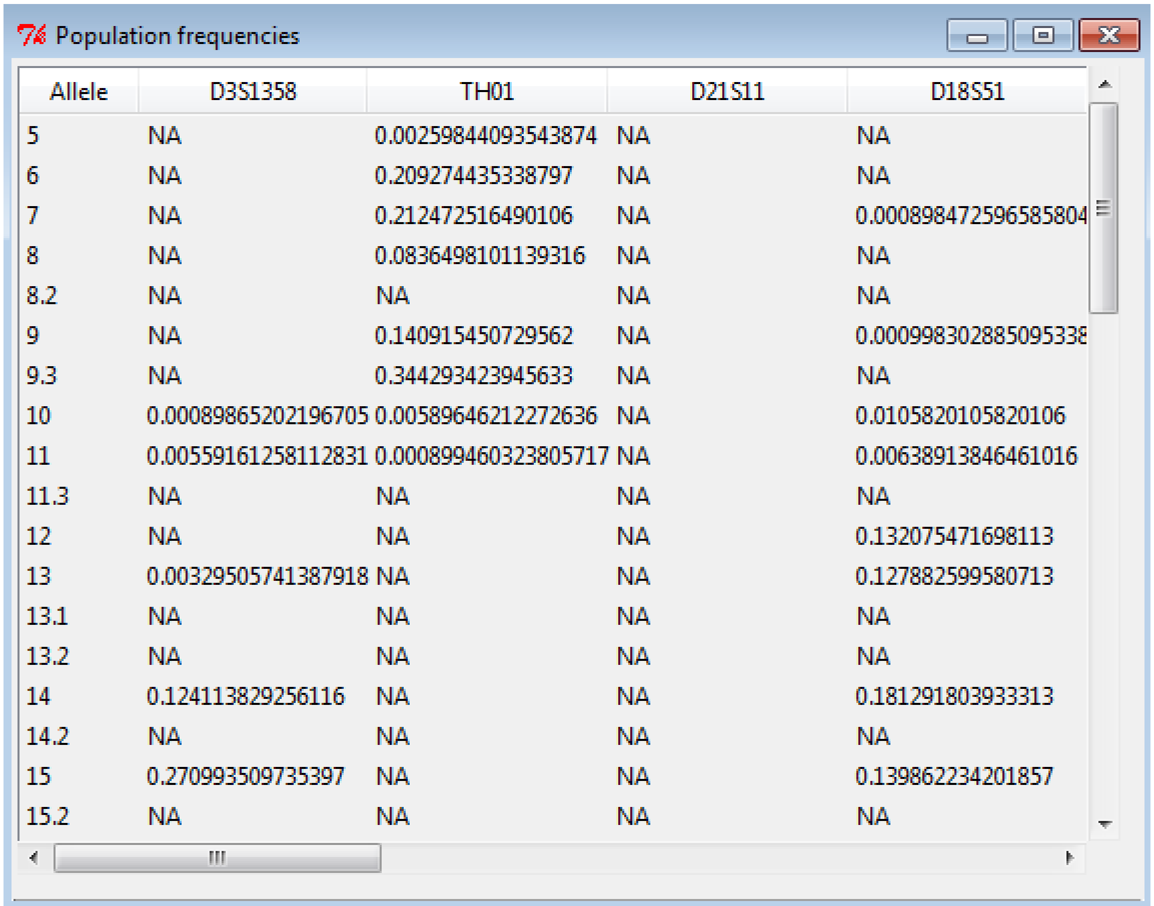


```
[1] "Raw fil import:"
                            Sample.Name   Marker Allele.1 Allele.2
1   00-JP0001-14_20142342311_NO-3241  D3S1358       14       15
2   00-JP0001-14_20142342311_NO-3241     TH01        7      9.3
3   00-JP0001-14_20142342311_NO-3241   D21S11       29       30
4   00-JP0001-14_20142342311_NO-3241   D18S51       13       17
5   00-JP0001-14_20142342311_NO-3241 D10S1248       12       13
6   00-JP0001-14_20142342311_NO-3241  D1S1656       11       14
7   00-JP0001-14_20142342311_NO-3241  D2S1338       17       19
8   00-JP0001-14_20142342311_NO-3241  D16S539       10       11
9   00-JP0001-14_20142342311_NO-3241 D22S1045       15       16
10  00-JP0001-14_20142342311_NO-3241      VWA       17       18
11  00-JP0001-14_20142342311_NO-3241  D8S1179       12       13
12  00-JP0001-14_20142342311_NO-3241      FGA       19       22
13  00-JP0001-14_20142342311_NO-3241   D2S441       11       10
14  00-JP0001-14_20142342311_NO-3241  D12S391       17       18
15  00-JP0001-14_20142342311_NO-3241  D19S433       13       14
16  00-JP0001-14_20142342311_NO-3241     SE33       15       21
17  00-JP0001-14_20142342311_NO-3241     AMEL        X        Y
18  00-JP0002-14_20142342311_NO-3242  D3S1358       15       18
19  00-JP0002-14_20142342311_NO-3242     TH01        6        9
20  00-JP0002-14_20142342311_NO-3242   D21S11       28     31.2
21  00-JP0002-14_20142342311_NO-3242   D18S51       13       18
22  00-JP0002-14_20142342311_NO-3242 D10S1248       13       13
23  00-JP0002-14_20142342311_NO-3242  D1S1656       15     18.3
24  00-JP0002-14_20142342311_NO-3242  D2S1338       25       25
25  00-JP0002-14_20142342311_NO-3242  D16S539       11       13
26  00-JP0002-14_20142342311_NO-3242 D22S1045       15       16
27  00-JP0002-14_20142342311_NO-3242      VWA       14       17
```

Figure 4: The figure shows the table format in the importing reference database file.

VIEW DATA:

- **View frequencies** (see figure 5 for the Norwegian SGMPlus population):
    - Creates a new window which shows the selected population frequencies in a table.
    - If any evidence profiles(s) are selected after evidence-import, the software makes a ‘false positive probability’ – plot for each selected profiles.
        - The plot (figure 6) shows the probability that a random individual **(‘false positive probability’**) matching at least (2*n-wildcardsize) up to 2*n alleles (MAC) with a **selected evidence** profile. Here **n** is number of considered loci (which are both in evidence and population frequencies) and wildcardsize is number of allowed mismatches (default is wildcardsize =7).
        - wildcardsize can be changed under “Frequencies” in Toolbar by changing value **Set number of wildcards in false positive match.**
    - Note:
        - Only allele-information in evidence-profiles are used.
        - New alleles which are not found in the selected population are assumed to have allele-frequency 0.



Allele	D3S1358	TH01	D21S11	D18S51
5	NA	0.00259844093543874	NA	NA
6	NA	0.209274435338797	NA	NA
7	NA	0.212472516490106	NA	0.000898472596585804
8	NA	0.0836498101139316	NA	NA
8.2	NA	NA	NA	NA
9	NA	0.140915450729562	NA	0.00099830288509533
9.3	NA	0.344293423945633	NA	NA
10	0.00089865202196705	0.00589646212272636	NA	0.0105820105820106
11	0.00559161258112831	0.000899460323805717	NA	0.00638913846461016
11.3	NA	NA	NA	NA
12	NA	NA	NA	0.132075471698113
13	0.00329505741387918	NA	NA	0.127882599580713
13.1	NA	NA	NA	NA
13.2	NA	NA	NA	NA
14	0.124113829256116	NA	NA	0.181291803933313
14.2	NA	NA	NA	NA
15	0.270993509735397	NA	NA	0.139862234201857
15.2	NA	NA	NA	NA

Figure 5: The figure shows the viewed frequencies for the Norwegian SGMPlus frequencies.

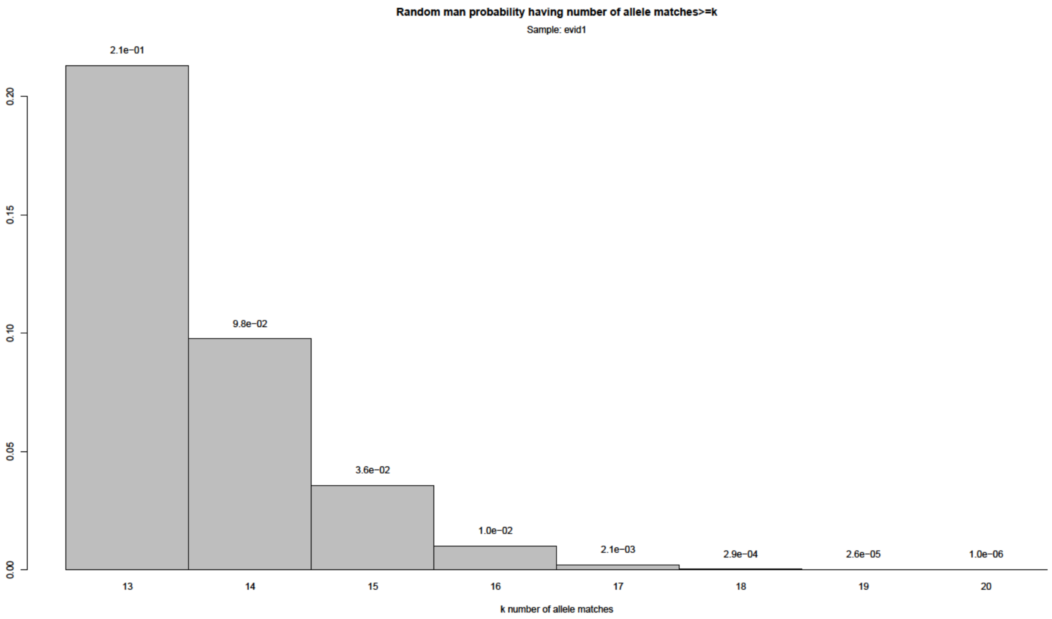
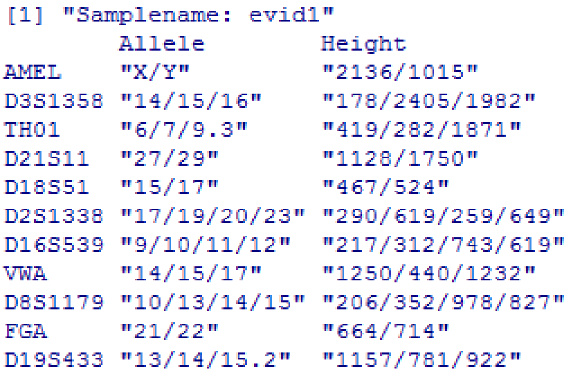


Figure 6: The figure shows the random match probability of matching with at least k number of alleles (in reference) with the observed alleles in evidence.

- **View evidence** (for selected evidence):

  ○ Prints imported alleles (and peak heights if any) for each selected evidence profile(s) (see figure 7).



```
[1] "Samplename: evid1"
        Allele          Height
AMEL    "X/Y"           "2136/1015"
D3S1358 "14/15/16"      "178/2405/1982"
TH01    "6/7/9.3"       "419/282/1871"
D21S11  "27/29"         "1128/1750"
D18S51  "15/17"         "467/524"
D2S1338 "17/19/20/23"   "290/619/259/649"
D16S539 "9/10/11/12"    "217/312/743/619"
VWA     "14/15/17"      "1250/440/1232"
D8S1179 "10/13/14/15"   "206/352/978/827"
FGA     "21/22"         "664/714"
D19S433 "13/14/15.2"    "1157/781/922"
```

Figure 7: The figure shows the printed alleles and heights in the imported evidence.

  ○ Plots EPG(s) (see figure 7) for each selected evidence profile(s)
    ▪ Requires that user have imported “Population frequencies”.
    ▪ The kit selected under ‘**Select kit**’ denotes the EPG format.
    ▪ Loci in evidence which are **inconsistent** with the ones in selected kit (or missing) are **not shown** in plot.
    ▪ Evidence profiles without peak heights for corresponding alleles are given with peak height equal 1.
  ○ Note:
    ▪ See ?plotEPG to see which kit-formats that are supported.

- Reference profiles can be imported as evidence profiles and shown in a EPG.

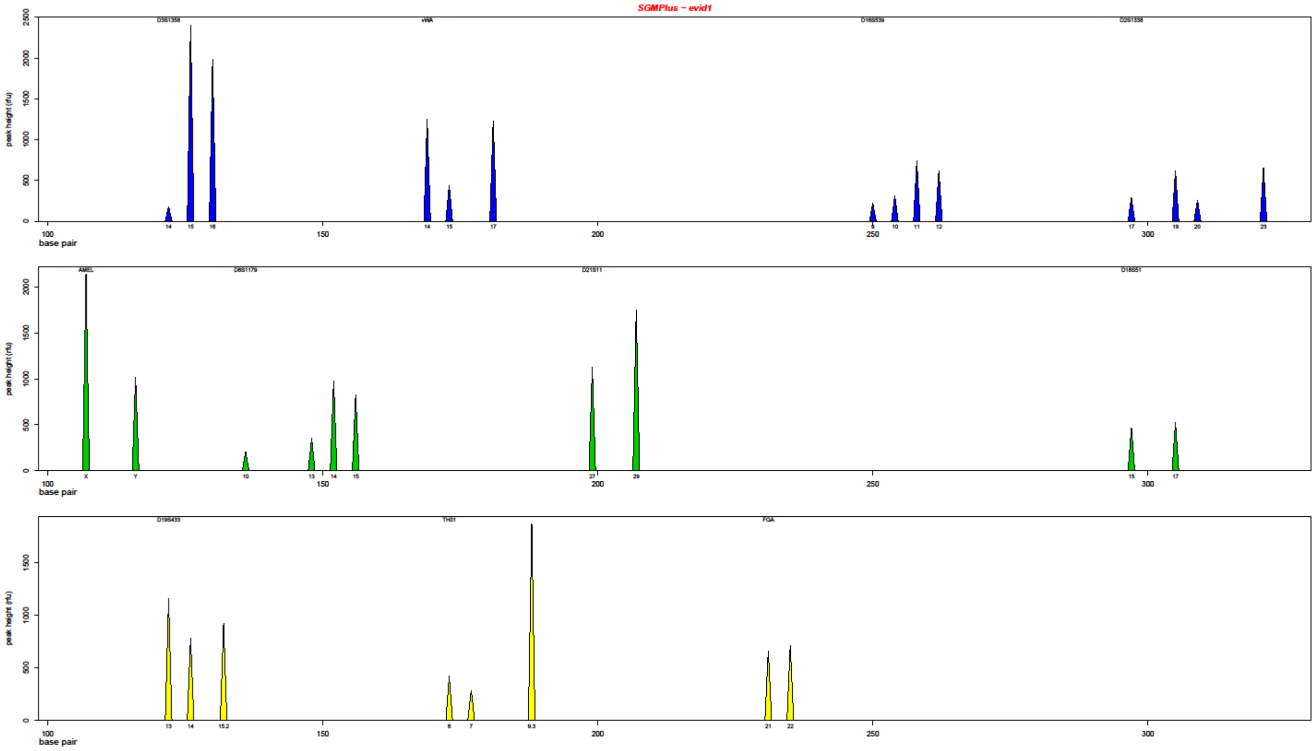
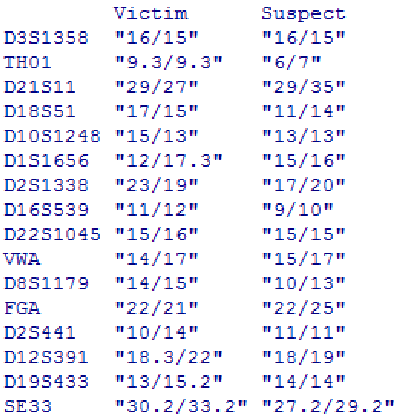


Figure 8: The figure shows the plotted EPG (on selected SGMPlus kit format) of the imported evidence stain.

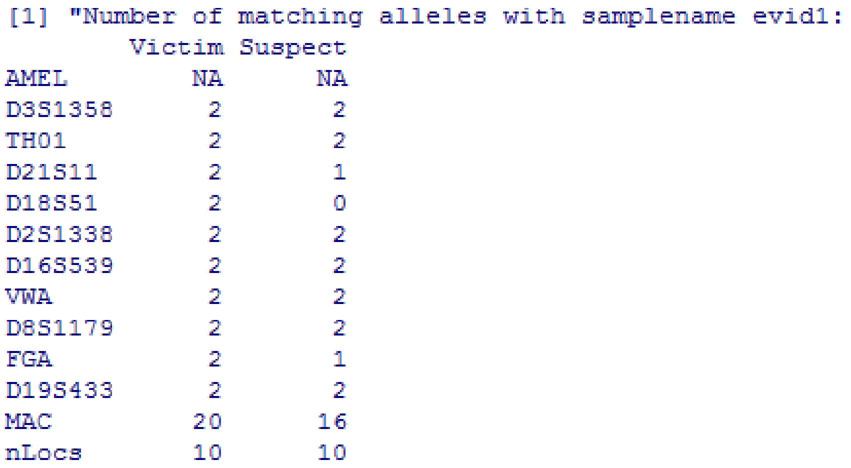
- **View reference** (for selected reference):
    - Prints imported genotypes for each selected reference profile(s) (figure 9).
    - If any evidence profiles(s) are selected after evidence-import, the software counts number of matching alleles (MAC) for each loci of the selected reference profiles, for each selected evidences (figure 10).
        - MAC = number of alleles for the reference which are included in the evidence.
        - nLocs = number of considered loci when counting MAC.



```
          Victim       Suspect
D3S1358   "16/15"      "16/15"
TH01      "9.3/9.3"    "6/7"
D21S11    "29/27"      "29/35"
D18S51    "17/15"      "11/14"
D10S1248  "15/13"      "13/13"
D1S1656   "12/17.3"    "15/16"
D2S1338   "23/19"      "17/20"
D16S539   "11/12"      "9/10"
D22S1045  "15/16"      "15/15"
VWA       "14/17"      "15/17"
D8S1179   "14/15"      "10/13"
FGA       "22/21"      "22/25"
D2S441    "10/14"      "11/11"
D12S391   "18.3/22"    "18/19"
D19S433   "13/15.2"    "14/14"
SE33      "30.2/33.2"  "27.2/29.2"
```

-

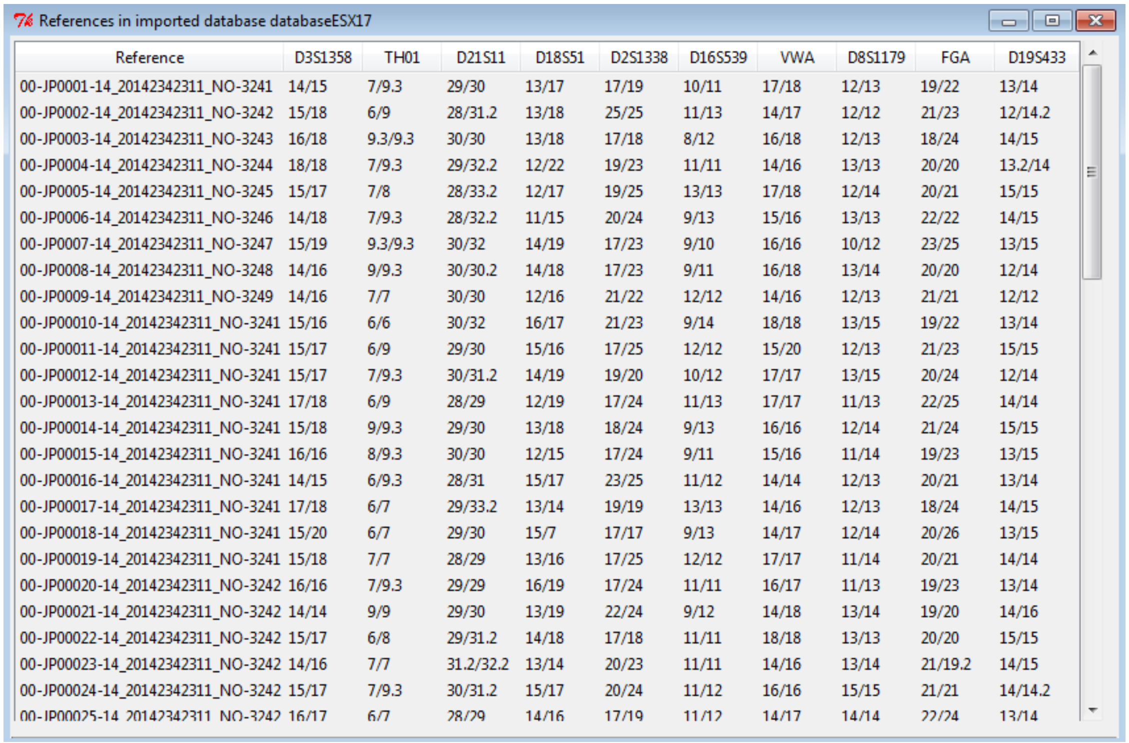
Figure 9: The figure shows the printed alleles of the imported reference profiles.



```
[1] "Number of matching alleles with samplename evid1:
        Victim Suspect
AMEL        NA      NA
D3S1358      2       2
TH01         2       2
D21S11       2       1
D18S51       2       0
D2S1338      2       2
D16S539      2       2
VWA          2       2
D8S1179      2       2
FGA          2       1
D19S433      2       2
MAC         20      16
nLocs       10      10
```

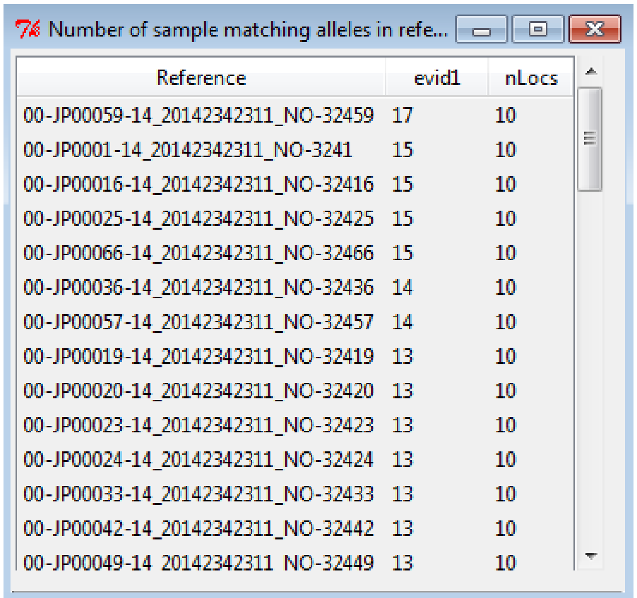
Figure 10: The figure shows number of matching alleles and total (MAC) with the imported and selected evidence stain. By combining the observed MAC and figure 7, the random match probability of observing MAC is useful for providing an extended version of "Random man not excluded"-statistics: The random match probability for Victim (MAC=20) becomes 1/1000000, while only 1/100 for Suspect (MAC=16).

- **View database** (see figure 11 for selected database):
  - ○ Creates a new window (for each selected database) which shows the genotypes for every reference in the database.
    - ▪ "NA" means that the genotype of a reference was missing.
  - ○ If any evidence profiles(s) are selected after evidence-import, the software counts number of matching alleles (MAC) for all references in the database against each of the selected evidences (see figure 12). The results are shown in a MAC-ranked table in a new window (for each selected database).
    - ▪ **MAC** = total number of alleles for the reference which are included in the evidence.
      - Summed over all selected evidences.
    - ▪ **nLocs** is number of reference-loci which has been used to evaluate the MAC.
  - ○ Note:
    - ▪ Max number of individuals to view in a database can be changed with selecting **Set maximum view-elements** under "Database search" in toolbar.



Reference	D3S1358	TH01	D21S11	D18S51	D2S1338	D16S539	VWA	D8S1179	FGA	D19S433
00-JP0001-14_20142342311_NO-3241	14/15	7/9.3	29/30	13/17	17/19	10/11	17/18	12/13	19/22	13/14
00-JP0002-14_20142342311_NO-3242	15/18	6/9	28/31.2	13/18	25/25	11/13	14/17	12/12	21/23	12/14.2
00-JP0003-14_20142342311_NO-3243	16/18	9.3/9.3	30/30	13/18	17/18	8/12	16/18	12/13	18/24	14/15
00-JP0004-14_20142342311_NO-3244	18/18	7/9.3	29/32.2	12/22	19/23	11/11	14/16	13/13	20/20	13.2/14
00-JP0005-14_20142342311_NO-3245	15/17	7/8	28/33.2	12/17	19/25	13/13	17/18	12/14	20/21	15/15
00-JP0006-14_20142342311_NO-3246	14/18	7/9.3	28/32.2	11/15	20/24	9/13	15/16	13/13	22/22	14/15
00-JP0007-14_20142342311_NO-3247	15/19	9.3/9.3	30/32	14/19	17/23	9/10	16/16	10/12	23/25	13/15
00-JP0008-14_20142342311_NO-3248	14/16	9/9.3	30/30.2	14/18	17/23	9/11	16/18	13/14	20/20	12/14
00-JP0009-14_20142342311_NO-3249	14/16	7/7	30/30	12/16	21/22	12/12	14/16	12/13	21/21	12/12
00-JP00010-14_20142342311_NO-3241	15/16	6/6	30/32	16/17	21/23	9/14	18/18	13/15	19/22	13/14
00-JP00011-14_20142342311_NO-3241	15/17	6/9	29/30	15/16	17/25	12/12	15/20	12/13	21/23	15/15
00-JP00012-14_20142342311_NO-3241	15/17	7/9.3	30/31.2	14/19	19/20	10/12	17/17	13/15	20/24	12/14
00-JP00013-14_20142342311_NO-3241	17/18	6/9	28/29	12/19	17/24	11/13	17/17	11/13	22/25	14/14
00-JP00014-14_20142342311_NO-3241	15/18	9/9.3	29/30	13/18	18/24	9/13	16/16	12/14	21/24	15/15
00-JP00015-14_20142342311_NO-3241	16/16	8/9.3	30/30	12/15	17/24	9/11	15/16	11/14	19/23	13/15
00-JP00016-14_20142342311_NO-3241	14/15	6/9.3	28/31	15/17	23/25	11/12	14/14	12/13	20/21	13/14
00-JP00017-14_20142342311_NO-3241	17/18	6/7	29/33.2	13/14	19/19	13/13	14/16	12/13	18/24	14/15
00-JP00018-14_20142342311_NO-3241	15/20	6/7	29/30	15/7	17/17	9/13	14/17	12/14	20/26	13/15
00-JP00019-14_20142342311_NO-3241	15/18	7/7	28/29	13/16	17/25	12/12	17/17	11/14	20/21	14/14
00-JP00020-14_20142342311_NO-3242	16/16	7/9.3	29/29	16/19	17/24	11/11	16/17	11/13	19/23	13/14
00-JP00021-14_20142342311_NO-3242	14/14	9/9	29/30	13/19	22/24	9/12	14/18	13/14	19/20	14/16
00-JP00022-14_20142342311_NO-3242	15/17	6/8	29/31.2	14/18	17/18	11/11	18/18	13/13	20/20	15/15
00-JP00023-14_20142342311_NO-3242	14/16	7/7	31.2/32.2	13/14	20/23	11/11	14/16	13/14	21/19.2	14/15
00-JP00024-14_20142342311_NO-3242	15/17	7/9.3	30/31.2	15/17	20/24	11/12	16/16	15/15	21/21	14/14.2
00-JP00025-14_20142342311_NO-3242	16/17	6/7	28/29	14/16	17/19	11/12	14/17	14/14	22/24	13/14

Figure 11: The figure shows the viewed references inside the imported ESX17 database which are presented only with SGMPlus profiles since the selected kit for the imported frequencies was SGMPlus_Norway.



Reference	evid1	nLocs
00-JP00059-14_20142342311_NO-32459	17	10
00-JP0001-14_20142342311_NO-3241	15	10
00-JP00016-14_20142342311_NO-32416	15	10
00-JP00025-14_20142342311_NO-32425	15	10
00-JP00066-14_20142342311_NO-32466	15	10
00-JP00036-14_20142342311_NO-32436	14	10
00-JP00057-14_20142342311_NO-32457	14	10
00-JP00019-14_20142342311_NO-32419	13	10
00-JP00020-14_20142342311_NO-32420	13	10
00-JP00023-14_20142342311_NO-32423	13	10
00-JP00024-14_20142342311_NO-32424	13	10
00-JP00033-14_20142342311_NO-32433	13	10
00-JP00042-14_20142342311_NO-32442	13	10
00-JP00049-14_20142342311_NO-32449	13	10

Figure 12: The figure shows the sorted references (in the reference database) with respect to MAC (total number of matching alleles) to the selected evidence.

INTERPRETATIONS:

- **Generate sample**:

    - Generates alleles using the population frequencies and draws peak heights for a specified hypothesis using the continuous model as described in the vignette.
    - Requires: Imported population frequencies.
    - Feature: Allele drop-out, Drop-in (with a peak height model) and stutter.

- **Deconvolution**:

    - Deconvolution ranks the most probable combined genotype profiles given **a specified hypothesis** and the Maximum Likelihood Estimates of the parameters in the continuous model (as given in the vignette).
    - Requires: Imported population frequencies and selection of at least one evidence profile with peak height information. References are optional to condition on in the hypothesis.
    - Feature: Model may handle replicates, allele drop-in, drop-out and stutter.

- **Weight-of-Evidence**:

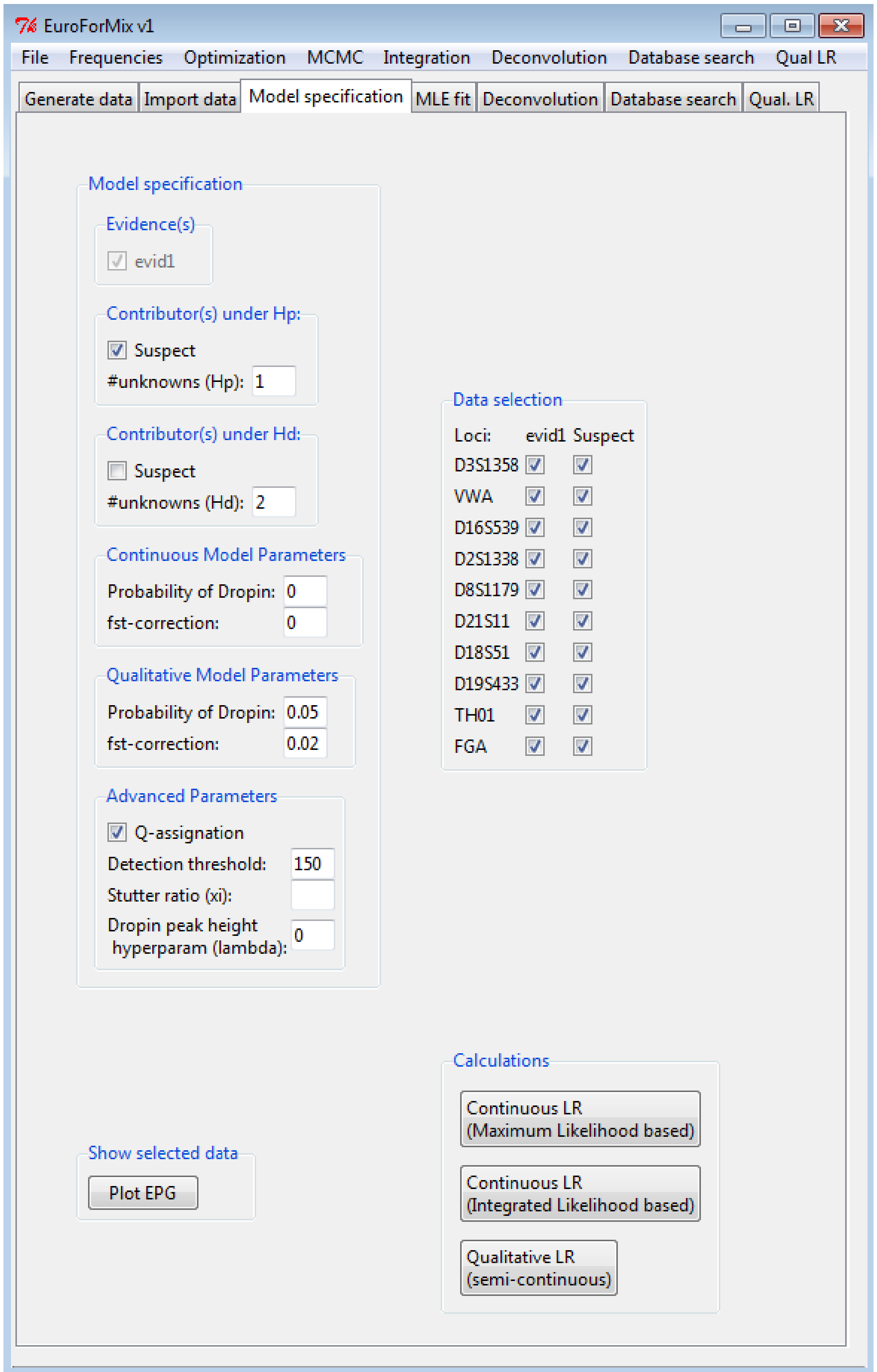
    - Weight-of-Evidence is done by comparing the Likelihood Ratio (LR) between the specified hypotheses Hp (prosecution) and Hd (defence) using the continuous model as given in the vignette.
    - Modules:
        - 1) ‘Continuous LR’ (Maximum Likelihood based)
            - Optimizes (maximum) the model parameters in the continuous model.
        - 2) ‘Continuous LR’ (Integrated Likelihood based)
            - Integrates out the model-parameters in the continuous model.
        - 3) ‘Qualitative LR’ (semi-continous)
            - Explores LR as a function of allele dropout probability parameter.

    - Requires:
        - Imported population frequencies, **at least one** evidence profile and **at least one** reference profile (suspect) to weight evidence for. Additional reference profiles are optional to condition on in the hypotheses.
        - ‘Continuous LR’ requires evidence(s) including peak heights, ‘Qualitative LR’ only requires allele data.
    - Feature:
        - The continuous model: Handles replicates, allele drop-in, drop-out, stutter and fst-correction.
        - The semi-continuous model: Handles replicates, allele drop-in, drop-out and fst-correction.

- **Database search:**

  - Does weight-of-evidence by comparing the Likelihood Ratio (LR) between the specified hypotheses Hj (reference j in database) and Hd (defence) using the continuous model as given in the vignette.
  - Modules:
    - 1) ‘Continuous LR’ (Maximum Likelihood based)
    - 2) ‘Continuous LR’ (Integrated Likelihood based)
    - 3) ‘Qualitatitve LR’ (Semi-continuous based)
  - Requires: Imported population frequencies, **at least one** evidence profile with **peak height** information and **at least one** reference-database. Reference profiles are optional to condition on in the hypotheses.
  - Feature: Model may handle replicates, allele drop-in, drop-out, stutter and fst-correction.
  - The continuous LR value is showed together with qualitative LR and MAC.

2. Model specification



Model specification

Evidence(s)
☑ evid1

Contributor(s) under Hp:
☑ Suspect
#unknowns (Hp): 1

Contributor(s) under Hd:
□ Suspect
#unknowns (Hd): 2

Continuous Model Parameters
Probability of Dropin: 0
fst-correction: 0

Qualitative Model Parameters
Probability of Dropin: 0.05
fst-correction: 0.02

Advanced Parameters
☑ Q-assignation
Detection threshold: 150
Stutter ratio (xi):
Dropin peak height hyperparam (lambda): 0

Data selection

Loci:	evid1	Suspect
D3S1358	☑	☑
VWA	☑	☑
D16S539	☑	☑
D2S1338	☑	☑
D8S1179	☑	☑
D21S11	☑	☑
D18S51	☑	☑
D19S433	☑	☑
TH01	☑	☑
FGA	☑	☑

Figure 13: The figure shows the <u>Model Specification</u> GUI page for **Weight-of-Evidence** based on Likelihood Ratio calculation.

MODEL SPECIFICATION

- **Evidence(s)**:
  - Shows selected evidence(s) from ‘Import data’.
  - All interpretations support **multiple replicates**.
    - Note: All replicates are assumed to have same parameter sets.

- **Contributors under Hp**
  - Case: **Weight-of-Evidence** or **‘Database search’):**
    - User may condition on selected references (from ‘Import data’) in the hypothesis Hp.
    - #unknowns under Hp: Denotes number of unknown contributors under the prosecution hypothesis Hp.
  - Case: **‘Database search’:**
    - The individual in the reference-database is already included in the hypothesis Hp.
  - Case: **Deconvolution** or **‘Generate sample’**:
    - This block is not considered, since Deconvolution only considers the model under Hd, and sample generation is done only under a specific hypothesis.

- **Contributors under Hd** (same for **all** cases):
  - User may condition on selected references (from ‘Import data’) in the hypothesis Hd.
  - #unknowns under Hd: Denotes number of unknown contributors under the prosecution hypothesis Hd.
  - Case: **Weight-of-Evidence** or **‘Database search’**:
    - References which are conditioned under Hp but not under Hd, will be assumed to be a ‘**known non-contributor’** under Hd (this is relevant when fst>0).

- **Continuous Model Parameters** and **Qualitative Model Parameters**:
  - The Continuous Model Parameter section is only used for “Continuous LR” Calculations, while Qualitative Model Parameters section is only used for ‘Qualitative LR’ Calculations.

  - ‘**Probability of drop-in**’: [0,1]
    - Assumed probability of a random allele drop-in to the evidence at a given locus. See vignette for more details.
    - This is default 0 for continuous models and 0.05 for qualitative models.

  - **fst-correction**: [0,1]
    - Assumed co-ancestry parameter assigned in the genotype probability for each contributor in the hypotheses. See vignette for more details.
    - This is default 0 for continuous models and 0.02 for qualitative models.

    - Case **‘Database search’:**
        - When doing database search with “Continuous LR” Calculations, the allele drop-in probability for the qualitative LR can be changed by **Set drop-in probability for qualitative model** under “Database search” in toolbar (default is 0.05). When doing database search with “Qualitative LR” Calculations, this value is ignored in favor of the specification under “Qualitative Model Parameters”.

    - Case **Generation** and **Deconvolution**:
        - The Qualitative Model Parameters section is removed.

- **Advanced Parameters**

    - **Q-assignation**:

        - If checked, all alleles **not** present in the evidence are considered as allele “99”. Its frequency will be given as the sum of the frequencies for all the “non-present” alleles.
        - If unchecked, the original alleles in the population are used as before.

    - ‘**Detection threshold’**: [0,->)

        - The threshold of required allele peak heights of whether an allele is present in the evidence or not.
            - Note: If peak heights in evidence are lower than the specified threshold, the corresponding alleles (and peak heights) below threshold **are removed** automatically. This may cause some loci to become empty.

    - ‘**Stutter ratio’**: [0,1]

        - Only used for ‘Continuous LR’ Calculations.
        - Stutter ratio is a constant parameter “**xi**” which denotes the proportion of peak heights from allele ‘a’ which is added to allele ‘a-1’. See vignette for more details.
            - If allele 22 with peak height y_22 is contributed by a contributor and allele 23 did not have any observed peak height, then the stutter contribution to allele 21 from allele 22 will be (**xi *** y_22).

    - ‘**Dropin peak height hyperparam’**: [0,1]

        - Only used for ‘Continuous LR’.
        - Assumed hyper-parameter to model the peak height of the dropped in allele caused by a ‘random allele drop-in’ if ‘**Probability of drop-in’**>0. See vignette for more details.

- **‘Database(s) to search’** (case**: ‘Database search**’)
    - Lists the selected imported reference-database(s) to do the database search for.

DATA SELECTION

- **Select/unselect loci**:
    - The user may select or unselect loci for each selected evidence(s) and reference(s) from "Import data"
    - If a locus has been unselected for any of the evidence(s) or reference(s), the unselected locus will not be evaluated at all.
    - Note: Evidence with more than 30 loci will not be able to be selected.

- **Missing data**:
    - Data with missing allele in any of the loci will automatically be deselected (inactivated) such that the corresponding loci will be unavailable to evaluate.
    - For continuous LR evaluation:
        - If peak heights (in any of the evidence(s)) are missing for any selected locus, the user gets a message about deselecting the issued loci before proceeding.

- **New alleles**:
    - If new alleles (does not exist in the population frequency table) occurs in the imported evidence or reference profile, the new alleles are assigned with allele frequency 'freq0'. 'freq0' is equal minimum observed frequency in population if N=0, or 'freq0'=5/(2N) where N is size of imported frequency database under "Frequencies" in Toolbar. The frequencies are after normalized.

SHOW SELECTED DATA

- **Plot EPG:**
    - **Prints** the selected evidence sample(s), reference(s) and considered population frequencies which are eventually used for further analysis **out to terminal**.
    - The selected evidence samples are shown in an EPG-plot.
        - Note: Alleles with corresponding peak heights below the specified "Detection Threshold" are removed.

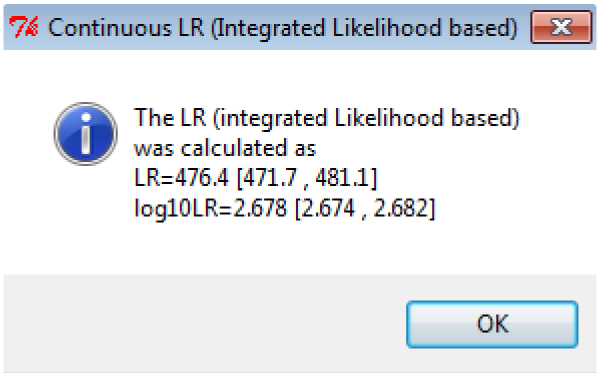
CALCULATIONS

- **'Continuous LR (Maximum Likelihood based) '** (case **Weight-of-Evidence** and **'Database search')**:
    - Maximizes the Likelihood of the unknown parameters in the continuous model given the assumed model so they attain maximum values for the specified hypothesis Hd (and Hp in case of Weight-of-Evidence).

    - The optimizer should return a global maximum. However, it may sometimes just return a local maximum. Number of start-points should be increased to ensure that the optimizer finds the global maximum of the Likelihood function. This can be changed under “Optimization” in Toolbar.
  - After calculation, the page ‘MLE fit’ is visited to present maximized results.

- **‘Continuous LR (Integrated Likelihood based)’** (case **Weight-of-Evidence** and **‘Database search’**):

  - Instead of optimizing the Likelihood of the unknown parameters, a **multivariate integration** over the unknown parameters are applied both under hypothesis Hp and Hd.
  - The accuracy of the integral depends on the specified ‘**relative error requirement’** (see vignette for details).
    - Can be changed under “Integration” in Toolbar. Default is 0.005.
  - In the output (see Figure 14), also the relative error of the LR is given in brackets.
  - The integral requires that an **upper boundary** for the parameters mu (amount of DNA) and sigma (coefficient of variation) are specified. As default these are 20000 and 1, respectively. These values may be changed under “Integration” in Toolbar. See vignette for details.
  - Calculates LR-values directly and avoids visiting the tab ‘MLE fit’.
    - Case **Weight-of-Evidence**: A message with LR pops up after calculation (see Figure 14).
    - Case **‘Database search’**: Database search results are shown directly after calculation (goes to tab ‘Database search’).
  - ‘Continuous LR (Integrated Likelihood based)’ is not possible for multiple replicates and large number of loci since it doesn’t evaluate on log-scale. Use the Maximum Likelihood based method instead if the other method goes wrong.



The LR (integrated Likelihood based)
was calculated as
LR=476.4 [471.7 , 481.1]
log10LR=2.678 [2.674 , 2.682]

Figure 14: The figure shows the calculated Weight-of-Evidence based the Integrated Likelihood based continuous LR for the specified model in Figure 13.

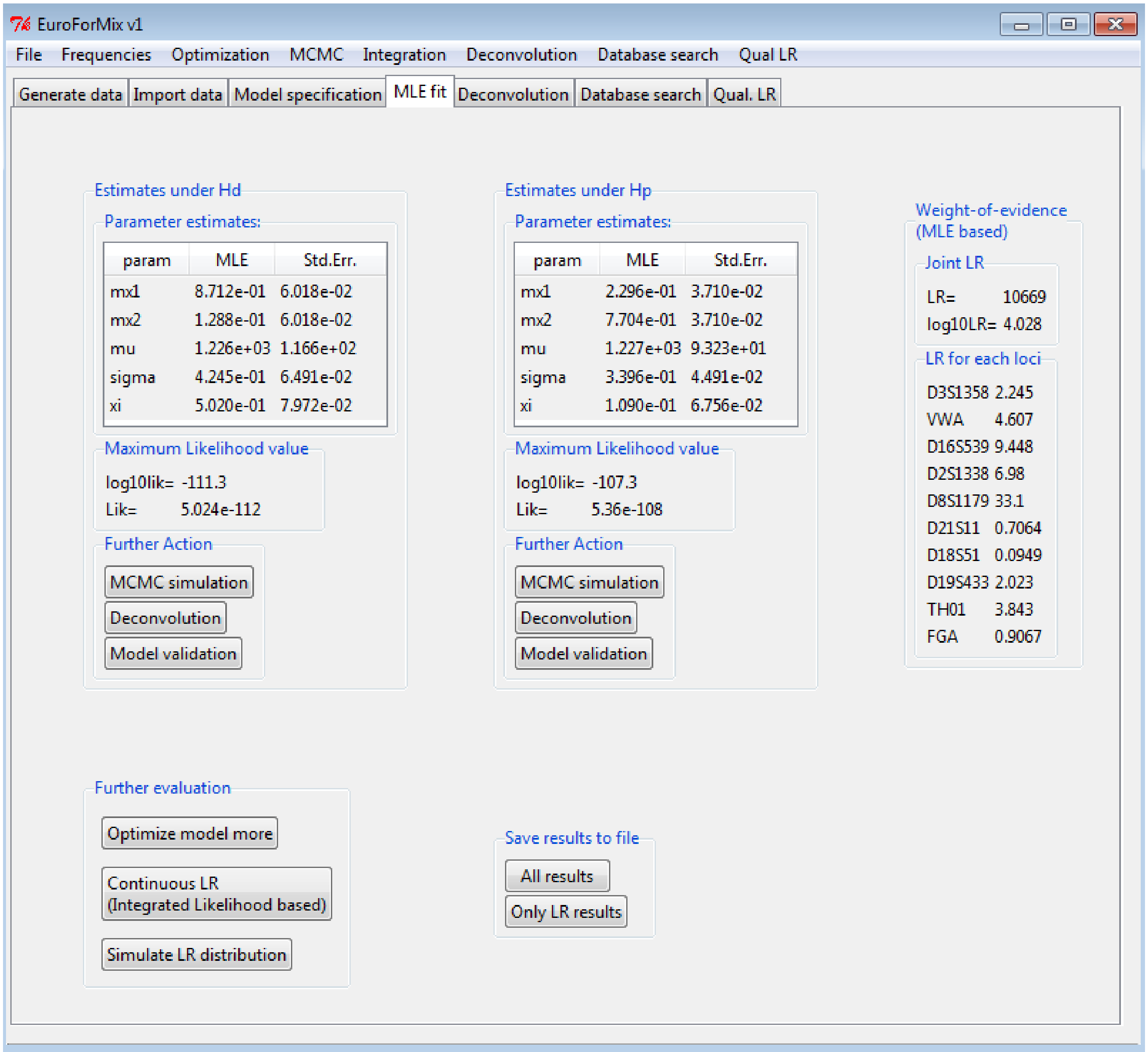
- **‘Qualitative LR (semi-continuous)’** (case **Weight-of-Evidence**)

  - Performs a semi-continuous procedure where the distribution of the ‘allele drop-out probability given number of observed alleles’ are utilized to infer a “conservative” LR.
    - The model is purely qualitative which means it is only based on allele-information.
  - Goes directly to page Qual. LR.

- **‘Generate sample’** (case **‘Generate sample’**):

  - A dataset (evidence sample and contributing references) will be randomly simulated under the specified model under “Model specification”.
  - Reference profiles may be imported and selected as assumed known in the hypothesis.
  - Detection threshold, stutter ratio, probability of drop-in and drop-in peak height hyperparam may all be used in the simulation (**fst** are not used).
  - The unknown contributor profiles under the hypothesis will be randomly generated using the selected population frequencies.
  - The simulated peak heights of the evidence in the dataset are entirely based on the continuous model for assumed values of the model-parameters (**mu,sigma,xi,mx)**. Default these are given as **mu**=1000, **sigma**=0.15, **xi**=0.1, **mx**=(C:1)/sum(C:1), where C is number of contributors.
  - Goes directly to page Generate data.

3. MLE fit: ('Continuous LR (Maximum Likelihood based)')



Estimates under Hd

Parameter estimates:

param	MLE	Std.Err.
mx1	8.712e-01	6.018e-02
mx2	1.288e-01	6.018e-02
mu	1.226e+03	1.166e+02
sigma	4.245e-01	6.491e-02
xi	5.020e-01	7.972e-02

Maximum Likelihood value

log10lik= -111.3

Lik= 5.024e-112

Further Action

Estimates under Hp

Parameter estimates:

param	MLE	Std.Err.
mx1	2.296e-01	3.710e-02
mx2	7.704e-01	3.710e-02
mu	1.227e+03	9.323e+01
sigma	3.396e-01	4.491e-02
xi	1.090e-01	6.756e-02

Maximum Likelihood value

log10lik= -107.3

Lik= 5.36e-108

Further Action

Weight-of-evidence (MLE based)

Joint LR

LR= 10669

log10LR= 4.028

LR for each loci

D3S1358	2.245
VWA	4.607
D16S539	9.448
D2S1338	6.98
D8S1179	33.1
D21S11	0.7064
D18S51	0.0949
D19S433	2.023
TH01	3.843
FGA	0.9067

Further evaluation

Save results to file

Figure 15: The figure shows the MLE-fit GUI page after doing **continuous LR (Maximum Likelihood based)** calculation (maximizing the continuous model with respect to the unknown parameters for each of the specified hypothesis in figure 13) for **Weight-of-Evidence**.

ESTIMATES UNDER Hd (and Hp for case: **Weight-of-Evidence**)

- **Parameter estimates**:

    - param: The unknown parameters in the model (see vignette for more details).
        - mx_i: Mixture-proportion for contributor 'i'.
        - mu: Expected amount of DNA.
        - sigma: Coefficient of variation.
        - xi: Stutter ratio (fraction of peak height that are stutter).

    - MLE: The optimized[1] parameters in the model which attains a maximum point of the likelihood function.

    - Std.Err.: The standard error of the parameter estimates in the model (see vignette for details).

- **Maximum Likelihood value**:
    - log10lik and Lik: The ten-logged and the original value of the Likelihood value attained from the optimization[1].

- **Further Action**:

    - **MCMC simulation** (see Figure 16):
        - Performs 'Markov Chain Monte Carlo (MCMC) random walk Metropolis' samples under the desired hypothesis.
            - Uses the mode and the covariance matrix attained from the optimization. See vignette for details.
        - The **first column** in the output shows the estimated posterior distributions for each of the unknown parameters in the model.
        - The **second column** in the output monitors the parameter samples in the simulation.
        - After sampling, the **acceptance rate** of the sampler is printed out to the terminal.
            - Acceptance ratio = number of accepted samples divided by number of proposed samples.
            - Ideally the acceptance rate should be around 0.2 to ensure that the parameter space has been fully explored.
                - Tweak '**variance of randomizer**' under MCMC in toolbar to change the acceptance rate.
        - User may **change number of required samples** in the simulation under 'MCMC' in toolbar.
        - The **purpose** of the MCMC simulation is to use it as an **exploratory tool** to see:
            - That the optimizer has found the global maximum.
            - The shape of the posterior distribution of the parameters.

[1] This may be only a local maximum point, not the global maximum (i.e. the Maximum Likelihood Estimate). Increase **number of start points** under "Optimization" in Toolbar to ensure a global maximum.

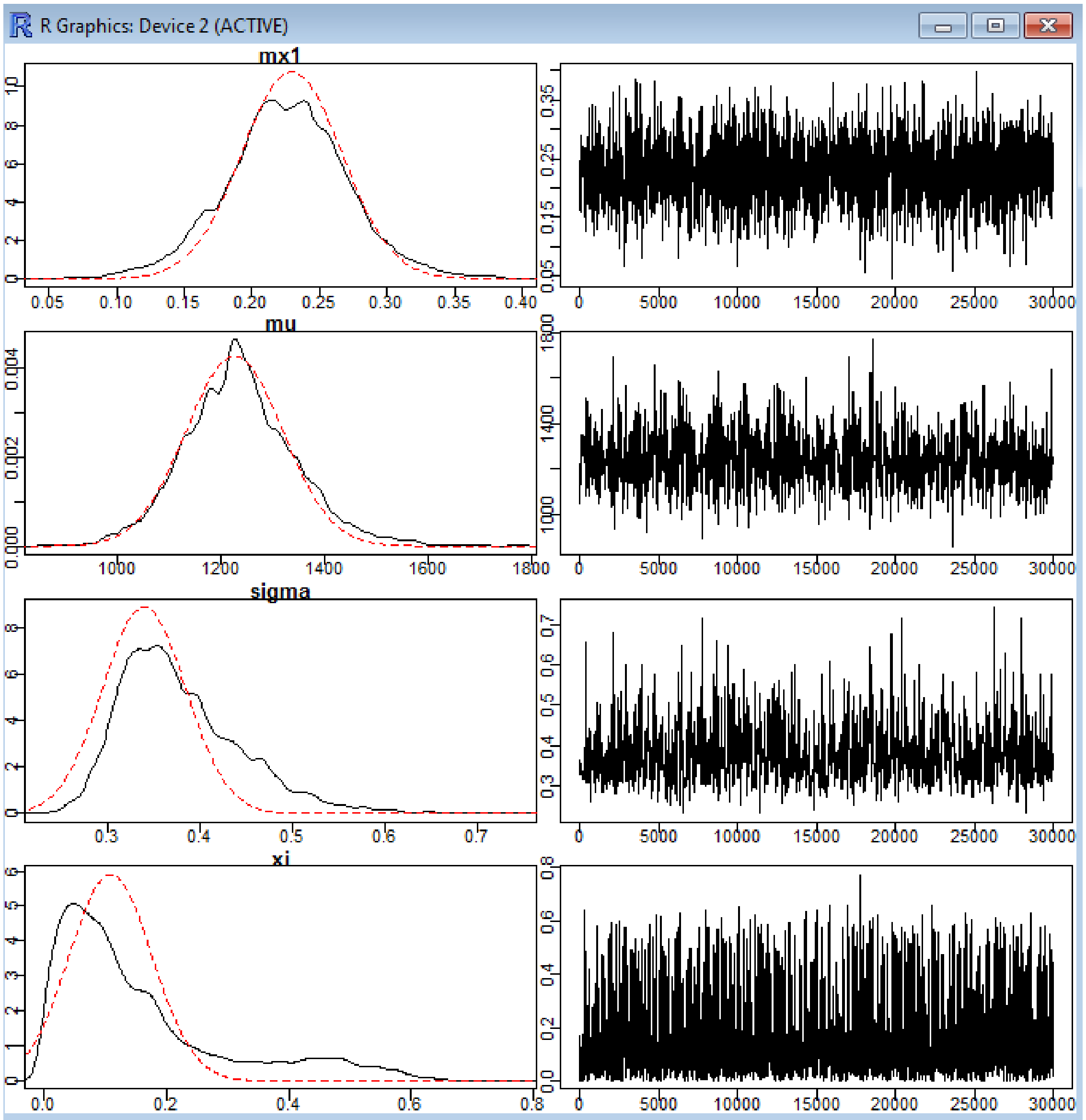


Figure 16: The figure shows the posterior density of the unknown parameters (first column) and corresponding iteration values (second column) from the MCMC method under the hypothesis Hp: “Suspect+1 unknown individual contributes to evidence evid1”. The acceptance ratio was given as 0.35.

- **Deconvolution**:
  - Performs “Deconvolution” under the desired hypothesis. (See Deconvolution (page 5) for details.

- **Model validation** (Figure 17):
  - Uses a statistical hypothesis test to reject whether the maximum likelihood fitted model fits the observed peak heights (i.e. whether the gamma model assumption is reasonable).
  - Estimates the cumulative probability of the observed peak heights conditional on the other peak heights (see vignette for more details).

    - Uses a one-sample Kolmogorov-Smirnov test to test if the observed cumulative probability deviates significant from the uniform distribution.
  - P-value from the test is printed out to terminal.
  - A textbox is shown when the P-value is lower than the significance level 0.05 (i.e. rejection of assumption).

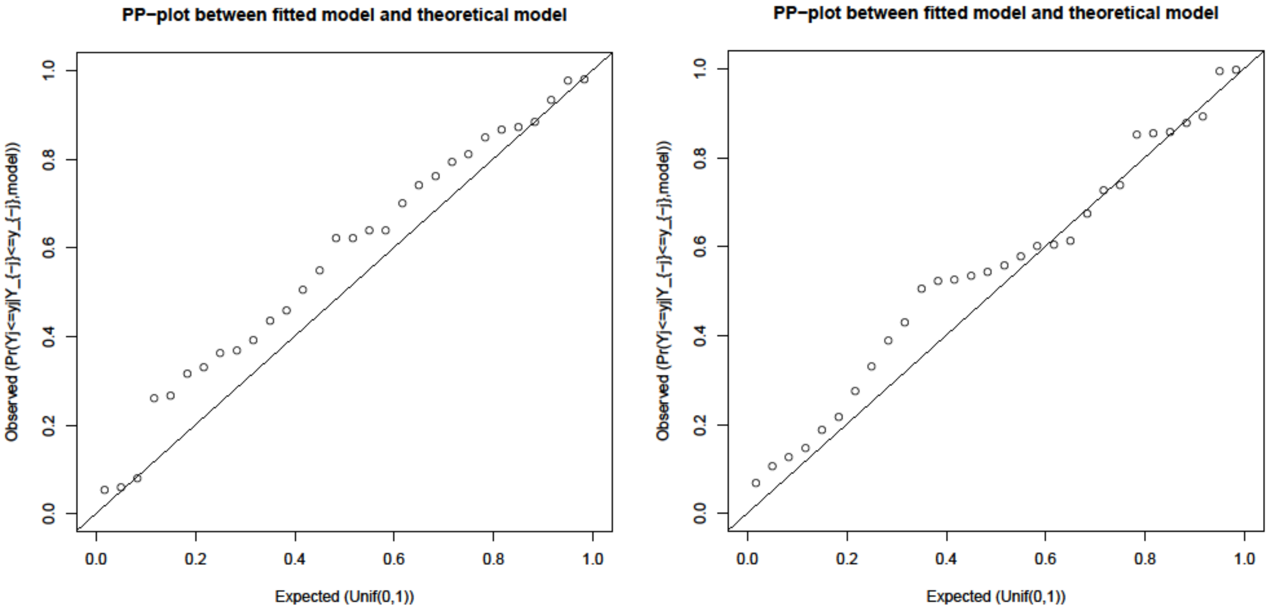


Figure 17: Left subplot shows the "**Model validation**" under hd with p-value 0.37. Right subplot is "**Model validation**" under hp with p-value 0.29.

WEIGHT-OF-EVIDENCE **(case Weight-of-Evidence)**

- Description:
  - The Weight-of-Evidence value is the ratio between the likelihoods of the two specified hypotheses Hp and Hd as specified in "Model specification".
  - The Weight-of-Evidence value is based on the continuous model as described in the vignette and handles allele drop-in, drop-out and stutter.
- **Join LR:**
  - LR: 'Likelihood value under optimization under Hp' divided by 'Likelihood value under optimization under Hd'
  - log10: The ten-logged value of LR.

- **LR for each loci:**

  - The LR for each loci separately (given the parameter-modes under Hp and Hd). See vignette for details.
  - Note: This will not be shown for evaluation of more than 30 loci

FURTHER EVALUATION

- **Optimize model more**:

  - The optimization procedure can be run again with the same specifications as selected in "Model specification" to ensure that a global maximum is attained.
  - It is recommended to do this and check that the optimized Likelihood value is not increased further.

- **Database search (case: 'Database search'):**

  - A database search with the specified continuous model will be applied. (See Database search for details.

- **'Continuous LR (Integrated Likelihood based)' (case Weight-of-Evidence)**

  - See CALCULATIONS under section "Model specification".

- **'Simulate LR distribution' (case Weight-of-Evidence)**

  - MCMC simulation will be applied both under Hp and Hd to provide a plot of a "Bayesian" distribution of the LR where the uncertainty of the parameters in the continuous model under both Hp and Hd are taken into account (see Figure 18).
    - Number of samples can be changed with **Set number of samples** under MCMC in Toolbar (default is 10000 samples).

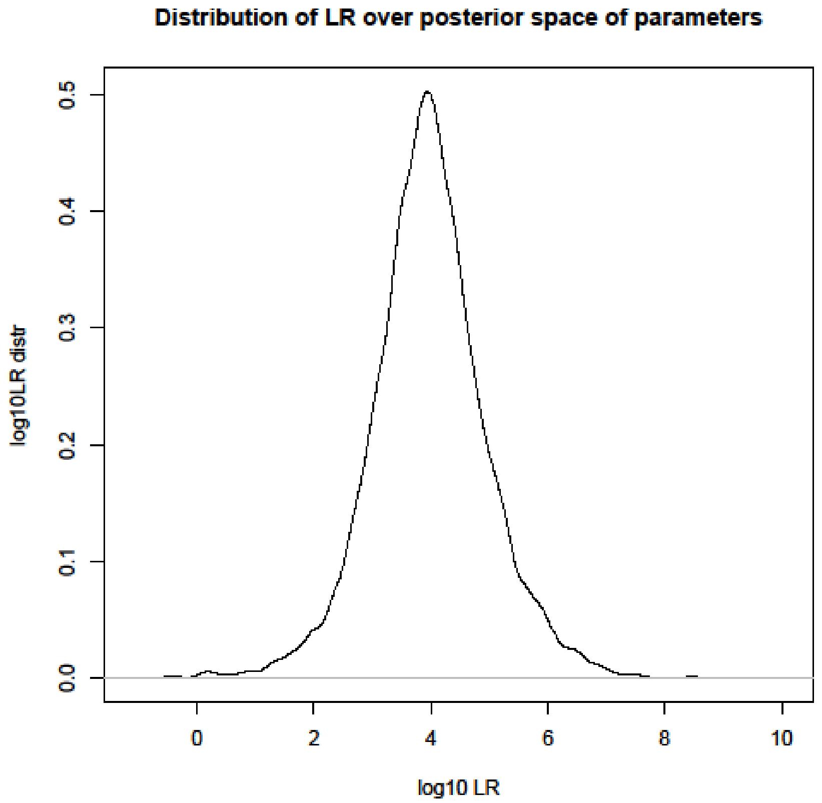
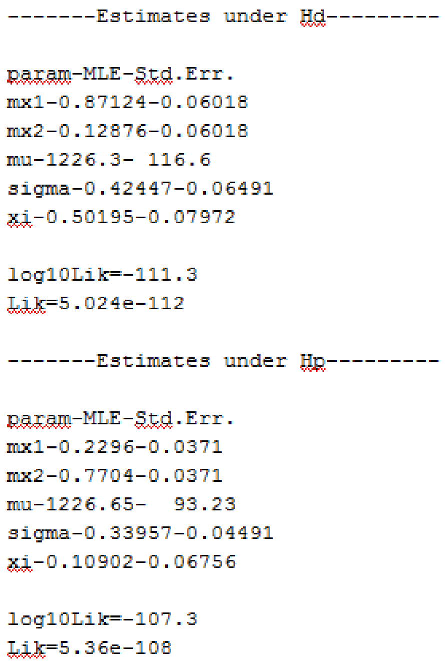


Figure 18: The plot shows the distributed LR where the *a posteriori* density of the parameters in the continuous model under both Hp and Hd are taken into account. *a posteriori* density are simulated using the **MCMC simulation** (Figure 16 shows only Hp)**.**

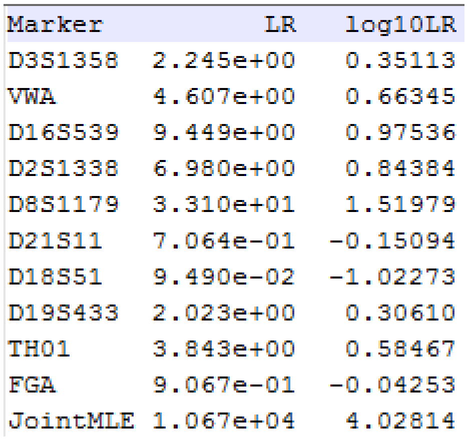
SAVE RESULTS TO FILE

- **'All results'**:
  - The parameter estimates with corresponding standard deviation errors estimates and the likelihood values will be printed to file for all hypotheses on page (see below).



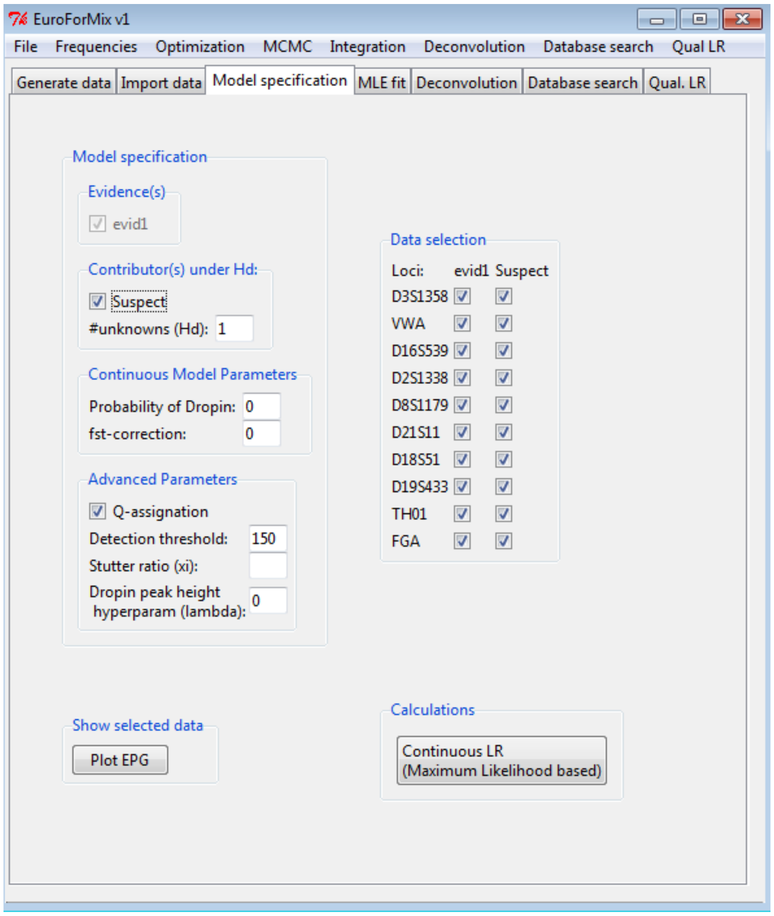
```
-------Estimates under Hd---------

param-MLE-Std.Err.
mx1-0.87124-0.06018
mx2-0.12876-0.06018
mu-1226.3- 116.6
sigma-0.42447-0.06491
xi-0.50195-0.07972

log10Lik=-111.3
Lik=5.024e-112

-------Estimates under Hp---------

param-MLE-Std.Err.
mx1-0.2296-0.0371
mx2-0.7704-0.0371
mu-1226.65-  93.23
sigma-0.33957-0.04491
xi-0.10902-0.06756

log10Lik=-107.3
Lik=5.36e-108
```

- **‘Only LR results’: (case Weight-of-Evidence)**
    - ○ The LR calculated values shown in WEIGHT-OF-EVIDENCE will be printed to file (see below).



Marker	LR	log10LR
D3S1358	2.245e+00	0.35113
VWA	4.607e+00	0.66345
D16S539	9.449e+00	0.97536
D2S1338	6.980e+00	0.84384
D8S1179	3.310e+01	1.51979
D21S11	7.064e-01	-0.15094
D18S51	9.490e-02	-1.02273
D19S433	2.023e+00	0.30610
TH01	3.843e+00	0.58467
FGA	9.067e-01	-0.04253
JointMLE	1.067e+04	4.02814

4. Deconvolution:



Evidence(s)

☑ evid1

Contributor(s) under Hd:

☑ Suspect

#unknowns (Hd): 1

Continuous Model Parameters

Probability of Dropin: 0

fst-correction: 0

Advanced Parameters

☑ Q-assignation

Detection threshold: 150

Stutter ratio (xi):

Dropin peak height hyperparam (lambda): 0

Data selection

Loci:	evid1	Suspect
D3S1358	☑	☑
VWA	☑	☑
D16S539	☑	☑
D2S1338	☑	☑
D8S1179	☑	☑
D21S11	☑	☑
D18S51	☑	☑
D19S433	☑	☑
TH01	☑	☑
FGA	☑	☑

Figure 19: The figure shows the <u>Model Specification</u> GUI page for doing **Deconvolution**. We condition on the suspect, and assume one unknown in the hypothesis. Our model assumes unknown “(n-1)- stutter” ratio, no allele drop-in and no theta-correction.

- Description:

  - Deconvolution is applied for a specific hypothesis Hd as shown in Figure 19.
  - The deconvolution conditions on the optimized parameters (i.e. the <u>MLE fit</u> in Figure 20) for the continuous model.
  - The deconvolution result shows (see Figure 21) a ranked list of the **posterior probabilities** of the combined genotype-profiles (see vignette for details).
  - Since the deconvolution is based on the continuous model it may handle multiple replicates, allele drop-in, drop-out and stutter.

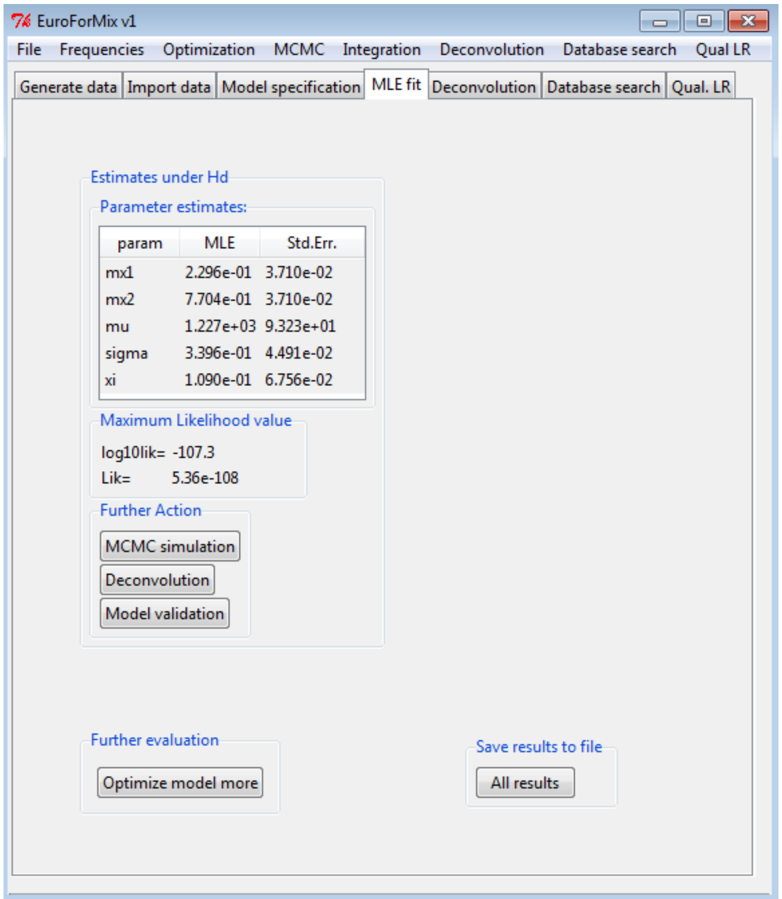
- **Table**:

  - The columns in the table (see Figure 21) show the resolved genotype for each contributor in the specified hypothesis (per locus).

  - The combined profiles are ranked due to their **posterior probabilities**.
  - The ranked elements in the table ensures that the sum of the **posterior probabilities** are at least 0.9999.
    - Can be changed under ‘Deconvolution’ in toolbar.
  - Maximum length of table is default 10000.
    - Can be changed under ‘Deconvolution’ in toolbar.
  - Note:
    - Having only sub-optimized parameters (in the **MLE fit**)will not give the most likely genotypes.
    - Q-assignation is recommended to use since dropped out alleles are equally threated and assigned as “99”.

- **Save table:**

  - The **full** table will be exported to a tabulator-separated text-file.



Estimates under Hd

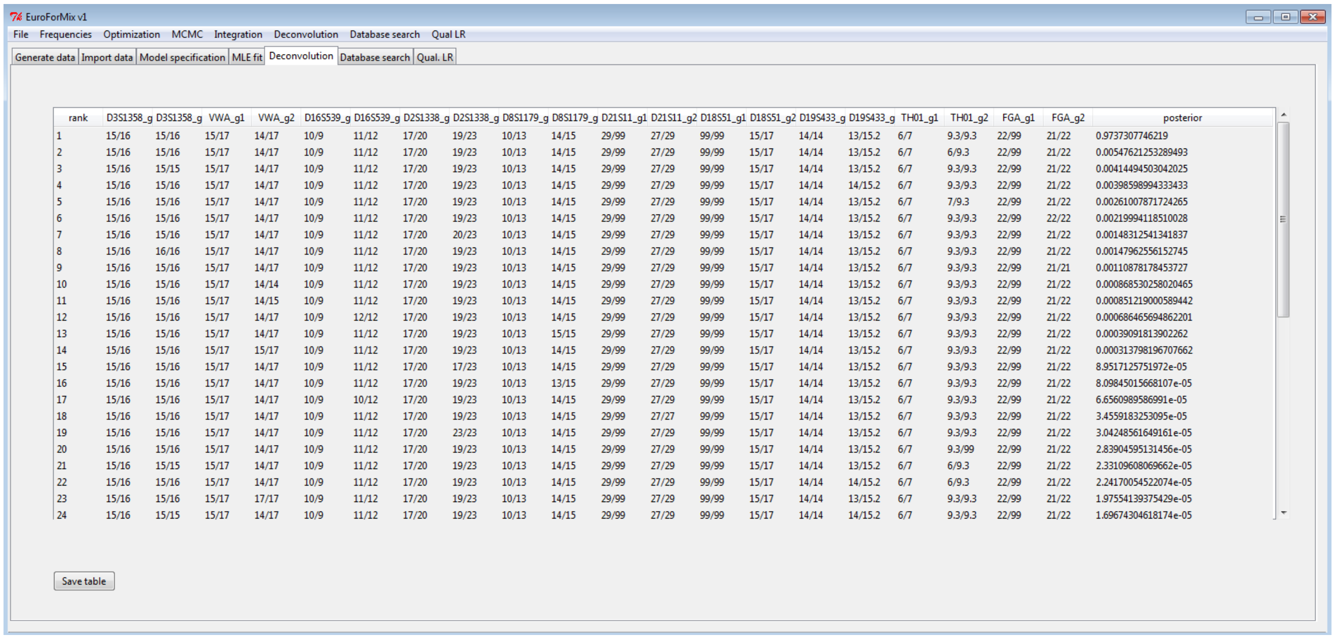
Parameter estimates:

param	MLE	Std.Err.
mx1	2.296e-01	3.710e-02
mx2	7.704e-01	3.710e-02
mu	1.227e+03	9.323e+01
sigma	3.396e-01	4.491e-02
xi	1.090e-01	6.756e-02

Maximum Likelihood value

log10lik= -107.3
Lik= 5.36e-108

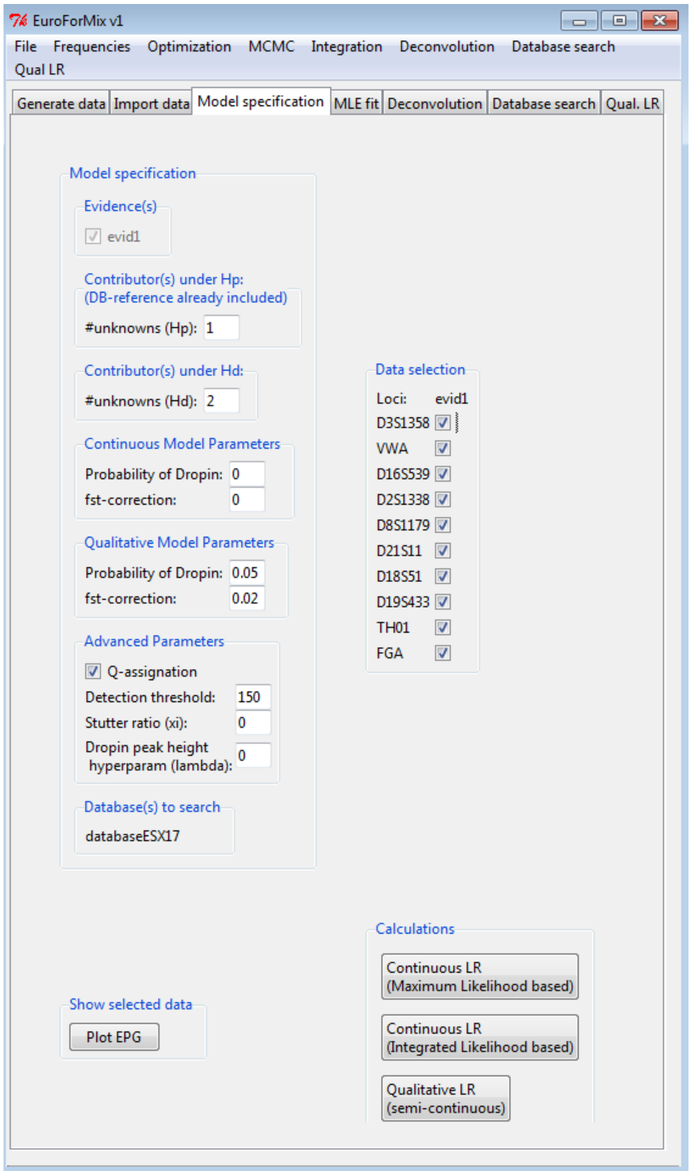
Figure 20: The figure shows the optimized parameters (i.e. the MLE fit) for the continuous model. The fitted model has the same “Further Action” possibilities as for “Weight-of-Evidence” and “Database search”.


EuroForMix v1
File Frequencies Optimization MCMC Integration Deconvolution Database search Qual LR
Generate data | Import data | Model specification | MLE fit | Deconvolution | Database search | Qual. LR

rank	D3S1358_g	D3S1358_g	VWA_g1	VWA_g2	D16S539_g	D16S539_g	D2S1338_g	D2S1338_g	D8S1179_g	D8S1179_g	D21S11_g1	D21S11_g2	D18S51_g1	D18S51_g2	D19S433_g	D19S433_g	TH01_g1	TH01_g2	FGA_g1	FGA_g2	posterior
1	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.9737307746219
2	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	6/9.3	22/99	21/22	0.00547621253289493
3	15/16	15/15	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.00414494503042025
4	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	14/15.2	6/7	9.3/9.3	22/99	21/22	0.00398598994333433
5	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	7/9.3	22/99	21/22	0.00261007871724265
6	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	22/22	0.00219994118510028
7	15/16	15/16	15/17	14/17	10/9	11/12	17/20	20/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.00148312541341837
8	15/16	16/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.00147962556152745
9	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/21	0.00110878178453727
10	15/16	15/16	15/17	14/14	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.000868530258020465
11	15/16	15/16	15/17	14/15	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.000851219000589442
12	15/16	15/16	15/17	14/17	10/9	12/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.000686465694862201
13	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	15/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.00039091813902262
14	15/16	15/16	15/17	15/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	0.000313798196707662
15	15/16	15/16	15/17	14/17	10/9	11/12	17/20	17/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	8.9517125751972e-05
16	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	13/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	8.09845015668107e-05
17	15/16	15/16	15/17	14/17	10/9	10/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	6.6560989586991e-05
18	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/27	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	3.4559183253095e-05
19	15/16	15/16	15/17	14/17	10/9	11/12	17/20	23/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	3.04248561649161e-05
20	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/99	22/99	21/22	2.83904595131456e-05
21	15/16	15/15	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	6/9.3	22/99	21/22	2.33109608069662e-05
22	15/16	15/16	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	14/15.2	6/7	6/9.3	22/99	21/22	2.24170054522074e-05
23	15/16	15/16	15/17	17/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	13/15.2	6/7	9.3/9.3	22/99	21/22	1.97554139375429e-05
24	15/16	15/15	15/17	14/17	10/9	11/12	17/20	19/23	10/13	14/15	29/99	27/29	99/99	15/17	14/14	14/15.2	6/7	9.3/9.3	22/99	21/22	1.69674304618174e-05

Figure 21: The figure shows the ranked table of deconvoluted genotype profiles for the unknown major contributor, when conditioning on the suspect profile. The table is ranked with respect to the posterior probability of different combined genotype profiles. The top ranked combined genotype profile is an outlier from the others which indicates that it is possible to extract the unknown profile (from figure 9 we see that this is a correct extraction).

5. Database search:



Model specification

Evidence(s)

☑ evid1

Contributor(s) under Hp:
(DB-reference already included)

#unknowns (Hp): 1

Contributor(s) under Hd:

#unknowns (Hd): 2

Continuous Model Parameters

Probability of Dropin: 0
fst-correction: 0

Qualitative Model Parameters

Probability of Dropin: 0.05
fst-correction: 0.02

Advanced Parameters

☑ Q-assignation
Detection threshold: 150
Stutter ratio (xi): 0
Dropin peak height hyperparam (lambda): 0

Database(s) to search

databaseESX17

Data selection

Loci:	evid1
D3S1358	☑
VWA	☑
D16S539	☑
D2S1338	☑
D8S1179	☑
D21S11	☑
D18S51	☑
D19S433	☑
TH01	☑
FGA	☑

Figure 22: The figure shows the GUI page of the model specification for doing database search on the database file "databaseESX17". Our model assumes no "(n-1)-stutter", no allele drop-in and no theta-correction.

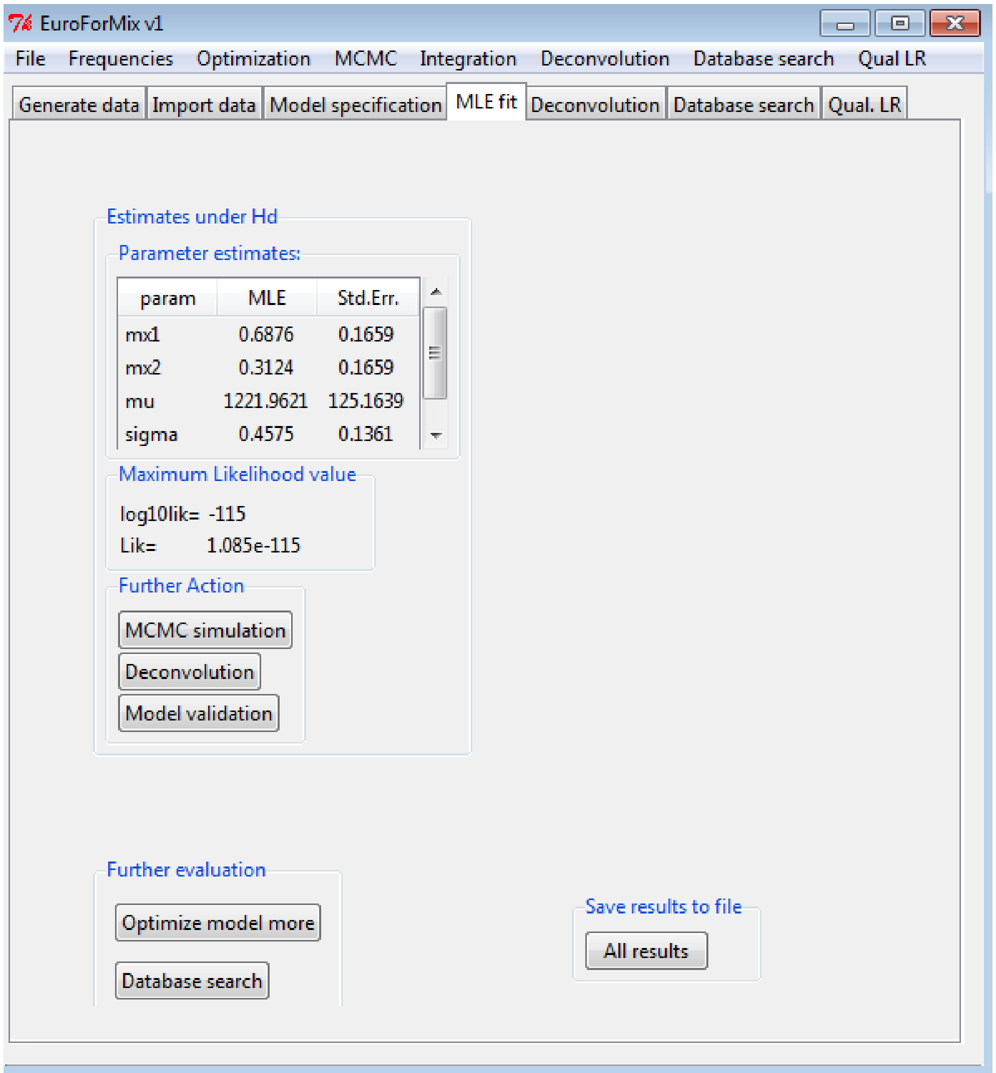
- Description:

  ○ The 'Database search' is very similar as the Weight-of-Evidence (see Figure 22) with the only difference in that each individual in the reference-database is assumed as a contributor in the hypothesis Hp. For each individual 'j' in reference-database we calculate a LR-value LRj.

    - The user may choose between using peak heights in a ‘Continuous LR’ (**Maximum Likelihood based** or **Integrated Likelihood based**)’ calculation or ignoring the peak heights in a ‘Qualitative LR’ calculation.

- When selecting ‘Continuous LR’:

    - ‘Qualitative LR’ is always calculated along with the ‘Continuous LR’ values.
    - The qualitative model assumes an allele drop-out parameter which is estimated.
    - The allele drop-in parameter in the qualitative model is set as default 0.05, but can be changed with “**Set drop-in probability for qualitative model**” under ‘Database search’ in the Toolbar.
    - No theta-correction is assumed in the qualitative model.
    - If “Continuous LR (Maximum Likelihood based)” calculation is used, the optimized parameters under the Hd -hypothesis are first shown (see Figure 23).



param	MLE	Std.Err.
mx1	0.6876	0.1659
mx2	0.3124	0.1659
mu	1221.9621	125.1639
sigma	0.4575	0.1361

log10lik= -115
Lik= 1.085e-115

Figure 23: The figure shows the optimized parameters (i.e. the MLE fit) for the continuous model under Hd (with specifications as given in Figure 22). The fitted model has the same “Further Action”

possibilities as for "Weight-of-Evidence" and "Deconvolution". The user must push "**Database search**" for doing the actual database searching.

- When selecting 'Qualitative LR':
  - The "**Set drop-in probability for qualitative model**" under 'Database search' in the Toolbar is ignored.
  - The qualitative model assumes an allele drop-out parameter which is estimated.
  - The 'Continuous LR' calculation is ignored.

- Note:
  - The 'Continuous LR' calculation is based on the **continuous model** as given in the vignette and hence may handle allele drop-in, drop-out and stutter.
  - Continuous LR (Integrated Likelihood based) is not possible to use for replicates.
  - The reason for showing the MLE fitted parameters under Hd (see Figure 23) for "Continuous LR (Maximum Likelihood based)" calculation is that the user should have the possibility to check if the parameter estimates under Hd seems reasonable so he can go back and change the model specification.

- **Table** (see Figure 24):
  - '**Reference name'** is name of individuals given in the reference-database.

  - The table shows the ranked individuals in the database due to the continuous LR values (**contLR**), qualitative LR values (**qualLR**), number of matching alleles (**MAC**) or number of evaluating loci (**nLocs**).

  - **qual.LR** (Qualitative LR (semi-continuous model))
    - Parameter for dropout probability is based on the median of 2000 samples from the 'distribution of dropout-probability'.
      - Number of required samples may be changed under 'Qual LR' in toolbar.
    - For multiple evidences, the mean of the median is used as the dropout probability parameter.
    - Assumes drop-in probability 0.05 as default. Can be changed under 'Database search' in toolbar.
    - Assumes no theta-correction.

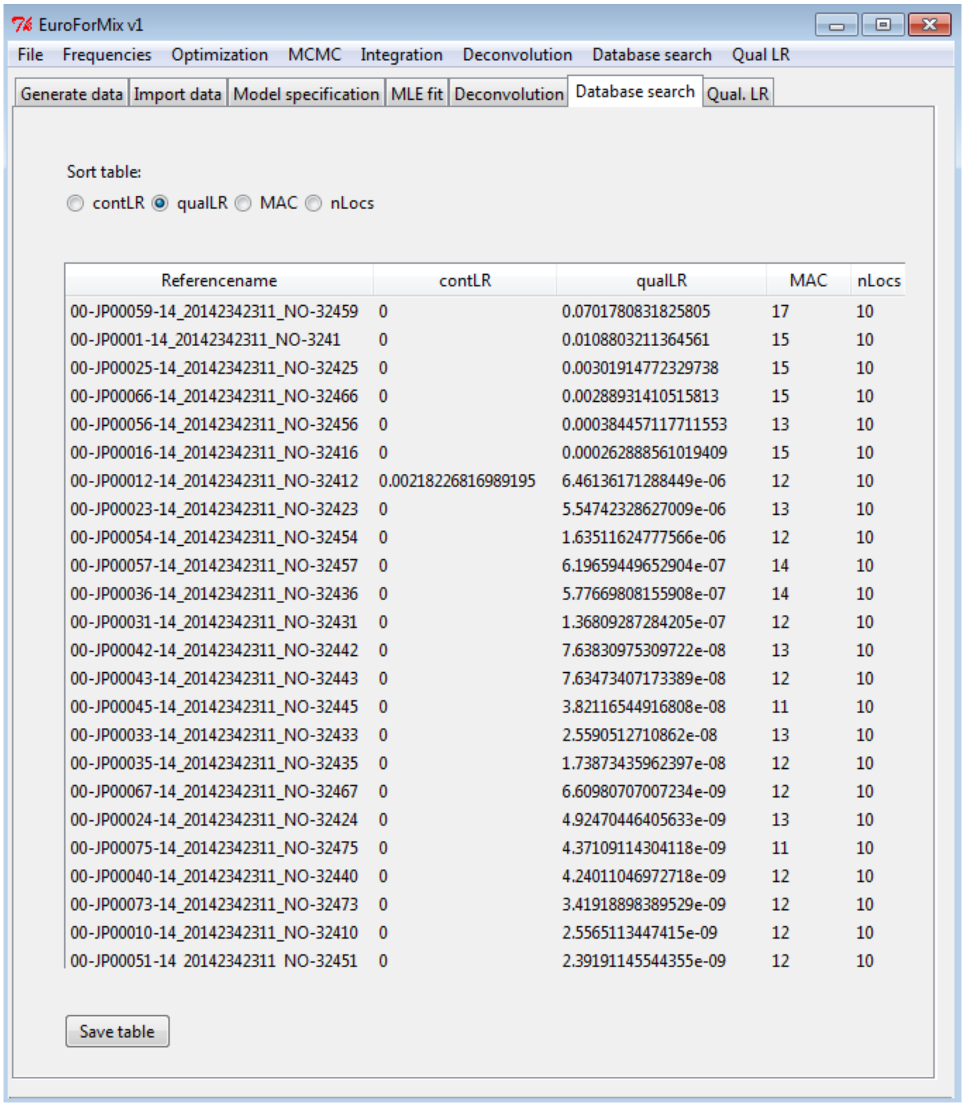
  - **MAC** (Matching allele counter) is number of alleles in the reference-profile which matches the evidence.
    - Note: MAC is summed over the considered evidences.

  - **nLocs** is number of loci in the reference-profile which are used to calculate the contLR,qualLR and MAC.
    - Note: Some references in the database may be missing loci which are presented in the evaluated evidence.

- Note:
  - Maximum number of elements to view a ‘Database search’ result table is 10000. This can be changed under ‘Database search’ in toolbar.
  - Putting fst>0 may be very time-consuming since we require that individual ‘j’ is a known non-contributor under Hd, and hence Hd is calculated for each individual in database.
  - If no allele drop-in is assumed under the continuous model, **cont.LR** is not calculated for the non-fitting individuals in the database.

- **Save table:**

  - The full table will be exported to a tabulator-separated text-file.



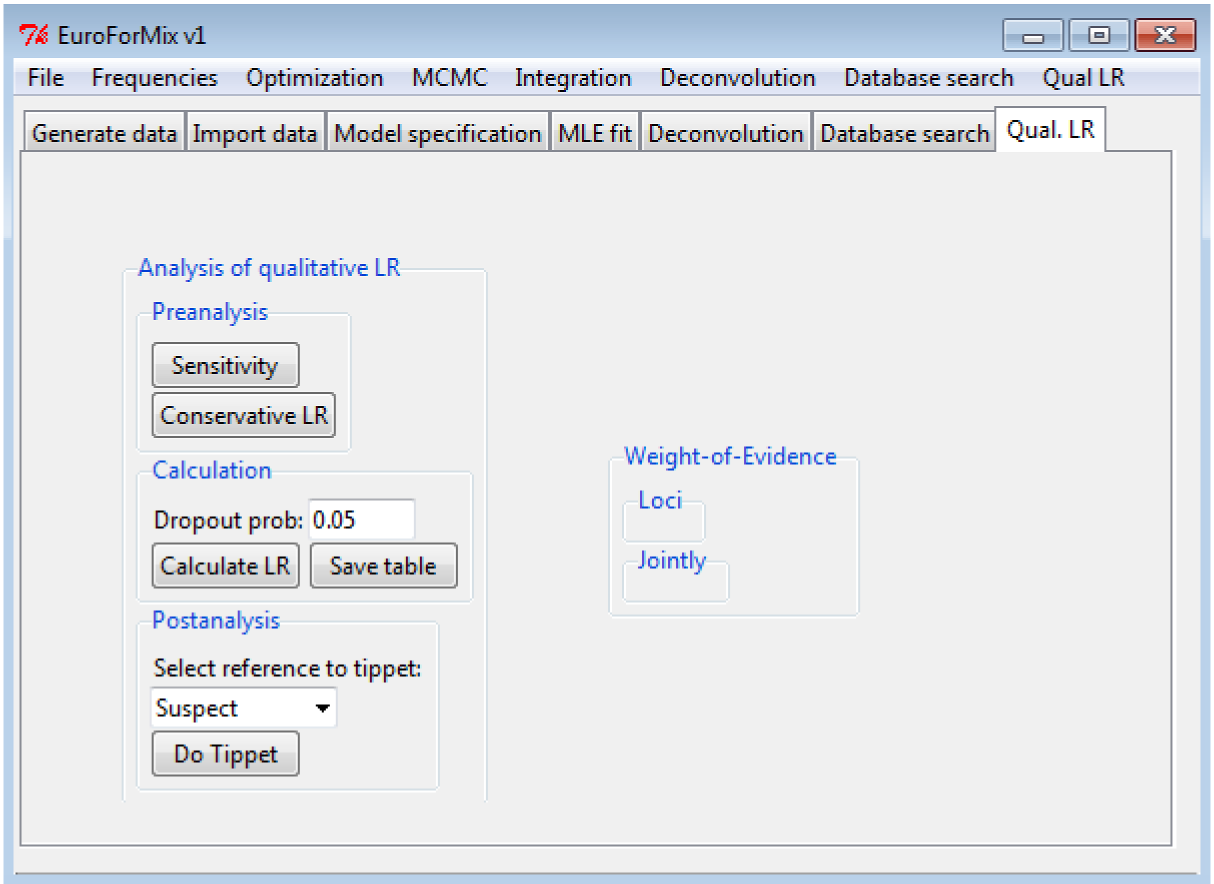
Sort table:

○ contLR ● qualLR ○ MAC ○ nLocs

Referencename	contLR	qualLR	MAC	nLocs
00-JP00059-14_20142342311_NO-32459	0	0.0701780831825805	17	10
00-JP0001-14_20142342311_NO-3241	0	0.0108803211364561	15	10
00-JP00025-14_20142342311_NO-32425	0	0.00301914772329738	15	10
00-JP00066-14_20142342311_NO-32466	0	0.00288931410515813	15	10
00-JP00056-14_20142342311_NO-32456	0	0.000384457117711553	13	10
00-JP00016-14_20142342311_NO-32416	0	0.000262888561019409	15	10
00-JP00012-14_20142342311_NO-32412	0.00218226816989195	6.46136171288449e-06	12	10
00-JP00023-14_20142342311_NO-32423	0	5.54742328627009e-06	13	10
00-JP00054-14_20142342311_NO-32454	0	1.63511624777566e-06	12	10
00-JP00057-14_20142342311_NO-32457	0	6.19659449652904e-07	14	10
00-JP00036-14_20142342311_NO-32436	0	5.77669808155908e-07	14	10
00-JP00031-14_20142342311_NO-32431	0	1.36809287284205e-07	12	10
00-JP00042-14_20142342311_NO-32442	0	7.63830975309722e-08	13	10
00-JP00043-14_20142342311_NO-32443	0	7.63473407173389e-08	12	10
00-JP00045-14_20142342311_NO-32445	0	3.82116544916808e-08	11	10
00-JP00033-14_20142342311_NO-32433	0	2.5590512710862e-08	13	10
00-JP00035-14_20142342311_NO-32435	0	1.73873435962397e-08	12	10
00-JP00067-14_20142342311_NO-32467	0	6.60980707007234e-09	12	10
00-JP00024-14_20142342311_NO-32424	0	4.92470446405633e-09	13	10
00-JP00075-14_20142342311_NO-32475	0	4.37109114304118e-09	11	10
00-JP00040-14_20142342311_NO-32440	0	4.24011046972718e-09	12	10
00-JP00073-14_20142342311_NO-32473	0	3.41918898389529e-09	12	10
00-JP00010-14_20142342311_NO-32410	0	2.5565113447415e-09	12	10
00-JP00051-14 20142342311 NO-32451	0	2.39191145544355e-09	12	10

Figure 24: The figure shows the table from the database search with specifications as given in Figure 22 based on ‘**Continuous LR’ (Maximum Likelihood based)”** calculations. The references are sorted due to the qualitative LR’s (which assumes allele drop-out probability 0.08 and allele drop-in probability 0.05).

6. Qual. LR:



Dropout prob: 0.05

Select reference to tippet: Suspect

Figure 25: The figure shows the GUI page where the weight-of-evidence evaluation based on the qualitative model is done.

- Description:
  - This module samples from the distribution of the '*allele drop-out probability given number of observed alleles*' to evaluate the qualitative LR automatically. Also sensitivity plot as a function of allele-dropout probability and random man tippet analysis is implemented.

PREANALYSIS

- **Sensitivity:**
  - Plots the log10LR as a function of allele-dropout probability (see Figure 26).
    - The upper probability range and number of ticks can be changed under 'Qual LR' in the toolbar.
  - Note:
    - Lower range in sensitivity is 1e-6 (something small).

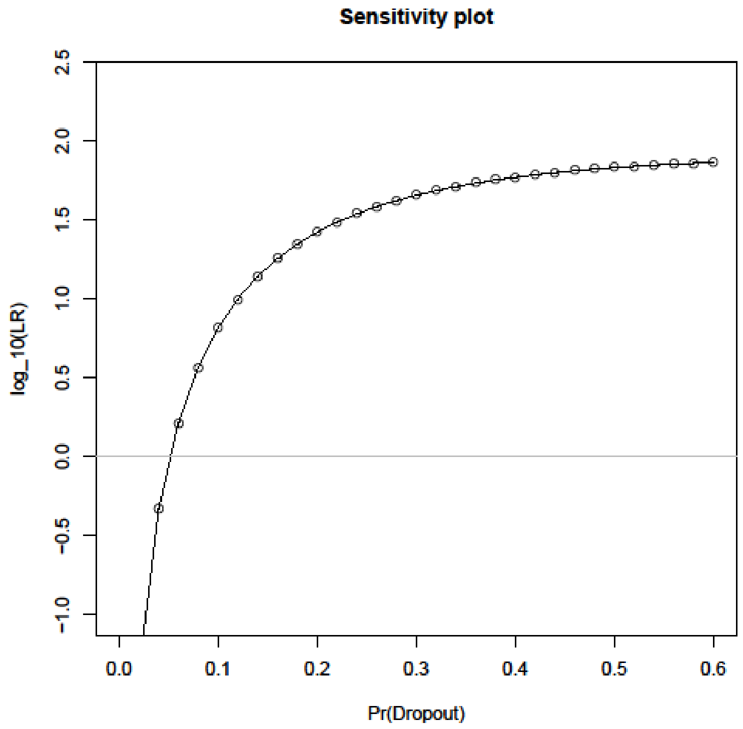
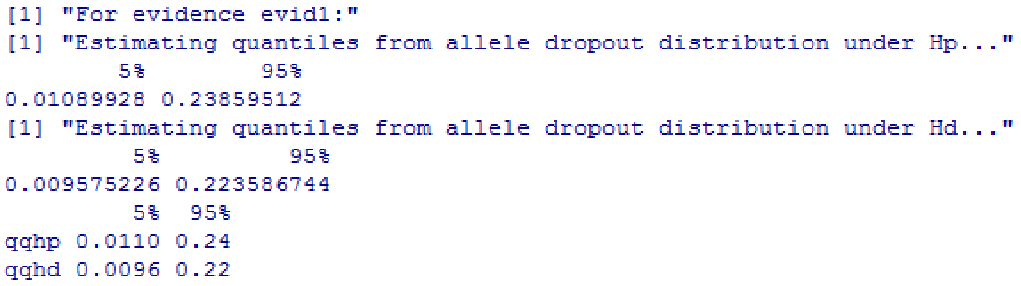


Figure 26: The figure shows the plot of Weight-of-evidence (Likelihood Ratio) as a function of allele drop-out probability.

- **Conservative LR:**

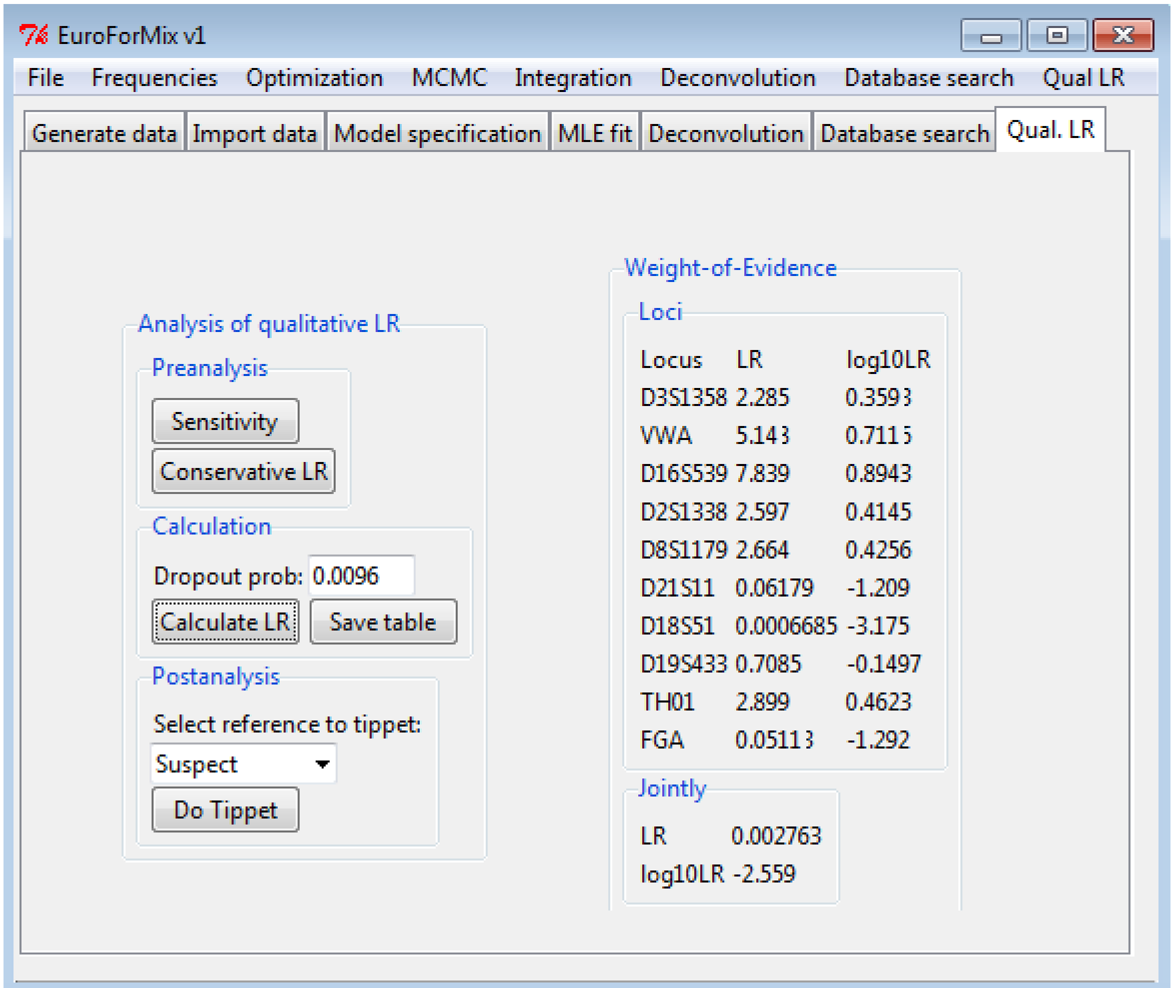
  - By sampling from the "*allele drop-out probability given number of observed alleles in the evidence"*- distribution for the hypothesis Hp and Hd, the most 'conservative' LR (i.e. smallest) is automatically calculated and printed (see Figure 27 and Figure 28).
    - The most "conservative" LR is found by following:
      - Take out the "alpha" and "1-alpha"-quantiles from the simulated 'allele-dropout probability distribution' under both Hp and Hd.
      - The quantile (under both Hp and Hd) which gives the lowest LR is the "conservative LR".
    - The significance level "alpha" is given 0.05 as default.
      - This can be changed under 'Qual LR' in the toolbar.
    - The number of required samples from the 'allele-dropout probability distribution' is given 2000 as default.
      - This can be changed under 'Qual LR' in the toolbar.
    - Note: If no samples are accepted from the allele-dropout probability distribution', an error-message is provided to the user.

  - When more evidence samples are imported, the most 'conservative LR' over all samples is considered.
    - The dropout probability quantiles are estimated for each of the evidence samples.



```
[1] "For evidence evid1:"
[1] "Estimating quantiles from allele dropout distribution under Hp..."
        5%         95%
0.01089928 0.23859512
[1] "Estimating quantiles from allele dropout distribution under Hd..."
         5%          95%
0.009575226 0.223586744
         5%  95%
qqhp 0.0110 0.24
qqhd 0.0096 0.22
```

Figure 27: The plot shows the sampled 5% and 95% quantiles of the distribution of the '*allele drop-out probability given number of observed alleles*'.



Dropout prob: 0.0096

Select reference to tippet: Suspect

Locus	LR	log10LR
D3S1358	2.285	0.3593
VWA	5.143	0.7113
D16S539	7.839	0.8943
D2S1338	2.597	0.4145
D8S1179	2.664	0.4256
D21S11	0.06179	-1.209
D18S51	0.0006685	-3.175
D19S433	0.7085	-0.1497
TH01	2.899	0.4623
FGA	0.05113	-1.292

LR 0.002763

log10LR -2.559

Figure 28: The plot shows the conservative Weight-of-Evidence values (Likelihood Ratios) after pushing "**Conservative LR**". The most conservative estimated allele drop-out probability-quantile from Figure 27 was the 5% quantile under Hd which gave 0.0096. Hence the table in this plot shows the LR inserted for this value.

CALCULATION

- **Dropout prob:**

  - The user may specify the assumed number of allele dropout-probability.

- **Calculate LR**

  - Instantly calculates the LR for the given user-specified allele dropout probability in "**Dropout prob**".

- **Save table:**

  - Saves the weight-of-evidence calculated LR results to a selected file.

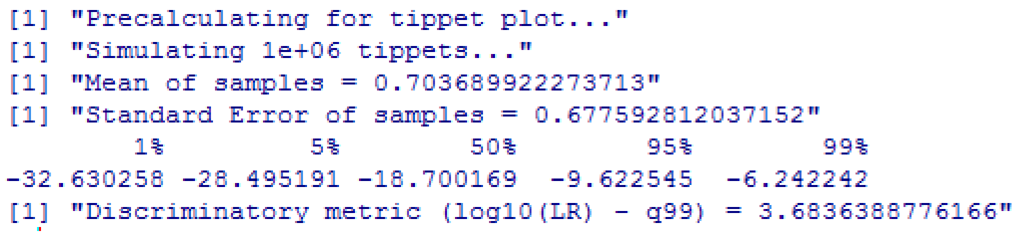
POSTANALYSIS

- **Selection of reference to tippet:**

  - A drop-down list of references which are conditioned under Hp but not under Hd.

- **Do Tippet:**

  - Random tippet samples are provided by replacing the selected reference (under the drop-down list in the hypothesis Hp) with a random individual from the population and then calculate his LR. A vast amount (default is 1e6) of random tippets is simulated to determine the tippet-distribution.
    - The mean, standard errors of LR and log10LR-quantiles (1%, 5%, 50%, 95%, 99%) are printed out to terminal (see Figure 29).
    - A plot of the cumulative distribution of log10LR will be shown (see Figure 30).
    - Number of tippets can be changed under 'Qual LR' in the toolbar.
  - If weight-of-evidence has been calculated:
    - The reporting LR for the "tipped individual" is superimposed as a blue line to the plot (see Figure 30).
    - The discriminatory metric (log10LR-q99%) is printed out to terminal (see Figure 29).
  - Note: Precalculations are always done previous to the tippet-sampling, therefore the number of tippets are only limited to make the plot.



```
[1] "Precalculating for tippet plot..."
[1] "Simulating 1e+06 tippets..."
[1] "Mean of samples = 0.703689922273713"
[1] "Standard Error of samples = 0.677592812037152"
        1%         5%        50%        95%        99% 
-32.630258 -28.495191 -18.700169  -9.622545  -6.242242 
[1] "Discriminatory metric (log10(LR) - q99) = 3.6836388776166"
```

Figure 29: The plot shows the printed tippet information to the terminal when replacing the "Suspect" in hypothesis Hp with a random man (a tippet). Number of tippets simulated, mean and standard errors of LR and log10LR-quantiles (1%, 5%, 50%, 95%, 99%) are printed out to terminal (see Figure 29). Also the discriminatory metric, the distance between the observed log10LR for the suspect and log10LR-99%-tippet-quantile is given.

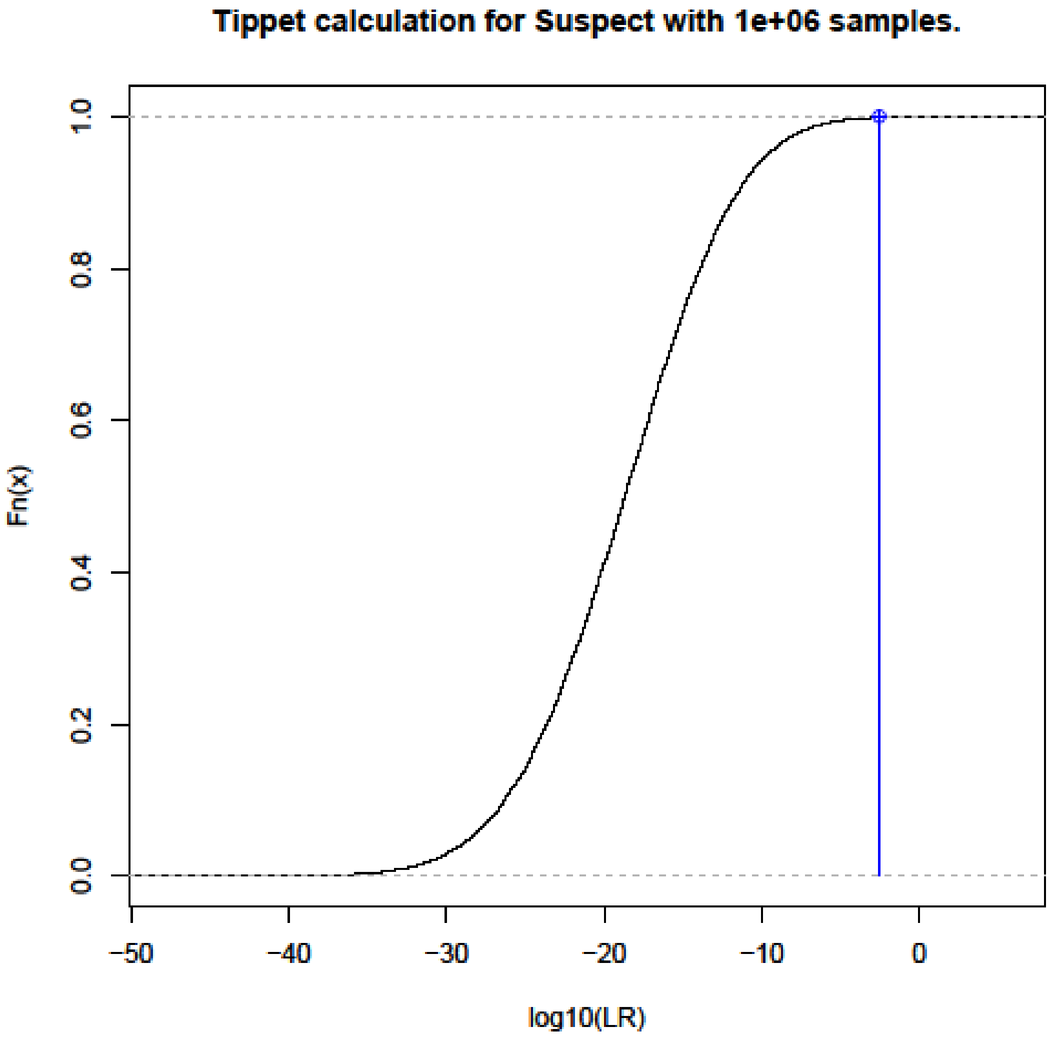
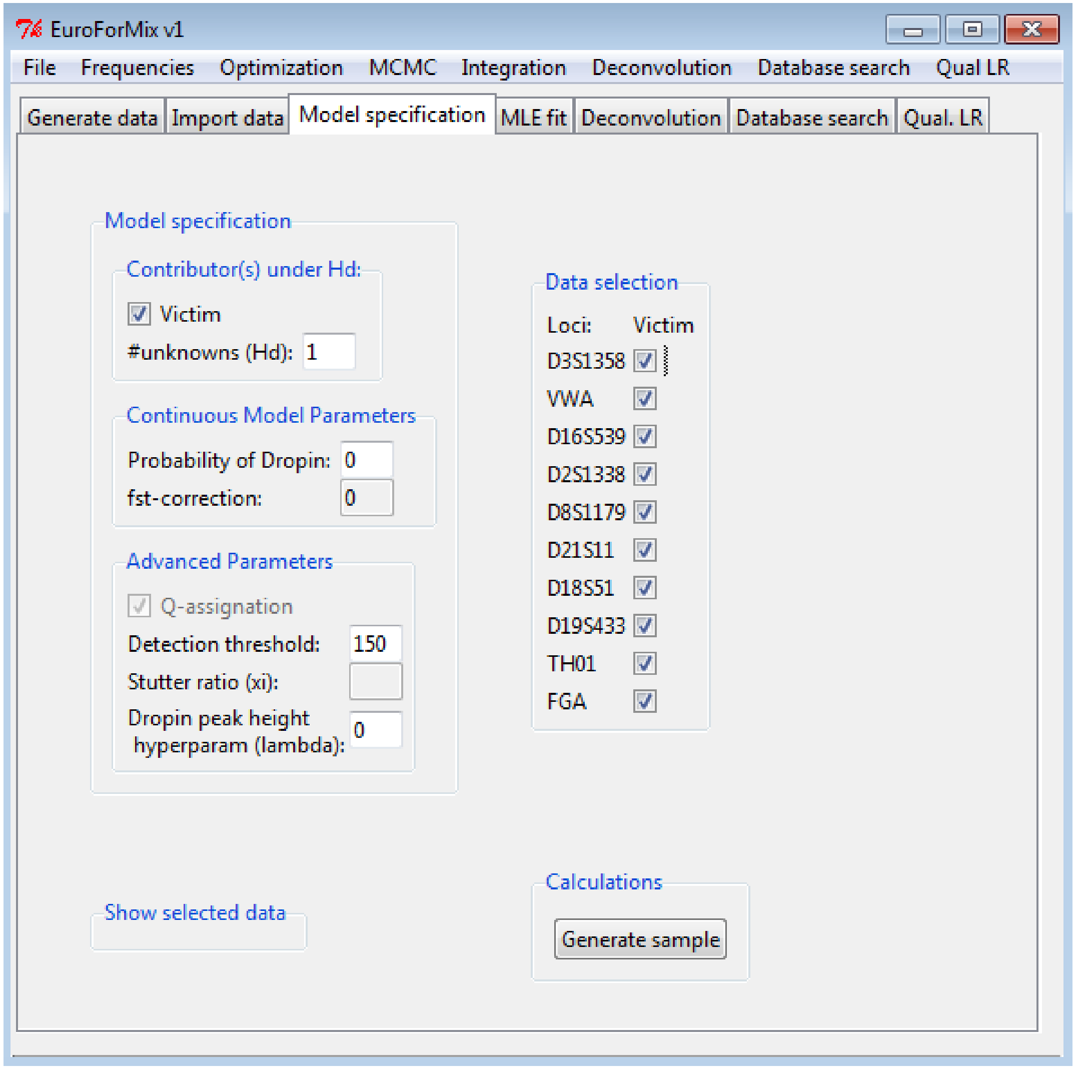


Figure 30: The figure shows a cumulative distribution of 1000000 log10LR tippets, where each tippet is based on replacing the "Suspect" in hypothesis Hp with a random man from the population. The reporting LR for the "tipped individual" (i.e. "Suspect in this case) is superimposed as a blue line to the plot.

7. Generate data:



Model specification

Contributor(s) under Hd:

☑ Victim

#unknowns (Hd): 1

Continuous Model Parameters

Probability of Dropin: 0

fst-correction: 0

Advanced Parameters

☑ Q-assignation

Detection threshold: 150

Stutter ratio (xi):

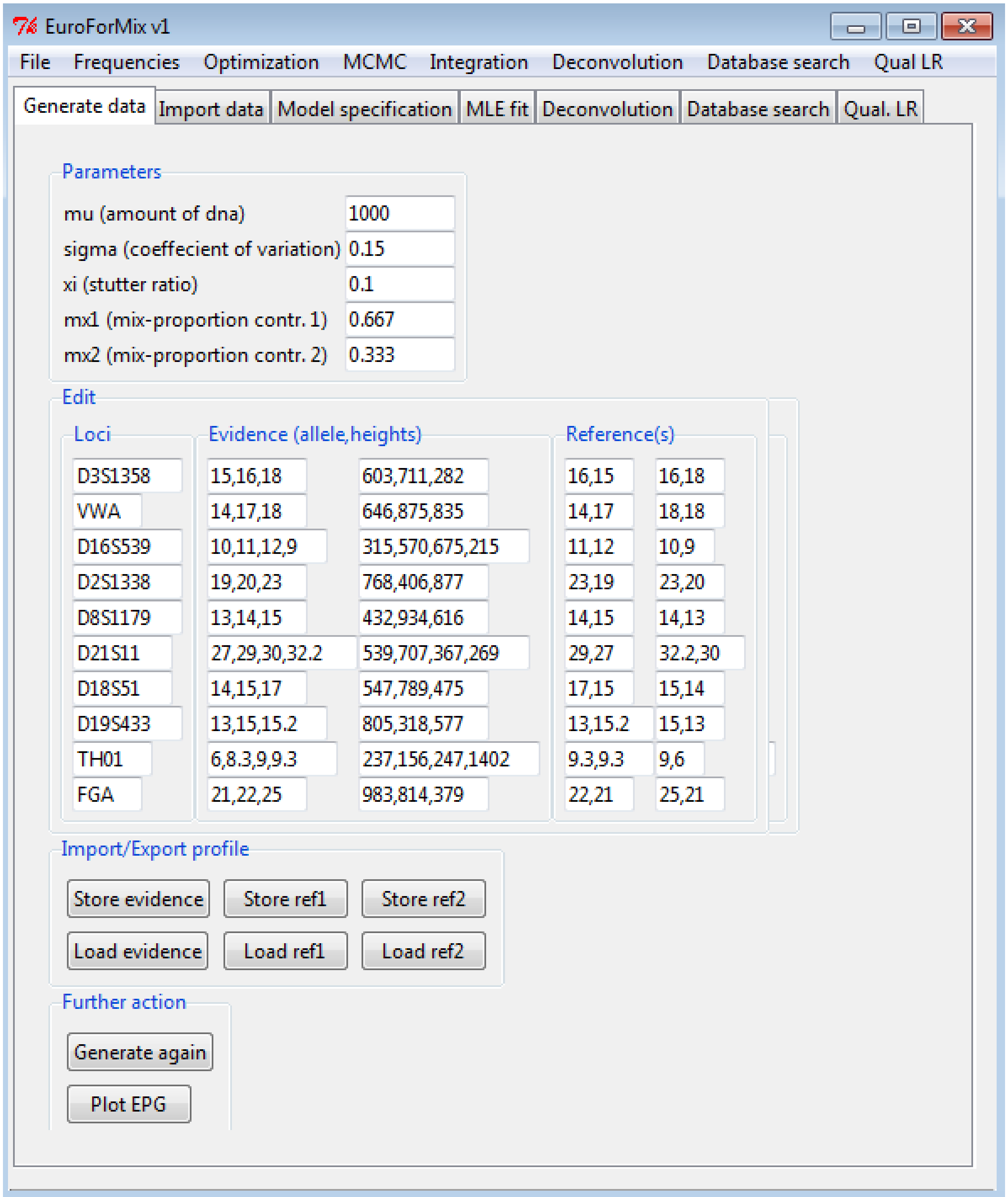
Dropin peak height hyperparam (lambda): 0

Data selection

Loci:	Victim
D3S1358	☑
VWA	☑
D16S539	☑
D2S1338	☑
D8S1179	☑
D21S11	☑
D18S51	☑
D19S433	☑
TH01	☑
FGA	☑

Figure 31: The figure shows the Model specification GUI page for generating allele with corresponding peak heights from the continuous model for a given specified model. From here we will generate data which are contributed from a known Victim profile and an unknown individual. We assume a detection threshold of 150 rfu and no allele drop-in is considered.

- Description:
    - Generates alleles using the population frequencies and simulates peak heights for a specified hypothesis (see Figure 31) using the continuous model.
    - The generation may simulate allele-dropout, drop-in (with a peak height model) and stutter (see Figure 32).
        - Allele-dropout is indirectly simulated by falling below the defined threshold.



Parameters

mu (amount of dna)	1000
sigma (coeffecient of variation)	0.15
xi (stutter ratio)	0.1
mx1 (mix-proportion contr. 1)	0.667
mx2 (mix-proportion contr. 2)	0.333

Edit

Loci	Evidence (allele,heights)		Reference(s)	
D3S1358	15,16,18	603,711,282	16,15	16,18
VWA	14,17,18	646,875,835	14,17	18,18
D16S539	10,11,12,9	315,570,675,215	11,12	10,9
D2S1338	19,20,23	768,406,877	23,19	23,20
D8S1179	13,14,15	432,934,616	14,15	14,13
D21S11	27,29,30,32.2	539,707,367,269	29,27	32.2,30
D18S51	14,15,17	547,789,475	17,15	15,14
D19S433	13,15,15.2	805,318,577	13,15.2	15,13
TH01	6,8.3,9,9.3	237,156,247,1402	9.3,9.3	9,6
FGA	21,22,25	983,814,379	22,21	25,21

Figure 32: The figure shows the Generate data GUI page which shows the generated alleles and corresponding peak heights (under **Evidence**) for the given selected set of parameters under **Parameters**. The true contributors are given under **Reference(s)**.

- **Parameters**:
  - **mu**: amount of DNA
  - **sigma:** coefficient of variance
  - **xi:** stutter ratio
  - **mx=(mx1,…, mxC):** mixture proportion for contributor 1,..,C.
    - Note: **mx** will be normalized if it's not already.

- **Edit**:
  - **Loci**: Loci name of the population frequency used to generate the dataset.
  - **Evidence**: The allele information is given in the left column while the peak height information is given in the right column. Each element **needs to be** separated with ",".
  - **Reference**: The alleles of the true contributors to the generate evidence is sequentially shown in each column.
  - All the loci names, evidence-allele and heights and reference-alleles may be edited before storing (See Figure 32).

- **Import/Export**:
  - **Save data**:
    - Stores the generated (and possible edited) evidence- or reference-profile to a file.
    - Extension .csv added automatically.

  - **Load data:**
    - Loads profiles from file into the selected entries (evidence or reference).
      - This is useful for generating random evidence samples where loaded references are conditioned on.
    - Note:
      - If any locus is missing from the loaded evidence or reference file, the edit-cell will be empty.
      - The order of the loci in the file does not matter.

- **Further action:**
  - Generate again**:** Make a new simulation of the evidence sample using the selected values of the parameters under **Parameters**.
  - Plot EPG: Plots the generated (and possible edited) evidence in a EPG-plot.
    - It will use the "kit" selected under "Import Data"-page.
    - See ?plotEPG to see which kit-formats that are supported in the EPG.

(C) To be implemented in a future version:

- Label the alleles of the selected references to the EPG-plot.
- Warning if exp(lik)=0 when lik>-Inf (happens for INT calculations)
- Empty loci will not be removed when imported to the software. They will be considered as a full dropped out loci in the evaluation.