Manual for EuroForMix v2.1

Author: Øyvind Bleka <oyvble.at.hotmail.com> and Peter Gill
Date: 26-02-2019
Version used: euroformix_2.1.0.zip
Data used "*Data set 0: Tutorialdata*" from www.euroformix.com/datasets
See webpage www.euroformix.com for details.

(A) Installation and running program:

1) Run R (>=3.1.0) in Windows (http://cran.r-project.org/).

2) Copy and run these commands in the R-software to download and install the required packages:
```
install.packages(c('gWidgetstcltk','forensim','cubature','XML','RCurl'))
```

3) To download and install *euroformix*:

   a. Download newest "euroformix_version.zip"-file and install it manually (using the original zip-file) in the the R-software using *Install package(s) from local zip files...*

4) Run these commands in the R-software to start the GUI:
```
library(euroformix)
efm()
```

It is strongly recommended to create a startup icon on your desktop because you will only need to carry out the above procedure once. To proceed, first close down R without saving the workspace.

(B) Creating a startup icon for EuroForMix on your desktop:

1) Launch R again

2) Run this code within your preferred R-version:
path <- file.path(Sys.getenv("USERPROFILE"),"Desktop",fsep="\\")
setwd(path) #If 'Desktop' is not the proper desktop name you may need to change it.
.First = function(x) {require(euroformix);efm()};save.image(".Rdata");

3) Creating the shortcut:
- Create a copy (CTRL + C and then CTRL + V) of the R icon on the desktop.
- Name the copy "EuroForMix" or something arbitrary.
- Right-click on icon. Select "Properties" and then go to the "Shortcut" tab (named something else for other languages).
- Remove the text within "Start in", click "Apply" and then close the window.
- You have now created an startup icon for EuroForMix.

(C) GUI

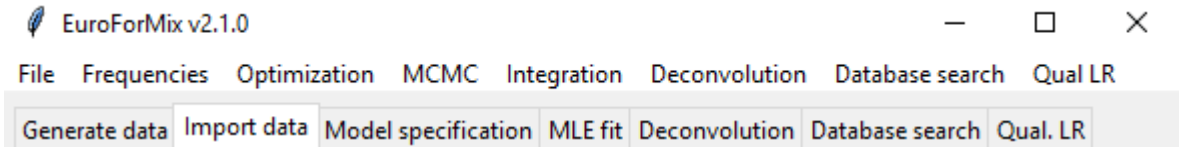


Figure 1: The Toolbar and Tabs for different steps of the analysis.

Sections:

Part (D) contains Mathematical details.

1- Toolbar

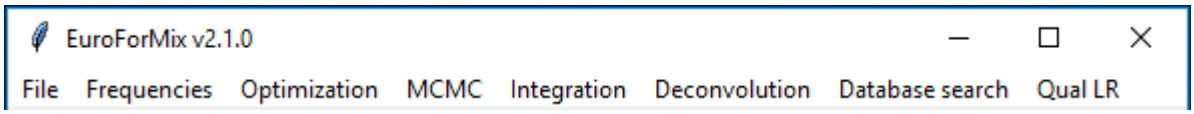


Figure 2: The Toolbar contains configurations and advanced model parameters for different kinds of analyses.

- File
  - **Set directory**: The user may select the working directory for the program.
  - **Open project**: The user may open an earlier project which is saved in a file in 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 opened at any time in the program.
    - Saving a project has the following advantages:
      - Large reference databases are stored efficiently (the required space for the database is drastically reduced).
        - *It is strongly recommended to use "save project" because it saves a lot of time if you need to re-evaluate the analysis. All the data are conveniently stored and can be reloaded instantly.*
  - **Settings:**

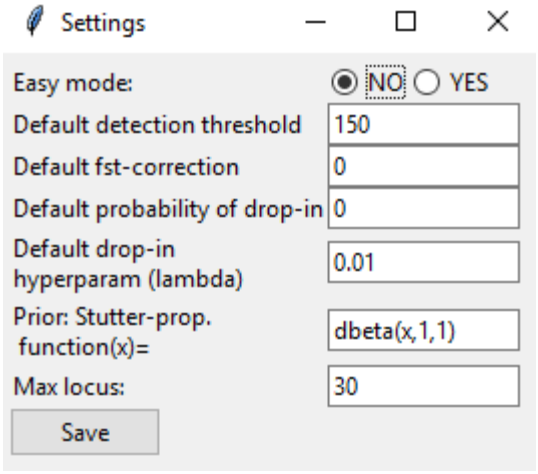


Figure 3: The setting window under "File->Settings" used to set advanced model parameters.

- **Easy mode**: NO gives all functionalities as before. YES introduces a reporting mode for calculating the Likelihood Ratio (disables buttons to guide the reporter).

- *Easy mode is strongly recommended for routine casework because it guides the user to carry out interpretation using a recommended path.*

- **Default Model Parameters**: The following parameters can only be selected here.

  - **Default detection threshold:** [0,->)
    - The limit of detection (LOD) threshold of required allele peak heights. Used to define whether an allele is present in the evidence or not.
    - 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.
    - Not considered if no peak heights are provided in the evidence.

  - **Default fst-correction**: [0,1]
    - Assumed co-ancestry parameter assigned in the genotype probability for each contributor in the hypotheses. See euroformix paper for more details.

  - **Default probability of drop-in:** [0,1]
    - Assumed probability of a random allele drop-in to the evidence at a given locus. See euroformix paper for more details.
    - If **Probability of drop-in'**>0 when consider 'Quantitative LR' the user needs to specify **Drop-in peak height hyperparam**>0.

  - **Default drop-in hyperparam (lambda)**: (0,1]
    - Only used for 'Quantitative LR' if **lambda** >0.
    - Assumed hyper-parameter to model the peak height of the dropped in allele caused by a 'random allele drop-in' (see **Figure 4**).
    - A default of 0.01 is suggested if the user has no data (although it is recommended that the user calculates this parameter from his/her own data). The parameter is calculated automatically when the user imports a dataset with the "Fit dropin data" button.

  - **Prior: Stutter-prop. function(x)= dbeta(x,1,1)**:
    - A prior density function for the stutter proportion parameter **Stutter-prop.**
    - The user can design his own density function over [0,1]
      - Default is a flat prior (specified through the beta distribution)

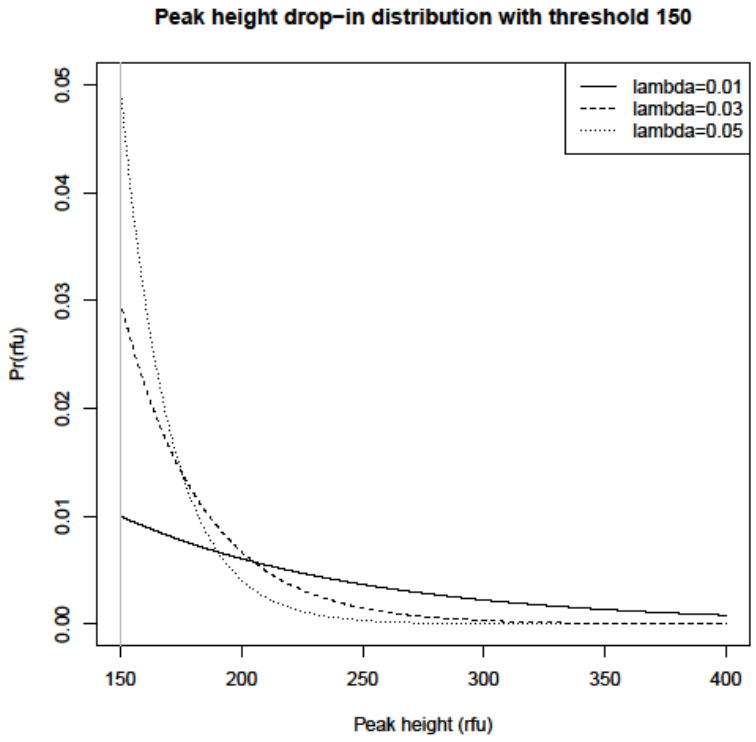


Figure 4: The figure shows the allele peak height drop-in distribution for three values of the lambda hyper-parameter. The distribution is expontial (RFU-threshold|lambda) (i.e. shifted exponential).

- **Max locus**: The maximum number of markers to present in the GUI (the "per-marker" info at "Model specification" and "MLE fit" pages). This is due to that the vertical resolution of the screen limits the number of markers which can be shown.

- **Save**: Click to save settings permanently*: The values are stored and loaded in the next session. This means that you can close EuroForMix and open it again without the loss of values.
    - *A loaded project restores settings used in the corresponding project.

- **Quit project**: When button is pushed, the user is given a 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, i.e. not in the frequency database, from imported files are found, these are assigned as freq0.
    - If N=0 (this is default), freq0 is equal to the minimum imported allele frequency.
    - If N>0, freq0='5/(2N)'.

- New alleles are updated to the population frequency database:
    - When a reference database is imported.
    - When interpretations are carried out ('Generate sample', Deconvolution, Weight-of-Evidence or 'Database search')
    - The allele frequencies used for an analysis will be presented in the LR report.
    - Frequencies will only be normalized when importing reference databases.
        - **WARNING**: Allele frequencies are no longer normalized (but the frequencies in used in the calculations still add to one).

- **Set minimum frequencies:** The user can specify the allele frequency for new alleles. See details above.

- **Set number of wildcards in false positive match**: The user may specify the 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 valid maximum optimizations to be obtained (hence the name "startpoints" is not precise in that perspective). Multiple optimizations are recommended to be obtained in order for the optimizer routine to ensure that the global maximum point is attained for the Maximum Likelihood Estimator (MLE). Default is 4.
    - An optimization is counted as valid if a concave maximum point has been obtained.
    - If the number of unknown contributors are more than 2, the optimization routine is carried out with parallelization. Then each optimization is carried out separately with a CPU core thread. If the number of available threads is less than the number of startpoints, the number of optimizations is set equal the number of available CPU core threads.

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

- **Set max number of iterations**: The user may set the maximum number of iterations carried out in the optimization. Default is 100.
    - NB: Setting this value too low may cause convergence problems.

- 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 1000. The program will carry out parallelization where this number will be carried out for each of the CPU core threads. The samples are combined afterwards. We recommend in total 5000 samples wherever possible – computation time is a limitation for samples with more than 3 unknowns.
        - Note that this is the number of samples used for the "LR sensitivity" analysis.

    - **Set variance of randomizer**: The user may set the variance parameter scalar used in the 'Markov Chain Monte Carlo (MCMC) random walk Metropolis'. Default is 10.
        - Note that this value should be tweaked so that the acceptance rate of the sampler is around ~0.2 (at least 0.05-0.3) (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 euroformix paper for details. Default is 0.1.

    - **Set maximum number of evaluations:** The user may set the maximum number of evaluations for calculating the integral. This number will override the relative error requirement if selected greater than 0. Default is 0.

    - **Set maximum of P.H.expectation**: The user may set upper limit of expected peak heights parameter (mu). See euroformix paper for details. Default is 20000.

    - **Set maximum of P.H.variability**: The user may set upper limit of the coefficient of variation of peak heights parameter (sigma). See euroformix paper for details. Default is 0.9.
        - **WARNING**: The user may experience non-convergent LR values if this value is too high. The user can then reduce it to for instance 0.6 and see if the problem is fixed.

- **Set maximum of stutter proportion**: The user may set upper limit of the backward-stutter proportion parameter (xi). More details about the stutter rate is given under 'Advanced Parameters' in the Model specification section. Default is 0.5.

- **Set likelihood-scaling to avoid zero**: The user may set an offset to the likelihood function (on log scale) for the quantitative LR (Bayesian based LR). The reason is that having large amount of data causes the likelihood value to be very small which causes underflow. Default is 700.
    - Note: The user may need to change this value close to the magnitude of the maximum likelihood.

- Deconvolution

    - **Set required summed probability**: The user may set the required summed posterior genotype-probability which the Joint deconvoluted list must contain. Default is 0.99.

    - **Set max listsize:** The user may set maximum number of elements in the Joint deconvoluted list. Default is 20.
        - The greater max listsize, the more time-consuming (and memory consuming) the search-algorithm behind will be.
        - This is also the maximum number of combined genotypes shown in the console.

- 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 this 'value', the more time-consuming it will become to show the table on the screen.
        - Note that the results table from the database search shows only the top 'value'-ranked elements.

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

    - **Set number of non-contributors**: The user may specify number of random non-contributor samples in the non-contributor analysis. Default is 1000.

- 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 median for the distribution of the '*allele drop-out probability given number of observed alleles*'. Default is 2000.

    - **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.

2- Importing data

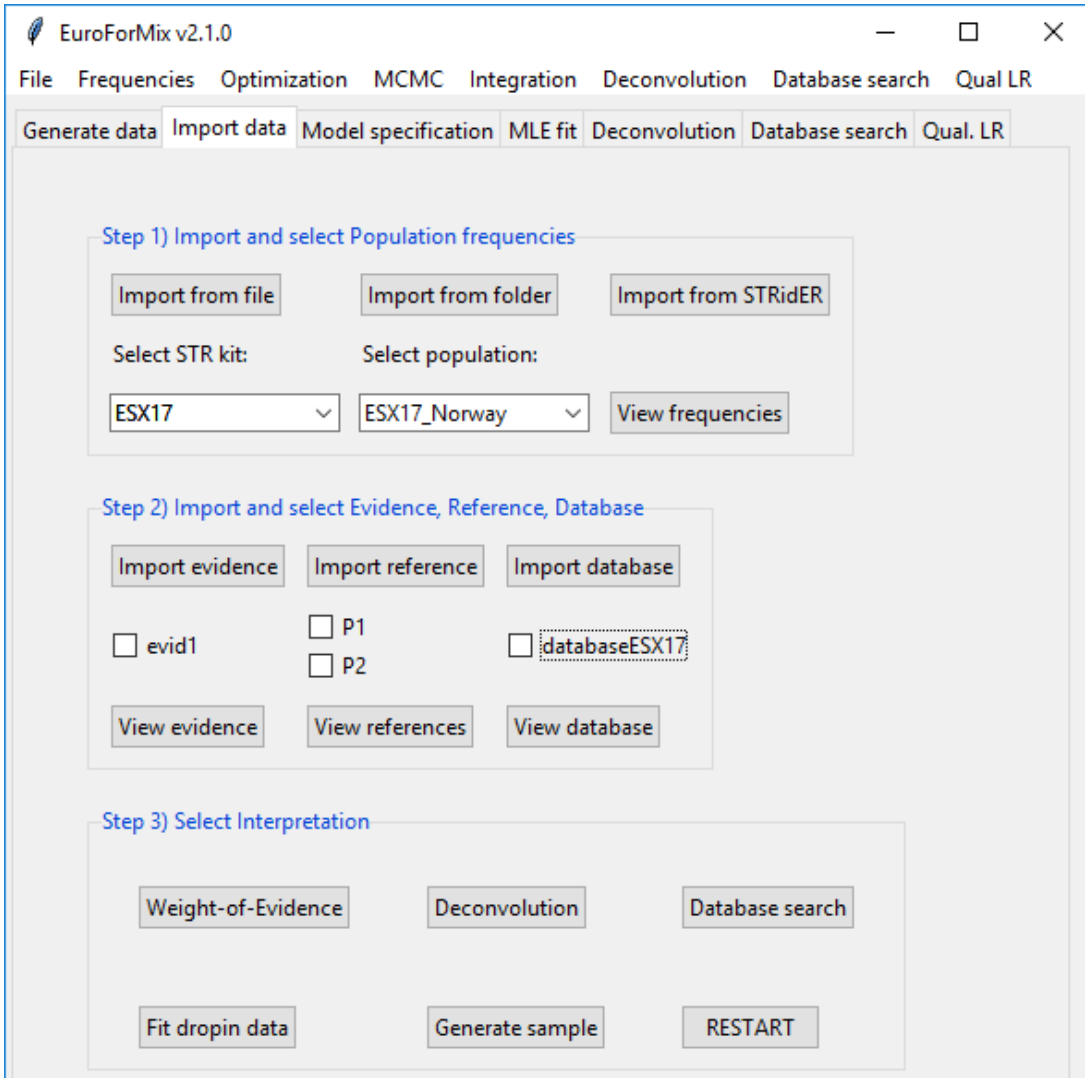


Figure 5: The figure shows the Import data page where the user can import population frequencies, evidence stains, reference profiles and reference databases.

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 X 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 X 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 R-console (see **Figure 6**). 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 Height.2 Height.3 Height.4 Height.5 Height.6   ADO UD1  X
1        evid1     AMEL        X        Y       NA       NA       NA       NA     2136     1015       NA       NA       NA       NA false  NA NA
2        evid1  D3S1358       14       15     16.0       NA       NA       NA      178     2405     1982       NA       NA       NA false  NA NA
3        evid1     TH01        6        7      9.3       NA       NA       NA      419      282     1871       NA       NA       NA false  NA NA
4        evid1   D21S11       27       29       NA       NA       NA       NA     1128     1750       NA       NA       NA       NA false  NA NA
5        evid1   D18S51       15       17       NA       NA       NA       NA      467      524       NA       NA       NA       NA false  NA NA
6        evid1 D10S1248       13       14     15.0       NA       NA       NA     1856      155     1045       NA       NA       NA false  NA NA
7        evid1  D1S1656       12       15     16.0     16.3     17.3       NA     1140      601      488      155     1877       NA false  NA NA
8        evid1  D2S1338       17       19     20.0     23.0       NA       NA      290      619      259      649       NA       NA false  NA NA
9        evid1  D16S539        9       10     11.0     12.0       NA       NA      217      312      743      619       NA       NA false  NA NA
10       evid1 D22S1045       15       16       NA       NA       NA       NA     1017      610       NA       NA       NA       NA false  NA NA
11       evid1      vWA       14       15     17.0       NA       NA       NA     1250      440     1232       NA       NA       NA false  NA NA
12       evid1  D8S1179       10       13     14.0     15.0       NA       NA      206      352      978      827       NA       NA false  NA NA
13       evid1      FGA       21       22       NA       NA       NA       NA      664      714       NA       NA       NA       NA false  NA NA
14       evid1   D2S441        9       10     11.0     14.0       NA       NA      200     3362     1168     3693       NA       NA false  NA NA
15       evid1  D12S391       18     18.3     19.0     21.0     22.0       NA      297     1446      751      171     1370       NA false  NA NA
16       evid1  D19S433       13       14     15.2       NA       NA       NA     1157      781      922       NA       NA       NA false  NA NA
17       evid1     SE33     29.2     30.2     33.2       NA       NA       NA      221      473      570       NA       NA       NA false  NA NA
```

Figure 6: The figure shows the table format in the importing evidence stain file.

- **Import from file**: The user can select a frequency file from the local system. The name of the selected file will be present under “Select population”.
  - Right click on the button to see the directory of example files (*FreqDatabases*) in the R installation location of euroformix.
  - Filename: There is no longer any requirement for the name of the frequency-files.
  - Required file format:
    - First column must contain allele-designations (header-name may be anything).
    - Other columns are frequency-information (header-name denotes the locus name and this is converted to capital letters).
  - Requirement for allele frequency values
    - The values in the file cannot be zero (instead keep the cells empty).

        - The values must sum up to one for each column.
            - Note: EuroForMix gives a warning if this is not the case.

- **Import from folder**:
    - Requires a separate folder (population-folder) with **only** frequency-files.
    - Filename: There is no longer any requirement for the name of the frequency-files.
    - Right-click on the button to automatically import the example frequency files from the installation location of euroformix (see previous).

- **Import from STRidER**:
    - This will import frequency tables found at the STRidER webpage.
    - Right-click to see the URL of where data are imported from.

- **Select STR kit:**
    - The user can at any time select the relevant kit in the drop-down menu. Here the kits are given by its *short name*.
        - Note: This will be the same names as obtained when running *getKit()* in the R-console (after loading *euroformix* package).
    - EuroForMix uses the kit info found in ~euroformix\extdata\kit.txt.
        - Note: If the relevant kit is not found, the user can contact help support or try to create a new kit.txt file.

- **Select population:**
    - The user must select one of the populations in the drop-down menu. Then corresponding allele frequencies will be used in all analyses.

- **Import Evidence/Reference** sample (see **Figures 6- 8**):

    - **Multiple** evidence or reference profiles are **allowed** in each file.
        - In evidence files:
            - "height" header is required for analysis: 'Deconvolution', 'Weight-of-Evidence' (quantitative 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.
            - Homozygote genotype may have an empty allele under 'Allele 2' (*updated from euroformix.0.6.2*). The user will get a notification if this occurs.
            - Note: The import function will not check whether number of alleles and corresponding peak heights are the same.
            - Loci without any allele-information (i.e. empty or dropped out), will also be imported.

```
[1] "Raw fil import:"
   SampleName   Marker Allele1 Allele2
1          P1  D3S1358    16.0    15.0
2          P1     TH01     9.3     9.3
3          P1   D21S11    29.0    27.0
4          P1   D18S51    17.0    15.0
5          P1 D10S1248    15.0    13.0
6          P1  D1S1656    12.0    17.3
7          P1  D2S1338    23.0    19.0
8          P1  D16S539    11.0    12.0
9          P1 D22S1045    15.0    16.0
10         P1      VWA    14.0    17.0
11         P1  D8S1179    14.0    15.0
12         P1      FGA    22.0    21.0
13         P1   D2S441    10.0    14.0
14         P1  D12S391    18.3    22.0
15         P1  D19S433    13.0    15.2
16         P1     SE33    30.2    33.2
17         P2  D3S1358    16.0    15.0
18         P2     TH01     6.0     7.0
19         P2   D21S11    29.0    35.0
20         P2   D18S51    11.0    14.0
21         P2 D10S1248    13.0    13.0
22         P2  D1S1656    15.0    16.0
23         P2  D2S1338    17.0    20.0
24         P2  D16S539     9.0    10.0
25         P2 D22S1045    15.0    15.0
26         P2      VWA    15.0    17.0
27         P2  D8S1179    10.0    13.0
28         P2      FGA    22.0    25.0
29         P2   D2S441    11.0    11.0
30         P2  D12S391    18.0    19.0
31         P2  D19S433    14.0    14.0
32         P2     SE33    27.2    29.2
```

Figure 7: The figure shows the table format for the imported reference file.

- **Import Reference Database (**see **Figure 8**):
  - Exactly same format as reference files.
  - Multiple database files may be imported (**must** be done one-at-a-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 need to be in same block but markers within a sample can be in a different order.
    - 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.
  - It is more efficient to import several small databases than one big.
  - Time usage to import a database file with 17 markes:
    - 1e6 profiles takes about 131 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:"

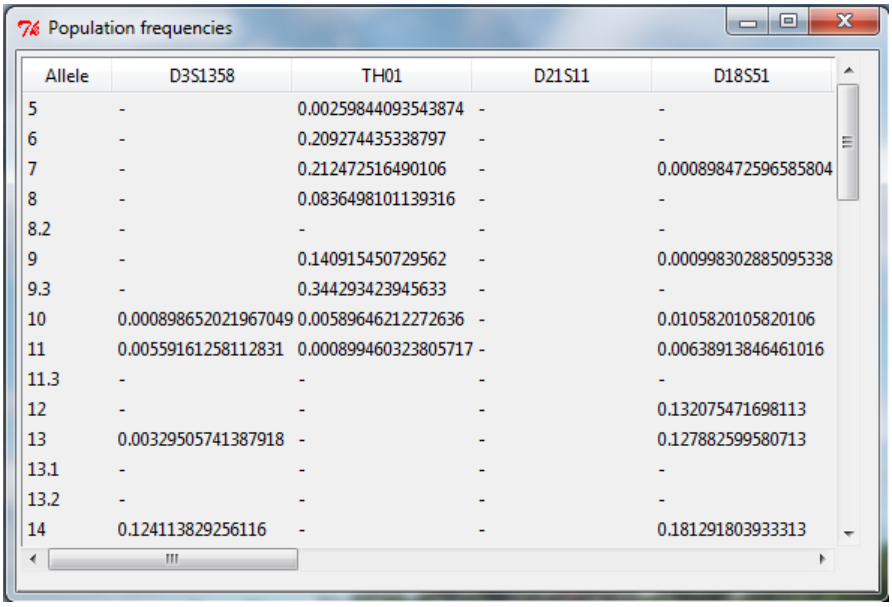
	Sample.Name	Marker	Allele.1	Allele.2
1	00-JP0001-14_20142342311_NO-3241	D3S1358	14	15
2	00-JP0001-14_20142342311_NO-3241	TH01	7	9.3
3	00-JP0001-14_20142342311_NO-3241	D21S11	29	30
4	00-JP0001-14_20142342311_NO-3241	D18S51	13	17
5	00-JP0001-14_20142342311_NO-3241	D10S1248	12	13
6	00-JP0001-14_20142342311_NO-3241	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 8: The figure shows the table format for the imported reference database file.

VIEW DATA

- **View frequencies** (see **Figure 9** for the Norwegian ESX17 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 an 'inclusion probability' plot for each of the selected profiles.
    - The plot (**Figure 10**) shows the exact probability[1] that a random reference profile (from population) **('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 the number of allowed mismatches (default is wildcardsize =5).
    - 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 is used.
    - New alleles which are not found in the selected population are assumed to have allele-frequency freq0 (see under **Frequencies** in section **1-Toolbar**).

[1] The formula is given in the section 'Exact random allele sharing with evidence stain' under (C) Supplementary.



Allele	D3S1358	TH01	D21S11	D18S51
5	-	0.00259844093543874	-	-
6	-	0.209274435338797	-	-
7	-	0.212472516490106	-	0.000898472596585804
8	-	0.0836498101139316	-	-
8.2	-	-	-	-
9	-	0.140915450729562	-	0.000998302885095338
9.3	-	0.344293423945633	-	-
10	0.000898652021967049	0.00589646212272636	-	0.0105820105820106
11	0.00559161258112831	0.000899460323805717	-	0.00638913846461016
11.3	-	-	-	-
12	-	-	-	0.132075471698113
13	0.00329505741387918	-	-	0.127882599580713
13.1	-	-	-	-
13.2	-	-	-	-
14	0.124113829256116	-	-	0.181291803933313

Figure 9: The figure shows the viewed frequencies for the Norwegian ESX17 population.

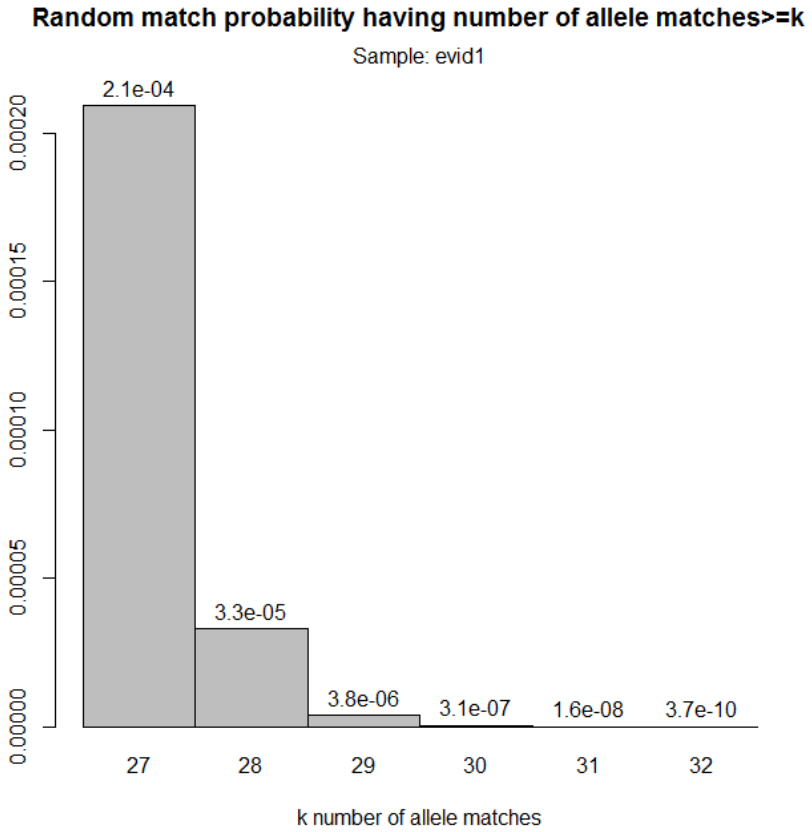


Figure 10: The figure shows the random probability of a match with at least k number of alleles (from a randomly chosen reference profile) compared with the observed alleles in evidence profile (wildcardsize=5).

- **View evidence** (for selected evidence):
  - Prints imported loci, along with allele designations (and peak heights if any) for each selected evidence profile(s) (**Figure 11**).

```
[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"
D10S1248 "13/14/15"            "1856/155/1045"
D1S1656  "12/15/16/16.3/17.3" "1140/601/488/155/1877"
D2S1338  "17/19/20/23"         "290/619/259/649"
D16S539  "9/10/11/12"          "217/312/743/619"
D22S1045 "15/16"               "1017/610"
VWA      "14/15/17"            "1250/440/1232"
D8S1179  "10/13/14/15"         "206/352/978/827"
FGA      "21/22"               "664/714"
D2S441   "9/10/11/14"          "200/3362/1168/3693"
D12S391  "18/18.3/19/21/22"    "297/1446/751/171/1370"
D19S433  "13/14/15.2"          "1157/781/922"
SE33     "29.2/30.2/33.2"      "221/473/570"
```

Figure 11: The figure shows the printed alleles and heights in the imported evidence.

- Plot EPG (**Figure 13**) and degradation plot (**Figure 12**) for each selected evidence profile(s)
  - The kit selected under '**Select STR kit**' denotes the EPG format.
  - Loci in evidence which are **inconsistent** with the ones in selected kit (or missing) are **not shown** in the EPG.
  - If reference profiles are imported and selected, they will be labeled together with the peak heights in the EPG plot (as shown in **Figure 13**).
  - The degradation plot shows points and a fitted regression line using sum peak heights at each marker (for the average fragment length).

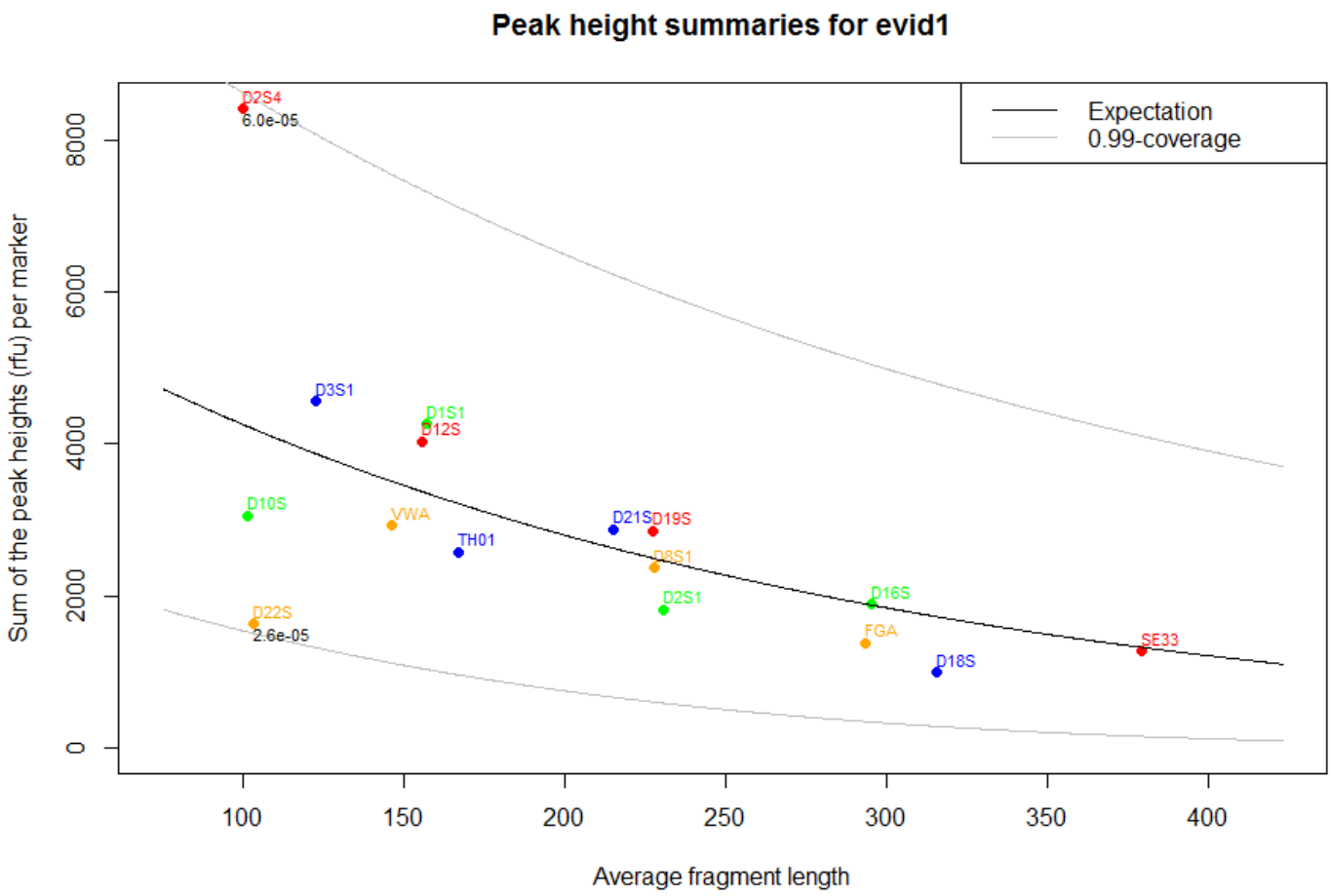


Figure 12: The figure shows peak height sum points *y* on the fitted gamma-regression model gamma(y,shape=2/sigma^2*beta^((x-125)/100),scale=mu*sigma^2), where x='average fragment length of observed alleles at the particular marker'. The model is fitted by inserting the maximum likelihood estimates for mu (P.H.expectation),sigma (P.H.variability) and beta (Degradation-slope.). The black solid line is the expectation of the fitted gamma-regression model, with corresponding 0.005- and 0.995-quantiles of the distribution. The text under the points are p-values found by inserting each point into the fitted gamma-regression based on all data points except for the corresponding inserted data point. These are only shown if they are smaller than 0.05/#loci.

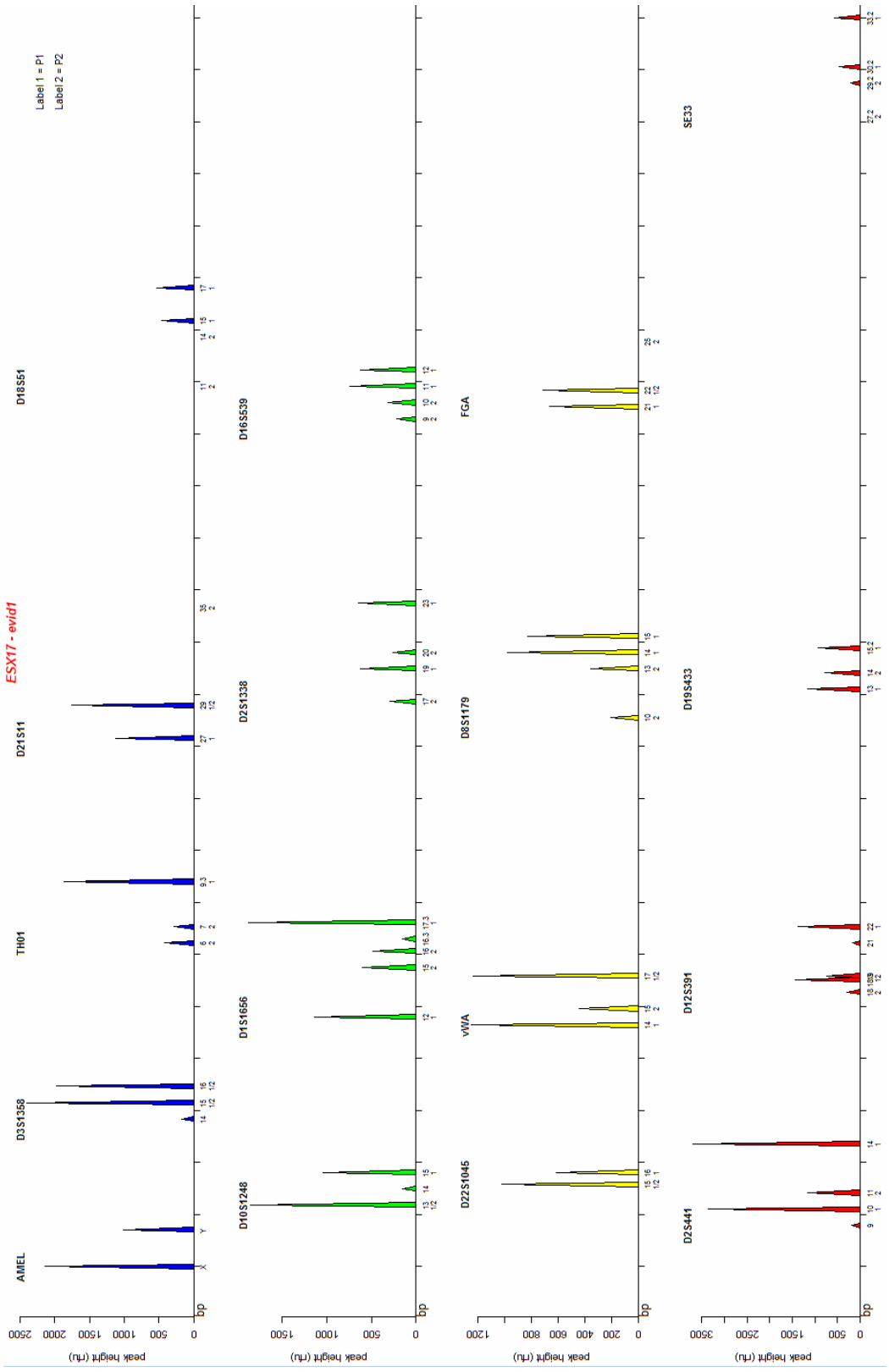
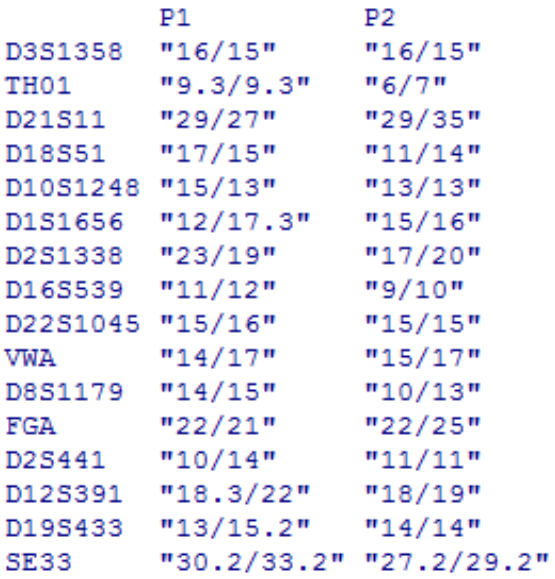


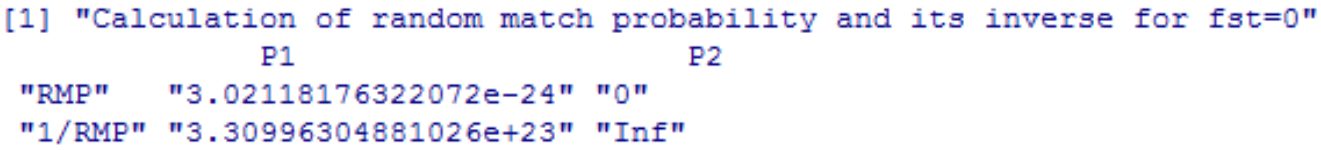
Figure 13: The figure shows the plotted EPG (on the selected SGMPlus kit format) of the imported evidence stain. The labels under the alleles show the imported and selected reference profiles.

- **View reference** (for selected reference):
  - Prints imported genotypes for each selected reference profile(s) (**Figure 14**).
  - If allele frequencies are imported and selected, the software will calculate the random match probability and its inverse, for each selected reference profile(s) (here the corresponding selected fst value will be taken into account) (**Figure 15**).
  - If any evidence profiles(s) are selected after evidence-import, the software counts number of matching alleles (MAC) for each locus of the selected reference profiles, for each selected evidence (**Figure 16**).
    - MAC = number of alleles for the reference which are included in the evidence.
    - nLocs = number of considered loci when counting MAC.



	P1	P2
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 14: The figure shows the printed alleles of the imported reference profiles.



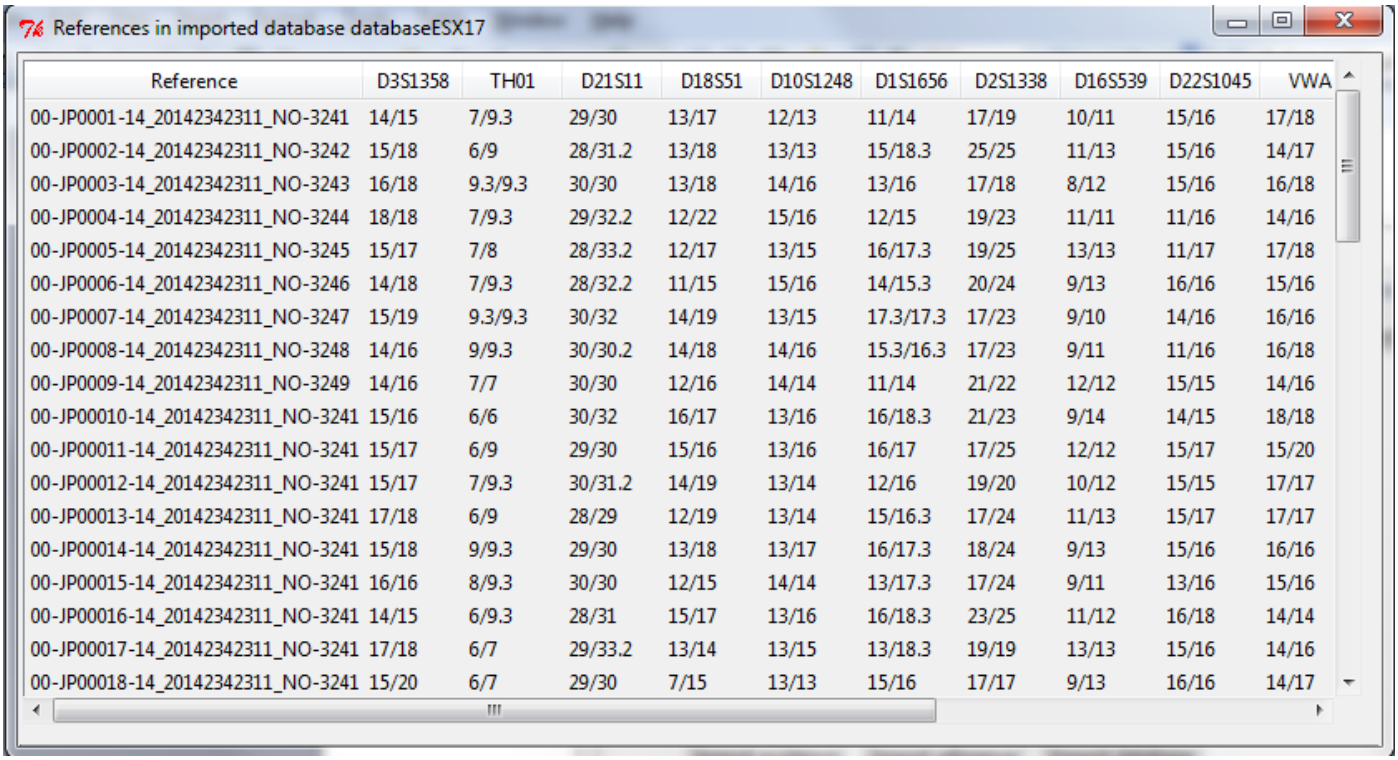
```
[1] "Calculation of random match probability and its inverse for fst=0"
             P1                      P2
 "RMP"   "3.02118176322072e-24" "0"
 "1/RMP" "3.30996304881026e+23" "Inf"
```

Figure 15: The figure shows the printed random match probabilities (and inverse) for each references.

```
[1] "Number of matching alleles with samplename evid1:"
         P1 P2
AMEL     NA NA
D3S1358   2  2
TH01      2  2
D21S11    2  1
D18S51    2  0
D10S1248  2  2
D1S1656   2  2
D2S1338   2  2
D16S539   2  2
D22S1045  2  2
VWA       2  2
D8S1179   2  2
FGA       2  1
D2S441    2  2
D12S391   2  2
D19S433   2  2
SE33      2  1
MAC      32 27
nLocs    16 16
```

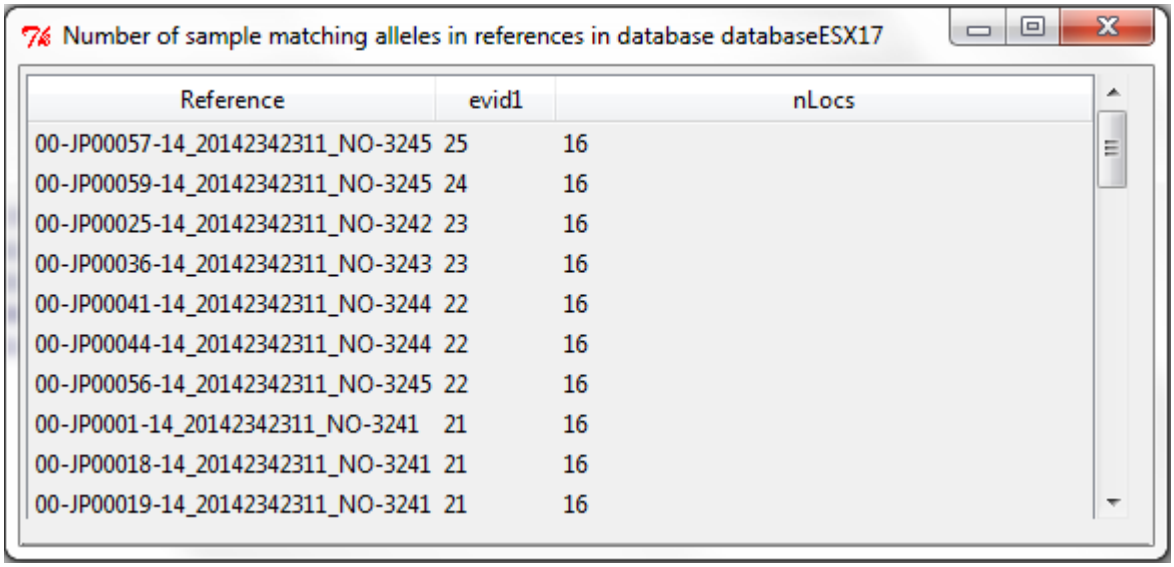
Figure 16: The figure shows number of matching alleles and total (MAC) between the imported references and selected evidence stain. By combining the observed MAC and **Figure 10**, the random match probability of observing at least MAC is useful to provide a 'more meaningful' version of "Random man not excluded"-statistics: The random match probability for P1 (MAC>=32) is 3.7e-10, while only 2.1e-4 for P2 (MAC>=27).

- **View database** (see **Figure 17** for selected database):
  - Creates a new window (for each selected database) which shows the genotypes for every reference in the database.
    - "-" means that the genotype of a reference was missing.
  - If any evidence profiles(s) are selected after evidence-import, the software counts the number of matching alleles (MAC) for all references in the database against each of the selected evidences (see **Figure 18**). 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.
    - **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.
    - Only loci within the selected kit will be shown and used in further calculations.



Reference	D3S1358	TH01	D21S11	D18S51	D10S1248	D1S1656	D2S1338	D16S539	D22S1045	VWA
00-JP0001-14_20142342311_NO-3241	14/15	7/9.3	29/30	13/17	12/13	11/14	17/19	10/11	15/16	17/18
00-JP0002-14_20142342311_NO-3242	15/18	6/9	28/31.2	13/18	13/13	15/18.3	25/25	11/13	15/16	14/17
00-JP0003-14_20142342311_NO-3243	16/18	9.3/9.3	30/30	13/18	14/16	13/16	17/18	8/12	15/16	16/18
00-JP0004-14_20142342311_NO-3244	18/18	7/9.3	29/32.2	12/22	15/16	12/15	19/23	11/11	11/16	14/16
00-JP0005-14_20142342311_NO-3245	15/17	7/8	28/33.2	12/17	13/15	16/17.3	19/25	13/13	11/17	17/18
00-JP0006-14_20142342311_NO-3246	14/18	7/9.3	28/32.2	11/15	15/16	14/15.3	20/24	9/13	16/16	15/16
00-JP0007-14_20142342311_NO-3247	15/19	9.3/9.3	30/32	14/19	13/15	17.3/17.3	17/23	9/10	14/16	16/16
00-JP0008-14_20142342311_NO-3248	14/16	9/9.3	30/30.2	14/18	14/16	15.3/16.3	17/23	9/11	11/16	16/18
00-JP0009-14_20142342311_NO-3249	14/16	7/7	30/30	12/16	14/14	11/14	21/22	12/12	15/15	14/16
00-JP00010-14_20142342311_NO-3241	15/16	6/6	30/32	16/17	13/16	16/18.3	21/23	9/14	14/15	18/18
00-JP00011-14_20142342311_NO-3241	15/17	6/9	29/30	15/16	13/16	16/17	17/25	12/12	15/17	15/20
00-JP00012-14_20142342311_NO-3241	15/17	7/9.3	30/31.2	14/19	13/14	12/16	19/20	10/12	15/15	17/17
00-JP00013-14_20142342311_NO-3241	17/18	6/9	28/29	12/19	13/14	15/16.3	17/24	11/13	15/17	17/17
00-JP00014-14_20142342311_NO-3241	15/18	9/9.3	29/30	13/18	13/17	16/17.3	18/24	9/13	15/16	16/16
00-JP00015-14_20142342311_NO-3241	16/16	8/9.3	30/30	12/15	14/14	13/17.3	17/24	9/11	13/16	15/16
00-JP00016-14_20142342311_NO-3241	14/15	6/9.3	28/31	15/17	13/16	16/18.3	23/25	11/12	16/18	14/14
00-JP00017-14_20142342311_NO-3241	17/18	6/7	29/33.2	13/14	13/15	13/18.3	19/19	13/13	15/16	14/16
00-JP00018-14_20142342311_NO-3241	15/20	6/7	29/30	7/15	13/13	15/16	17/17	9/13	16/16	14/17

Figure 17: The figure shows the viewed references from the imported ESX17 database



Reference	evid1	nLocs
00-JP00057-14_20142342311_NO-3245	25	16
00-JP00059-14_20142342311_NO-3245	24	16
00-JP00025-14_20142342311_NO-3242	23	16
00-JP00036-14_20142342311_NO-3243	23	16
00-JP00041-14_20142342311_NO-3244	22	16
00-JP00044-14_20142342311_NO-3244	22	16
00-JP00056-14_20142342311_NO-3245	22	16
00-JP0001-14_20142342311_NO-3241	21	16
00-JP00018-14_20142342311_NO-3241	21	16
00-JP00019-14_20142342311_NO-3241	21	16

Figure 18: The figure shows the sorted references (in the reference database) with respect to MAC (total number of matching alleles) compared to the selected evidence.

INTERPRETATIONS

- **Weight-of-Evidence**:

  - Weight-of-Evidence is carried out by comparing the Likelihood Ratio (LR) between the specified hypotheses Hp (prosecution) and Hd (defense) using the quantitative model as given in the euroformix paper. There are a number of options as follows:
  - Modules:
    1) 'Quantitative LR' (Maximum Likelihood based)
       - Optimizes (maximum) the model parameters in the continuous model.
    2) 'Quantitative LR' (Bayesian based)
       - Integrates out the model parameters in the continuous model.
    3) 'Qualitative LR' (semi-continuous) – Mirrors the LRmix module but also contains the maximum likelihood based method.

    Note that 'Easy Mode' uses module 1) 'Continuous LR' (Maximum Likelihood based).

  - Requirements:
    - Imported population frequencies, **at least one** evidence profile and **at least one** reference profile (e.g. a 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.

  - Features:
    - The quantitative model: Handles replicates, allele drop-in, allele drop-out, backward-stutter, fst-correction and degradation.
    - The qualitative model: Handles replicates, allele drop-in, allele drop-out (equal across contributors) and fst-correction.

- **Deconvolution**:

  - Deconvolution ranks the most probable combined genotype profiles given **a specified hypothesis** based on a maximum likelihood fitted quantitative model (as given in the euroformix paper).

  - 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, allele drop-out, backward-stutter, fst-correction and degradation.

- **Database search:**

    - Carries out 'weight-of-evidence' tests by comparing the Likelihood Ratio (LR) between the specified hypotheses Hj (reference j in database) and Hd (defense) using the quantitative model as given in the euroformix paper.

    - Modules:
        1) 'Quantitative LR' (Maximum Likelihood based)
        2) 'Quantitative LR' (Bayesian based)
        3) 'Qualitative 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, backward-stutter, fst-correction and degradation.
        - The quantitative LR value is shown together with qualitative LR and MAC.

- **Fit drop-in data:**
    - When clicking the button the user must select a text-file which contains drop-in peak heights in a separated format (";","","","\t","\n").
    - This will fit the lambda parameter for the shifted exponential drop-in function, and the value of the lambda parameter in the "Settings" will automatically be updated accordingly.
    - A histogram of the peak heights will be present in a plot together with the fitted model.
    - **WARNING**: Remember to specify your detection threshold in "Settings" before doing this, since the estimated lambda depends on this.

- **Generate sample:**
    - Generates alleles using the population frequencies and draws peak heights for a specified hypothesis using the quantitative model as described in the euroformix paper.
    - Requires: Imported population frequencies.
    - Feature: All the parameters in the quantitative model (EFM).

- **Restart**
    - Simply restarts the program.

3- Model specification

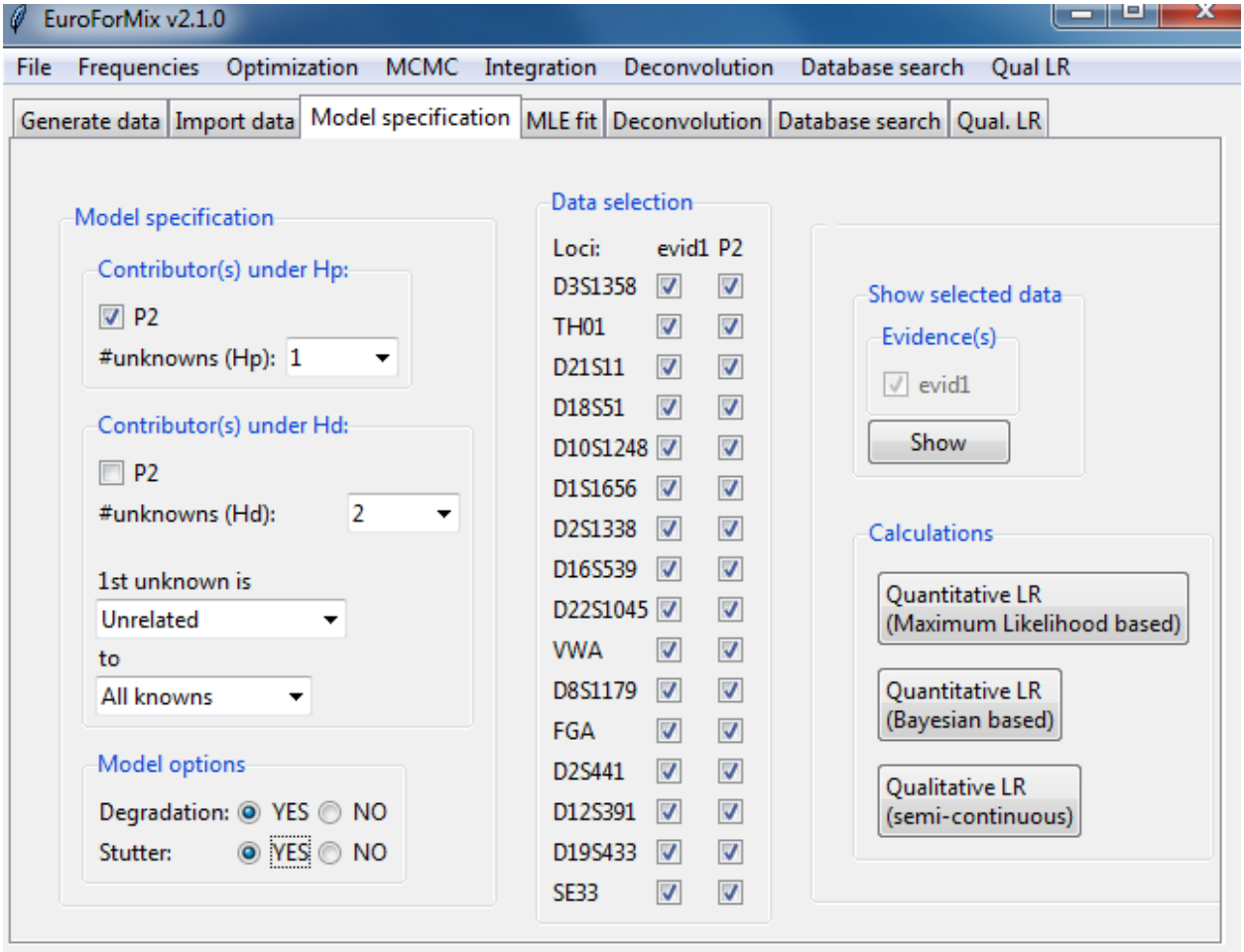


Figure 19: The figure shows the Model Specification page for **Weight-of-Evidence**

MODEL SPECIFICATION

The model specification tab is invoked from several different routes. From the 'Import data' tab the options that can be followed are the buttons: Generate sample, Weight of evidence, Database search and Deconvolution. The effect and properties of each case are as follows:

- **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 carried out 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 defense hypothesis Hd.

    - **Relatedness module** (updated from v2):
        - The user can now specify relationships between the 1st unknown contributor (only under Hd) to an imported reference sample.
        - Supported relationships:
            - Unrelated (this is default).
            - Parent/Child, Sibling, Uncle/Nephew, Grandparent/Grandchild, Half-sibling, Cousin.
        - The relatedness model will be used for the following features:
            - The quantitative LR calculations (also for database searching).
            - Non-contributor tests: The non-contributors will be random 1st unknown individuals as specified under Hd.
            - LR sensitivity, MCMC simulations, Deconvolution, Model validation, Model fitted P.H.
        - The relatedness model is not (yet) implemented for the following features:
            - Qualitative LR and its non-contributor analysis.
            - The "Generate data" module.
        - Note:
            - If a relationship is specified, the user must select which of the imported reference samples the unknown is related to.
            - The given relationship will be indicated in the "report".
            - Theta/fst correction is taken into account for this module.

    - Case: **Weight-of-Evidence** or **'Database search'**: References which are conditioned under Hp but not under Hd, will be assumed to be **known non-contributors** under Hd (this is relevant when fst>0).

    - Case **'Database search':** When doing a database search with "Quantitative 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).

- **Model options**

    - **Stutter:** Boolean of incorporating a model for (n-1) backward-stutters by including parameter **Stutter-prop.** (xi):
        - **Stutter-prop.** is a constant parameter which denotes the expected fraction of the contributor expected peak height moved from allele *a* to allele *a-1*. See euroformix paper for more details.

    - **Degradation**: Boolean of incorporating a global degradation model where the slope is determined with parameter **Degrad.slope** (beta):
        - The expected peak height of a specific allele is modelled to be proportional to beta^{(s-125)/100}, where s is the fragment length of the corresponding allele.

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 is missing or has been unselected for any of the evidence(s) or reference(s), the missing/unselected locus will not be evaluated at all.
    - Note: At default there is a limitation of 30 loci that can be selected. This can be changed under "File->Settings->Max locus".

- **Missing data**:

    - Missing markers in evidence samples (but present in references) will not be evaluated (this deactivates the selection ticks). However, EuroForMix handles markers that have fully dropped out in the evidence profile, hence these markers should still be included in the evidence sample file (but with no alleles).
    - If marker information for references samples is missing (this deactivates the selection ticks). For such markers the program will substitute the reference with an unknown contributor.

- **New alleles**:

    - If alleles that do not exist in the population allele frequency table occur in the imported evidence, the new alleles are assigned with allele frequency 'freq0'. freq0 can be specified in several ways:

        1. 'freq0' is equal to the minimum observed allele frequency in the population table if N=0, or 'freq0'=5/(2N) otherwise where N is number of individuals used to create the imported frequency database. This can be changed manually under "Frequencies->**Set size of frequency database**" in Toolbar.
        2. freq0 is the frequency set by the user in "Frequencies -> **Set minimum frequency**".

    - **WARNING**: The population frequencies are not normalized after adding new allele frequencies to the population frequencies.

SHOW SELECTED DATA

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

- **Show:**
    - **Prints the following to the R-console:** The selected evidence sample(s), reference(s) and considered population frequencies which are used for further analysis.
    - The selected evidence samples are shown in an EPG-plot.
        - Note: Alleles with corresponding peak heights below the specified "Detection threshold" (indicated as the horizontal gray lines) are removed.

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

CALCULATIONS

- **'Quantitative LR (Maximum Likelihood based) '** (case **Weight-of-Evidence** and **'Database search')**:

- Maximizes the likelihood with respect to the unknown parameters in the quantitative model 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, by chance. Number of start-points should be sufficiently large to ensure that the optimizer always finds the global maximum of the Likelihood function. This can be changed under "Optimization->Set number of startpoints" in Toolbar.
    - After calculation, the page 'MLE fit' is visited to present results.

- **'Quantitative LR (Bayesian based)'** (case **Weight-of-Evidence** and **'Database search'**):

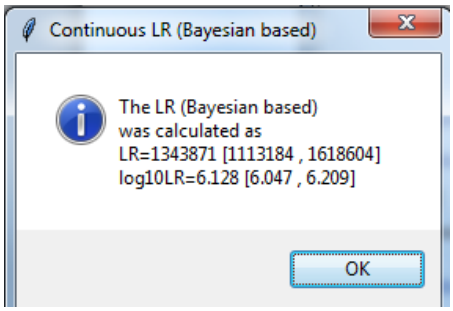


Figure 20: The figure shows the calculated Weight-of-Evidence based on the Bayesian based quantitative LR for the specified model in **Figure 19**.

- Instead of maximizing the likelihood of the unknown parameters, a **numerical integration** over the unknown parameters are applied both under hypothesis Hp and Hd. The ratio becomes the estimated LR value (marginalized with respect to the model parameters).

- The accuracy of the integrals depends on the specified '**relative error requirement'** (see euroformix paper for details).
    - Can be changed under "Integration" in Toolbar. Default is 0.1.

- The integral requires that an **upper boundary** for the parameters P.H. expectation (mu), P.H. variability (sigma) and Stutter-prop. (xi) are specified. The default values for these are 20000, 0.9 and 0.5, respectively. These values may be changed under "Integration" in Toolbar. See euroformix paper for details.

- Calculates LR-values directly and avoids visiting the tab 'MLE fit'.
    - Case **Weight-of-Evidence**: A message with the estimated LR, with the relative errors given in brackets, pops up after calculation (see **Figure 20**).
        - See the euroformix paper for details of how the relative errors are calculated.

- Case **'Database search'**: Database search results based on the estimated LR are shown directly after calculation (goes to tab 'Database search').

- Sometimes the likelihood values become zero because of underflow (happens for multiple replicates or large number of loci). To avoid this, the user can change the value under "Integration-> Set likelihood-scaling to avoid zero". It is recommended that this value is set close to the magnitude of the maximum likelihood.

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

  - Performs a semi-continuous procedure (mirrors the LRmix module) where the distribution of the 'allele drop-out probability given the number of observed alleles' are utilized to infer a "conservative" LR.
    - The model is purely qualitative which means that it is only based on allele-designation information.
    - Here it is also possible to obtain the LR based on the maximum likelihood method.

  - Goes directly to page Qual. LR.

- **'Generate sample'** (case **'Generate sample'**):
  - Push **'Generate sample'** button under the 'Import data' tab – this opens the Model specification tab.
  - A dataset (an evidence sample and the contributing references) will be randomly simulated under the specified model under "Model specification" (relatedness is not yet implemented).
  - Reference profiles may be imported and selected as assumed known in the hypothesis.
  - The settings for Detection threshold, probability of drop-in and drop-in peak height hyperparam will be used in the simulation (fst-correction is not used).
  - The unknown contributor profiles under the hypothesis will be randomly generated using the selected population frequencies (the relatedness module is not used).
  - The simulated peak heights of the generated evidence sample are entirely based on the quantitative model for assumed values of the model-parameters (**mu=P.H.expectation, sigma=P.H.variability, xi=Stutter-prop., mx=mixture prop., beta=Degrad.slope)**. Default values are **mu**=1000, **sigma**=0.15, **xi**=0.1, **beta**=1, **mx**=(C:1)/sum(C:1), where C is number of contributors.
  - Once the model is specified, push button 'Generate sample' in the 'model specification' tab. The output goes directly to page Generate data. Turn to section 7 for a full description of this page.
  - Notice: This module does not include fst-correction or relatedness module.

4- MLE fit: (Maximum Likelihood based)
('Quantitative LR used in 'easy mode')

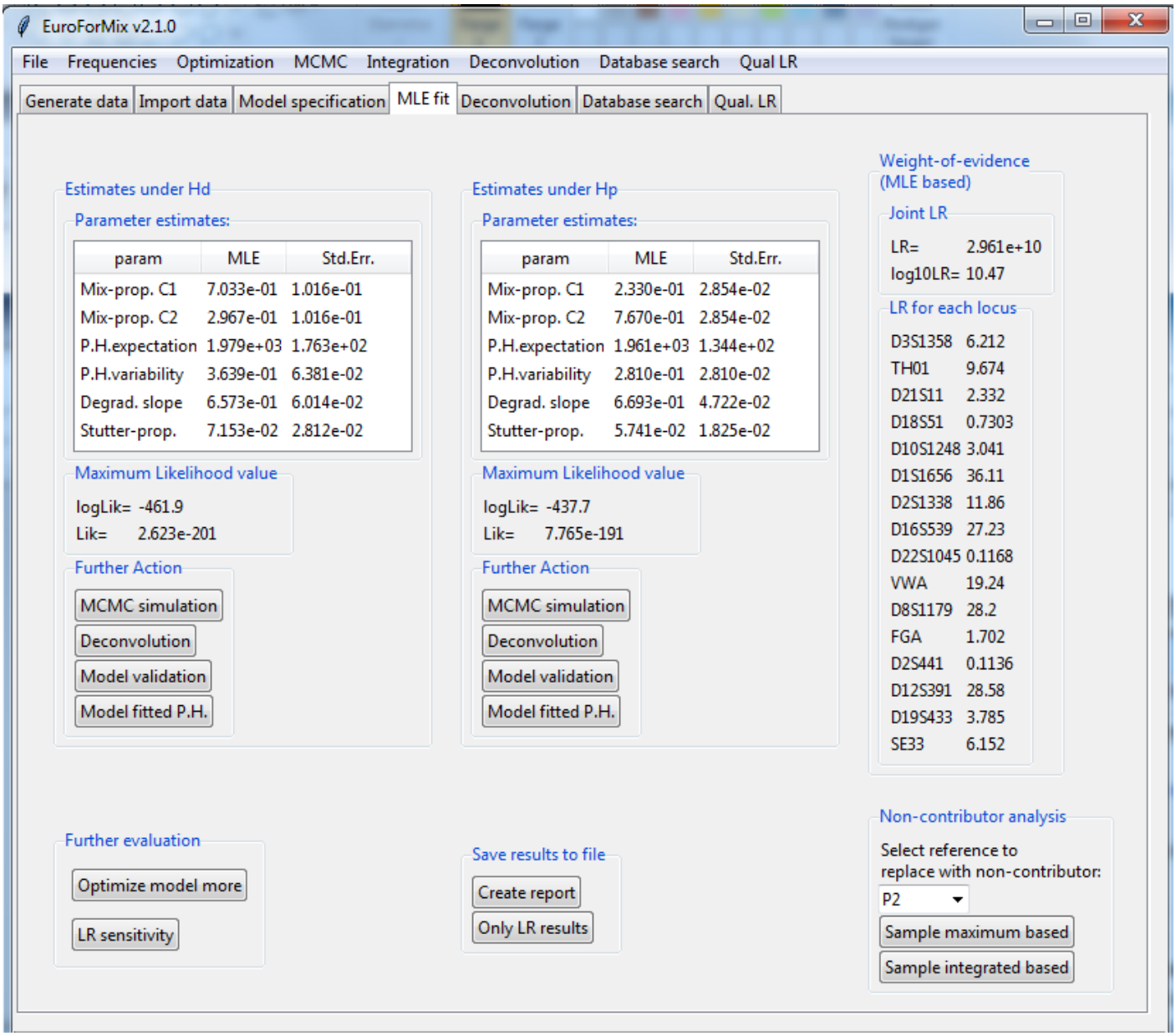


Figure 21: The figure shows the MLE-fit page after running the **Quantitative LR (Maximum Likelihood based)** calculation (maximizing the quantitative model with respect to the unknown parameters for each of the specified hypothesis in **Figure 19**) for **Weight-of-Evidence**.

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

- Parameter estimates:

- param: The unknown parameters in the model (see euroformix paper for more details).
  - Mix-prop. Ci (Mxi): Mixture-proportion for contributor 'i'.
  - P.H.expectation (mu): Expectation of the peak height for a single heterozygote (Mix-prop=1) allele without degradation (at 125 bp or beta=1).
  - P.H.variability (sigma): Coefficient of variation of the peak height for a single heterozygote (Mix-prop=1) allele without degradation (at 125 bp or beta=1).
  - Stutter-prop (xi): A global parameter related to backward-Stutter proportion. The expected fraction of peak height that are stutter.
  - Degrad.slope (beta): A parameter related to the degree of decaying degradation global for all contributors.

- MLE: The optimized[2] parameters in the model which attain a maximum point of the likelihood function.

- Std.Err.: The standard error of the parameter estimates in the model. These are based on the hessian matrix returned from the optimizer function nlm.

- Maximum Likelihood value:

- log10lik and Lik: The logged (10-base) and the original value of the Likelihood value attained from the optimization[1].

- Further Action:

- **MCMC simulation** (see **Figure 22**):
  - Performs 'Markov Chain Monte Carlo (MCMC) Metropolis Hastings' sampling under the desired hypothesis using parallelization (one chain for each CPU threads).
  - The MCMC proposal function is based on a normal distribution with the mode as and the covariance matrix attained from the optimization as parameters.
  - The sampling is done blockwise, with one block for mixture proportions, one block for peak height parameters (P.H.expectation, P.H.variability, Degrad.slope), and one block for the stutter-prop. parameter.
  - 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 R-console.
  - Acceptance rate = number of accepted samples divided by number of proposed samples.

[2] 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.

- Tweak '**variance of randomizer**' under MCMC in toolbar to change the acceptance rate.
    - Ideally the acceptance rate should be around 0.2 (0.05-0.3) to ensure that the parameter space has been fully explored.
- The user may **change number of required samples** in the simulation under 'MCMC->Set number of samples' (n) in toolbar.
    - This is the number of samples for one chain and one "block". E.g. with n=1000, 4 CPU threads available and a model including stutter, the MCMC will perform 3*4*1000 = 12000 iterations.
- The **purpose** of the MCMC simulation is to use it as an **exploratory tool** to show:
    - That the optimizer has found the global maximum.
    - The shape of the posterior distribution of the parameters.

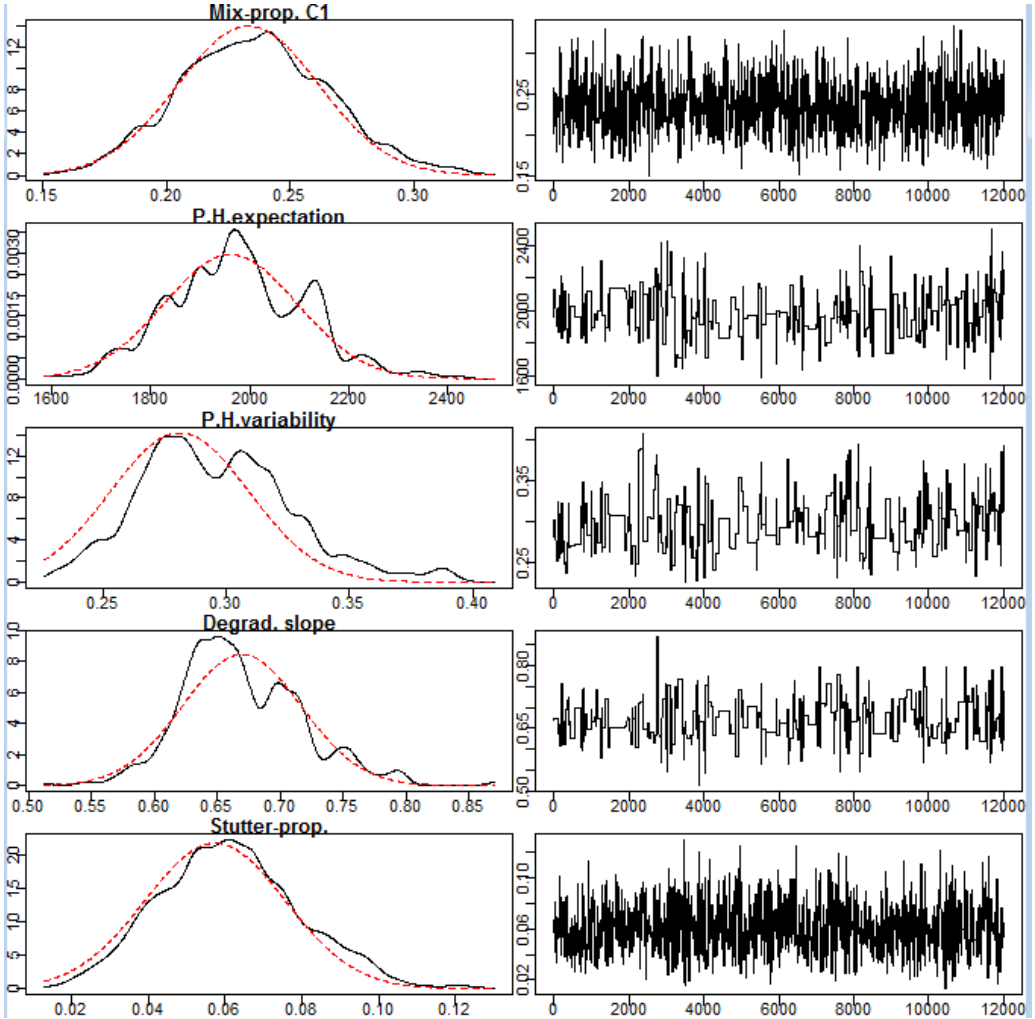


Figure 22: The figure shows the posterior density of the unknown parameters (first column) and corresponding iteration values (second column) from the MCMC method under the hypothesis Hp: "Reference P1+1 unknown individual contributes to evidence evid1".

- **Deconvolution**:
  - Performs “Deconvolution” under the desired hypothesis, where the genotypes for the unknown contributors are ranked with respect to the posterior probability (based on the quantitative likelihood function and allele frequencies). See section 5-Deconvolution.

- **Model validation** (**Figure 23**):
  - Estimates the cumulative probability of the observed peak heights conditional on the other peak heights. These probabilities are compared with the theoretical underlying model (see euroformix paper for more details).
  - In theory the cumulative probabilities follows a uniform distribution, if the underlying density model is "reasonable" (null-hypothesis) – giving a straight line in the plot.
  - The j-th largest (out of n) observed probability (y-axis) is distributed as beta(j, n - j + 1) when observed probabilities are independently uniform(0,1).
  - Quantiles of the beta-distribution are shown as the envelopes in **Figure 22.** The black lines are the 0.005 and 0.995 quantiles, while the red lines are the Bonferroni-adjusted 0.005/n and 0.995/n quantiles.
  - The significance level is set initially by the user (pop-up window). Default is 0.01.

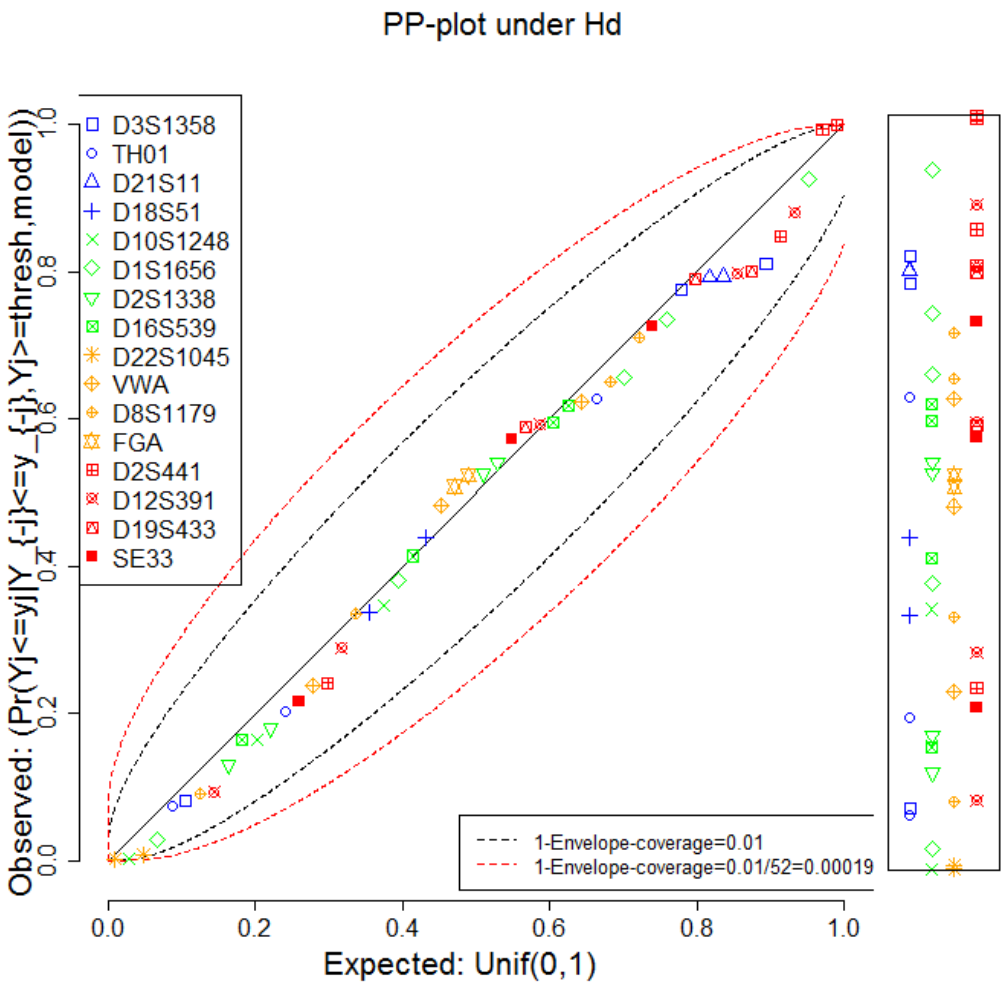


Figure 23: The figure shows “**Model validation**” under Hd.

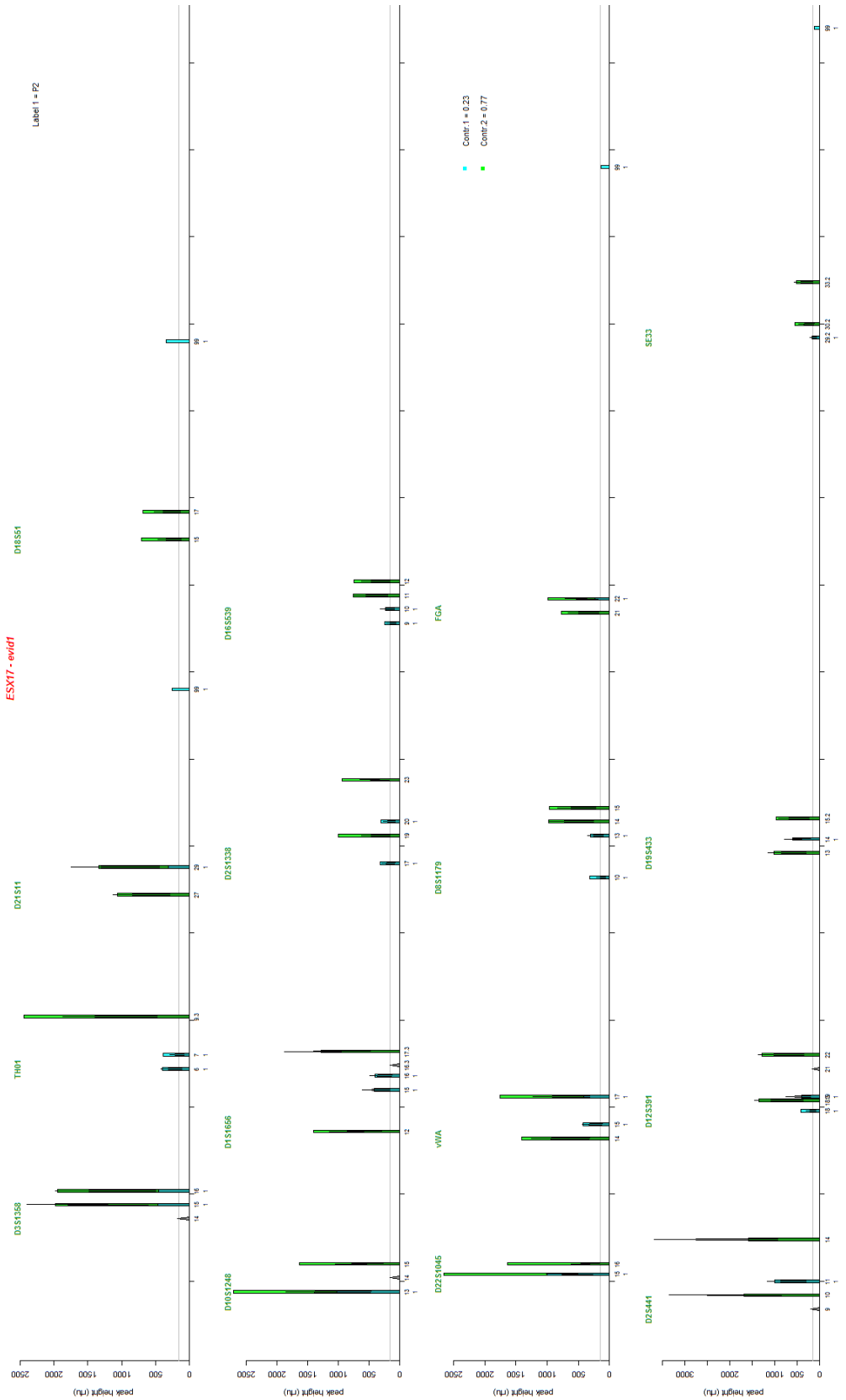


Figure 24: The figure shows the results when clicking *Model fitted P.H.* under Hp.

- **Model fitted P.H.** (**Figure 24**):
  - This button gives a plot which presents the expected peak heights for each contributor in color bars, which are superimposed on top of the peak heights.
  - The expectations are conditioned on the maximum likelihood estimates of the parameters and the most likely genotype for the unknown contributors (see section 5-Deconvolution).
  - If the joint probability of the unknown genotypes is above 0.95 the locus name is colored green, and colored orange if between 0.9 and 0.95 and otherwise red.
  - Drop-out alleles for contributors are presented as "99".

WEIGHT-OF-EVIDENCE (the output of the MLE fit page)

- Description:

  - The LR values are calculated as the ratio between the maximized likelihoods of the two specified hypotheses Hp and Hd as specified in "Model specification".

  - The likelihood function is based on the quantitative model as described in the euroformix paper.

- **Joint 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 locus:**

  - The LR for each locus is provided separately (given the parameter-modes under Hp and Hd).
  - Note: By default there is a limitation of 30 loci to visualize this. This limitation can be changed under "File -> Settings -> Max locus".

FURTHER EVALUATION

- **Optimize model more**:

  - If required, 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 will never return a lower likelihood).

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

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

- **'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 quantitative model under both Hp and Hd are taken into account (see **Figure 25**).
        - Number of samples can be changed with **Set number of samples** under MCMC in Toolbar (default is 1000 samples). See "**MCMC simulation**" for more details.
        - MCMC will vary between runs for the same data. This variation reduces as the number of samples increases.

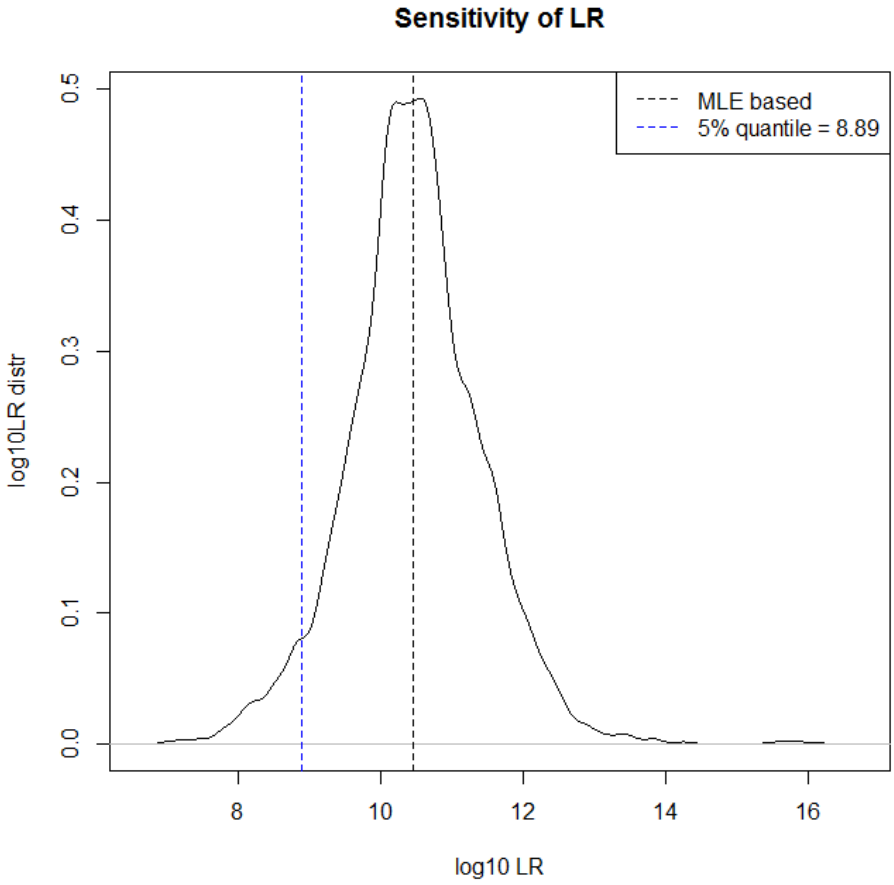
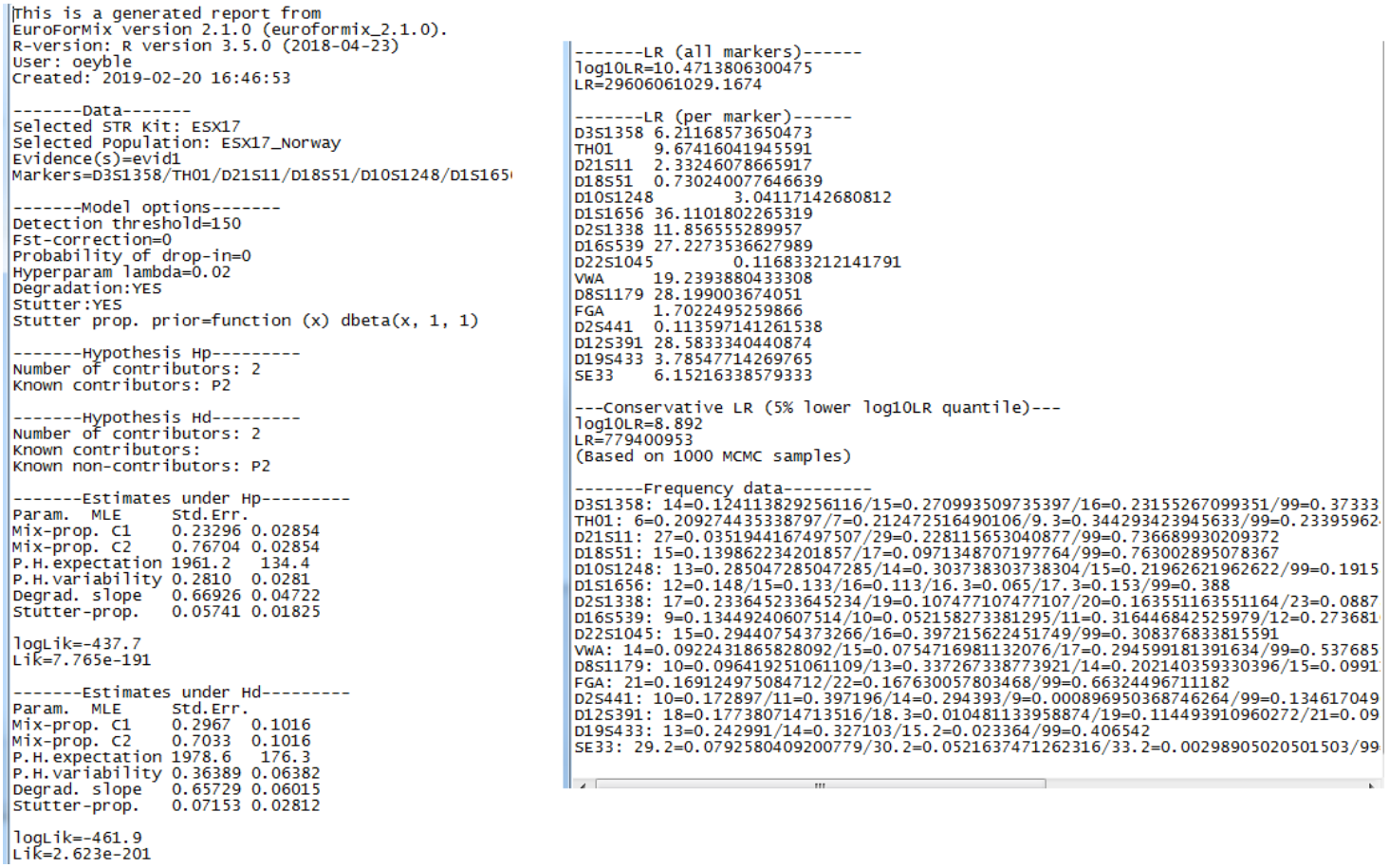


Figure 25: The plot shows the distributed log10 Lik under Hp (upper left), log10 Lik under Hd (upper right) and the combined distribution log10 LR (lower) where the *a posteriori* density of the parameters in the quantitative model under both Hp and Hd are taken into account. *a posteriori* density are simulated using the **MCMC simulation** (**Figure 22** shows only Hp)**.** The black dotted line is the maximum likelihood based LR.

SAVE RESULTS TO FILE

'Create report':

- The parameter estimates with corresponding standard deviation errors estimates and the likelihood values will be printed to file for all hypotheses on page (see **Figure 26**).



```
This is a generated report from
EuroForMix version 2.1.0 (euroformix_2.1.0).
R-version: R version 3.5.0 (2018-04-23)
User: oeyble
Created: 2019-02-20 16:46:53

-------Data-------
Selected STR Kit: ESX17
Selected Population: ESX17_Norway
Evidence(s)=evid1
Markers=D3S1358/TH01/D21S11/D18S51/D10S1248/D1S165

-------Model options-------
Detection threshold=150
Fst-correction=0
Probability of drop-in=0
Hyperparam lambda=0.02
Degradation:YES
Stutter:YES
Stutter prop. prior=function (x) dbeta(x, 1, 1)

-------Hypothesis Hp---------
Number of contributors: 2
Known contributors: P2

-------Hypothesis Hd---------
Number of contributors: 2
Known contributors:
Known non-contributors: P2

-------Estimates under Hp---------
Param.   MLE      Std.Err.
Mix-prop. C1     0.23296 0.02854
Mix-prop. C2     0.76704 0.02854
P.H.expectation 1961.2   134.4
P.H.variability 0.2810   0.0281
Degrad. slope    0.66926 0.04722
Stutter-prop.    0.05741 0.01825

logLik=-437.7
Lik=7.765e-191

-------Estimates under Hd---------
Param.   MLE      Std.Err.
Mix-prop. C1     0.2967   0.1016
Mix-prop. C2     0.7033   0.1016
P.H.expectation 1978.6    176.3
P.H.variability 0.36389 0.06382
Degrad. slope    0.65729 0.06015
Stutter-prop.    0.07153 0.02812

logLik=-461.9
Lik=2.623e-201
```

```
-------LR (all markers)------
log10LR=10.4713806300475
LR=29606061029.1674

-------LR (per marker)------
D3S1358 6.21168573650473
TH01    9.67416041945591
D21S11  2.33246078665917
D18S51  0.730240077646639
D10S1248        3.04117142680812
D1S1656 36.1101802265319
D2S1338 11.856555289957
D16S539 27.2273536627989
D22S1045        0.116833212141791
VWA     19.2393880433308
D8S1179 28.199003674051
FGA     1.7022495259866
D2S441  0.113597141261538
D12S391 28.5833340440874
D19S433 3.78547714269765
SE33    6.15216338579333

---Conservative LR (5% lower log10LR quantile)---
log10LR=8.892
LR=779400953
(Based on 1000 MCMC samples)

-------Frequency data---------
D3S1358: 14=0.124113829256116/15=0.270993509735397/16=0.23155267099351/99=0.37333
TH01: 6=0.209274435338797/7=0.212472516490106/9.3=0.344293423945633/99=0.23395962
D21S11: 27=0.0351944167497507/29=0.228115653040877/99=0.736689930209372
D18S51: 15=0.139862234201857/17=0.0971348707197764/99=0.763002895078367
D10S1248: 13=0.285047285047285/14=0.303738303738304/15=0.21962621962622/99=0.1915
D1S1656: 12=0.148/15=0.133/16=0.113/16.3=0.065/17.3=0.153/99=0.388
D2S1338: 17=0.233645233645234/19=0.107477107477107/20=0.163551163551164/23=0.0887
D16S539: 9=0.13449240607514/10=0.052158273381295/11=0.316446842525979/12=0.273681
D22S1045: 15=0.29440754373266/16=0.397215622451749/99=0.308376833815591
VWA: 14=0.0922431865828092/15=0.0754716981132076/17=0.294599181391634/99=0.537685
D8S1179: 10=0.096419251061109/13=0.337267338773921/14=0.202140359330396/15=0.0991
FGA: 21=0.169124975084712/22=0.167630057803468/99=0.66324496711182
D2S441: 10=0.172897/11=0.397196/14=0.294393/9=0.000896950368746264/99=0.134617049
D12S391: 18=0.177380714713516/18.3=0.010481133958874/19=0.114493910960272/21=0.09
D19S433: 13=0.242991/14=0.327103/15.2=0.023364/99=0.406542
SE33: 29.2=0.0792580409200779/30.2=0.0521637471262316/33.2=0.0029890502050150 3/99
```

Figure 26: The stored information in the text file generated from **'Create report'**.

- **'Only LR results'**: **(case Weight-of-Evidence)**
  - The LR calculated values shown in WEIGHT-OF-EVIDENCE will be printed to file (**Figure 26**).

Marker	LR	log10LR
D3S1358	6.212e+00	0.7932
TH01	9.674e+00	0.9856
D21S11	2.332e+00	0.3678
D18S51	7.302e-01	-0.1365
D10S1248	3.041e+00	0.4830
D1S1656	3.611e+01	1.5576
D2S1338	1.186e+01	1.0740
D16S539	2.723e+01	1.4350
D22S1045	1.168e-01	-0.9324
VWA	1.924e+01	1.2842
D8S1179	2.820e+01	1.4502
FGA	1.702e+00	0.2310
D2S441	1.136e-01	-0.9446
D12S391	2.858e+01	1.4561
D19S433	3.785e+00	0.5781
SE33	6.152e+00	0.7890
JointMLE	2.961e+10	10.4714

Figure 27: The stored information in **'Only LR results'** is the calculated Likelihood Ratio values given the fitted Maximum Likelihood parameters under each of the inferred hypotheses for each locus together with the joint LR.

NON-CONTRIBUTOR ANALYSIS

- "**Sample maximum based"**
  - Sampled random individuals are calculated with the MLE based Likelihood Ratio, i.e. **Quantitative LR (Maximum Likelihood based)**

- **"Sample integrated based"**
  - Sampled random individuals are calculated with the integrated based Likelihood Ratio, i.e. **quantitative LR (Bayesian based)**

- **Select reference to replace with non-contributor:**
  - A drop-down list of references which are conditioned under Hp but not under Hd.

- **Sample non-contributors:**
  - Random non-contributor 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/her LR. In default, 1e3 of random non-contributors are simulated to determine the LR distribution of non-contributors.
    - The mean, standard errors of LR, proportion of LR greater than zero and one, and log10LR-quantiles (50%, 95%, 99%, max) are printed out to R-console.
    - A plot of the cumulative distribution of log10LR will be shown (**Figure 28**).

- Number of non-contributors can be changed under ‘Database search’ in the toolbar.

- Setting fst>0 may be very time-consuming since we require that the sampled non-contributor individual is a known non-contributor under Hd, and hence the likelihood value for Hd is calculated for each sample.

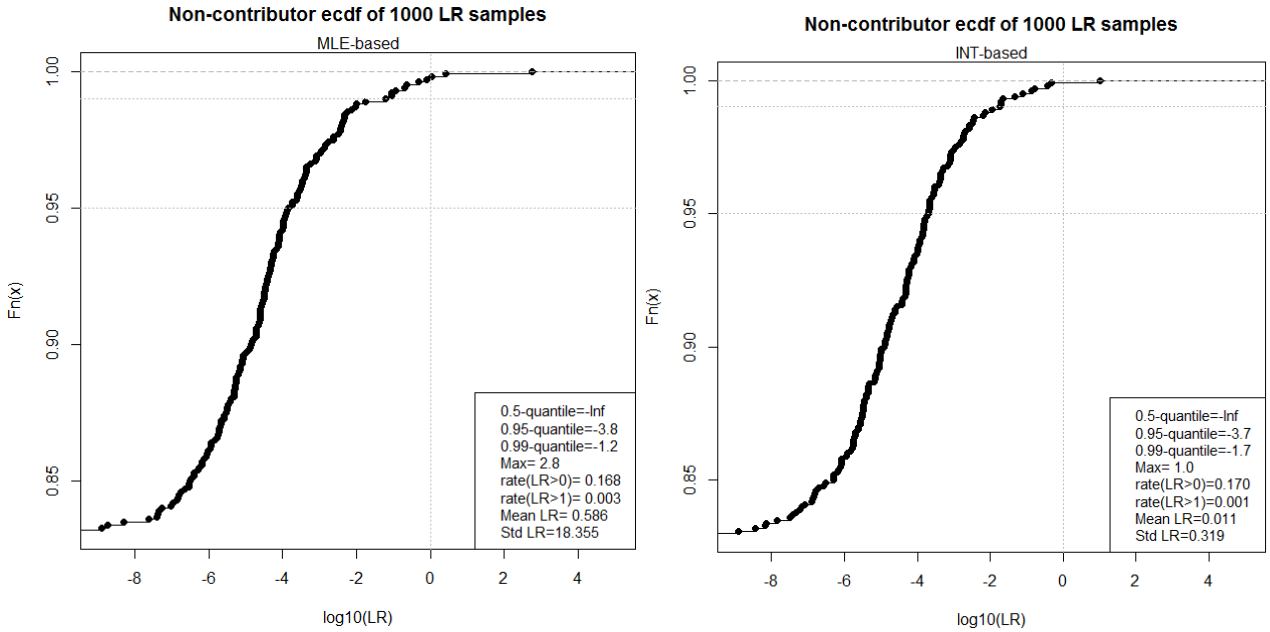


Figure 28: The plot shows the cumulative distribution of 1000 non-contributing individuals replacing Suspect in hypothesis Hp based on the fitted MLE model (left) and for the Integrated based model (right). The mean and standard errors of LR, proportion of LR greater than zero and one, and log10LR-quantiles (50%, 95%, 99%, max) based on the simulated non-contributors are given in the plot as well.

5- Deconvolution:

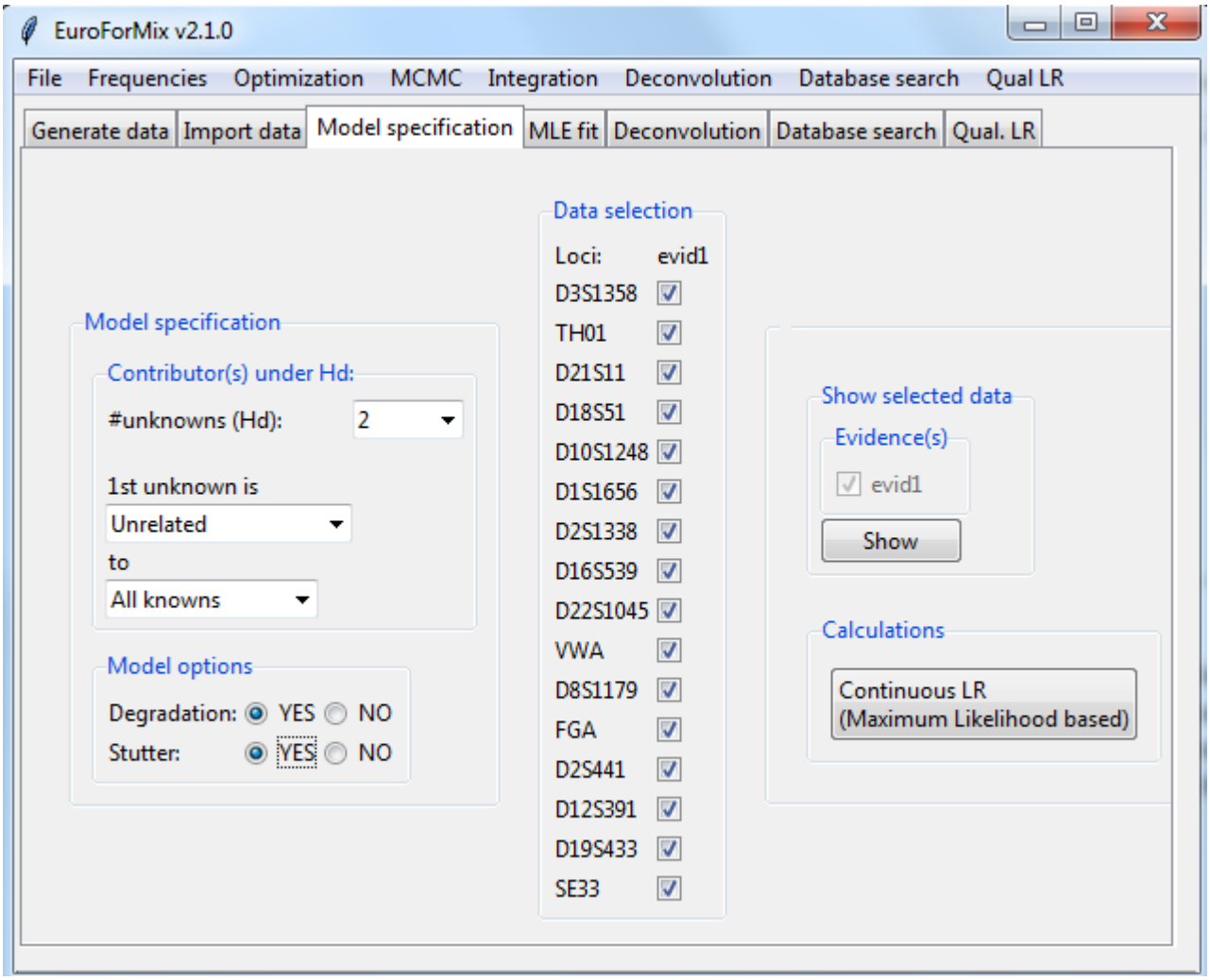


Figure 29: The figure shows the Model Specification page for doing **Deconvolution**. We condition on the suspect, and assume one unknown in the hypothesis. Our model assumes unknown backward-stutter rate, no allele drop-in, no theta-correction and no degradation.

- Description:

  - Deconvolution is applied for a specific hypothesis Hd as shown in **Figure 29**. Click "Quantitative LR (Maximum Likelihood based)" to proceed.
  - For a given optimized model, either Hd or Hp (**case Weight-of-Evidence**), the user must click on "Deconvolution" under "Further Action".
  - The deconvolution conditions on the optimized parameters (i.e. the MLE fit in **Figure 30**) for the quantitative model. Hence the deconvolution may handle multiple replicates, allele drop-in, drop-out, backward-stutter, degradation, theta-correction and relatedness (if considered) .

- The results tables will show different types of probabilities which are useful for deconvolution (**Top Marginal (Figure 31), All Joint (Figure 32), All Marginal (G) (Figure 33), All Marginal (A) (Figure 34)**). The probabilities gives a quantification of how "certain" different genotypes/alleles are for the different contributors at different loci".

Figure 30: The figure shows the optimized parameters (i.e. the MLE fit) for the quantitative model. The fitted model has the same "Further Action" possibilities as for "Weight-of-Evidence" and "Database search" in order to optimize the model.

- **Tables**:

    - **Top Marginal (Figure 31)**: Gives the top genotype with corresponding probability (most likely) marginalized for each contributor and each locus. “TopGenotype_Ck” gives the most likely genotype for contributor k (same order as in **MLE fit**), with corresponding probability under “probability_Ck”). The “ratioToNextGenotype_Ck” column gives the ratio of the largest probability (i.e. probability_Ck) to the second largest probability. The probabilities become one for known contributors.

    - **All Joint** (**Figure 32**): A ranked table of the combined genotype profiles for all contributors (C1,…,CK) with corresponding probabilities, given for each locus. The probabilities become one for known contributors.

    - **All Marginal (G) (Figure 33)**: A ranked table of the genotype profiles for each of the contributors, for each locus. The probabilities become one for known contributors.

    - **All Marginal (A) (Figure 34)**: A the ranked table of single alleles for each of the contributors for each locus. The probabilities become one for known contributors.

    - Maximum length of table is default 20.
        - Can be changed under ‘Deconvolution-> Set max listsize’ in toolbar.

    - The allele named as 99 represents alleles which are not in the evidence (or potential stutters if stutter-model is assumed).

- **Save table:**

    - The corresponding table will be exported to a tabulate-separated text-file.

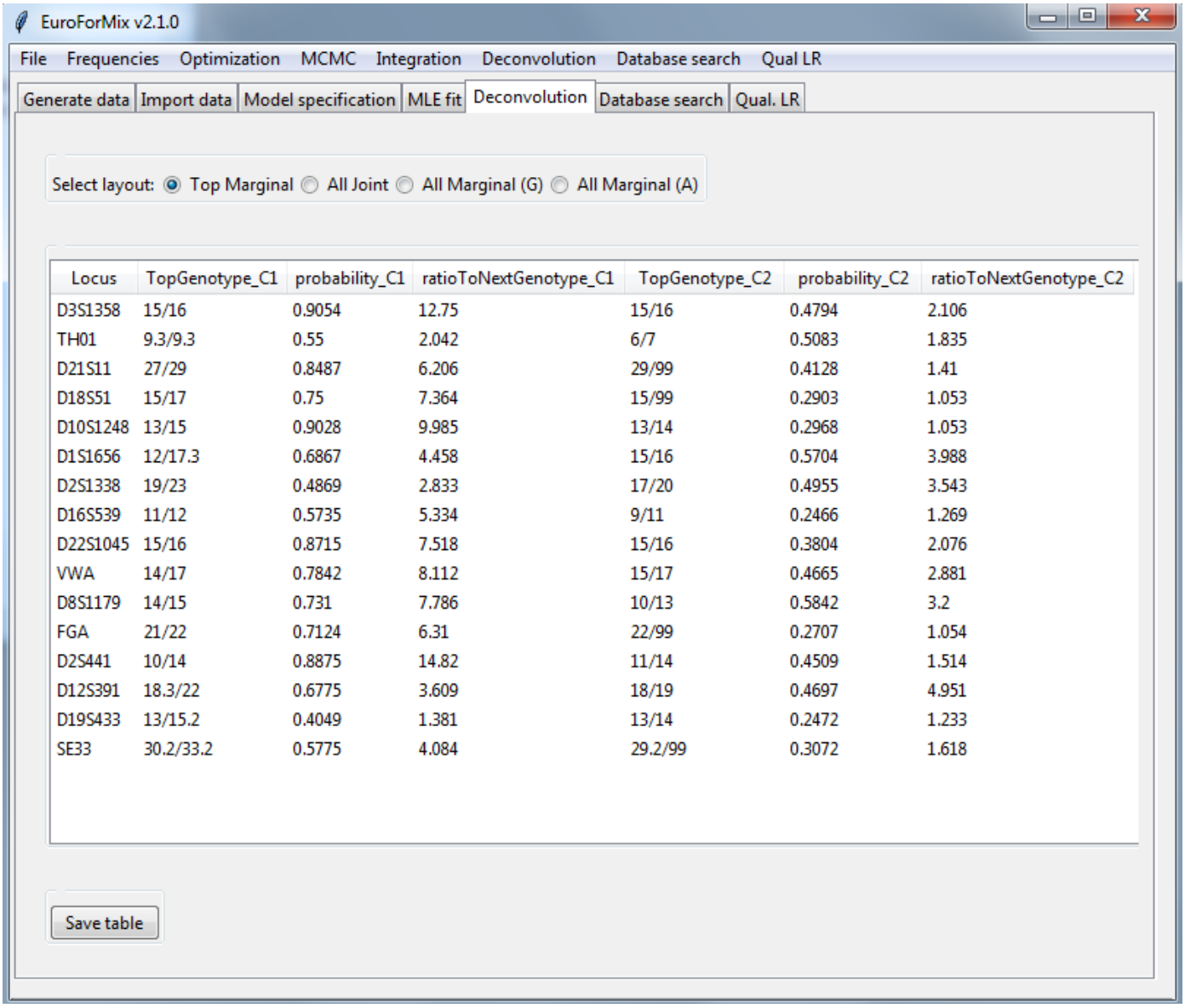


Figure 31: The figure shows **Top Marginal,** the top ranked genotypes (TopGenotype) for each contributor per loci, with corresponding probabilities **probability_Ck,** for each of the contributors, k=1,…,K. ratioToNextGenotype is the ratio of the largest probability (i.e. **probability_C**) to the second largest probability.

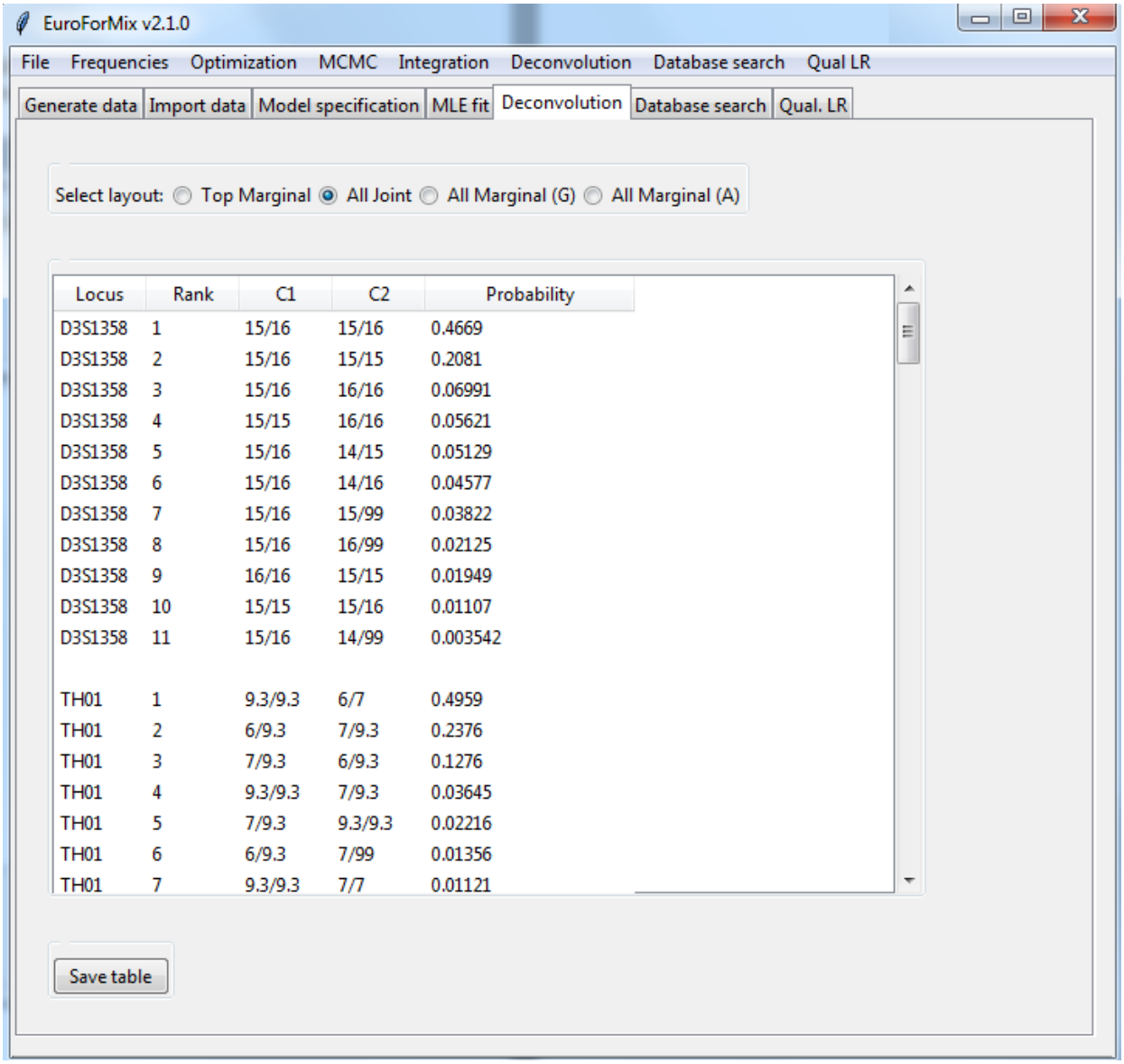


Figure 32: The figure shows **All Joint,** the ranked table of the combined genotype profiles for all the contributors combined (here, C1 and C2) for each locus.

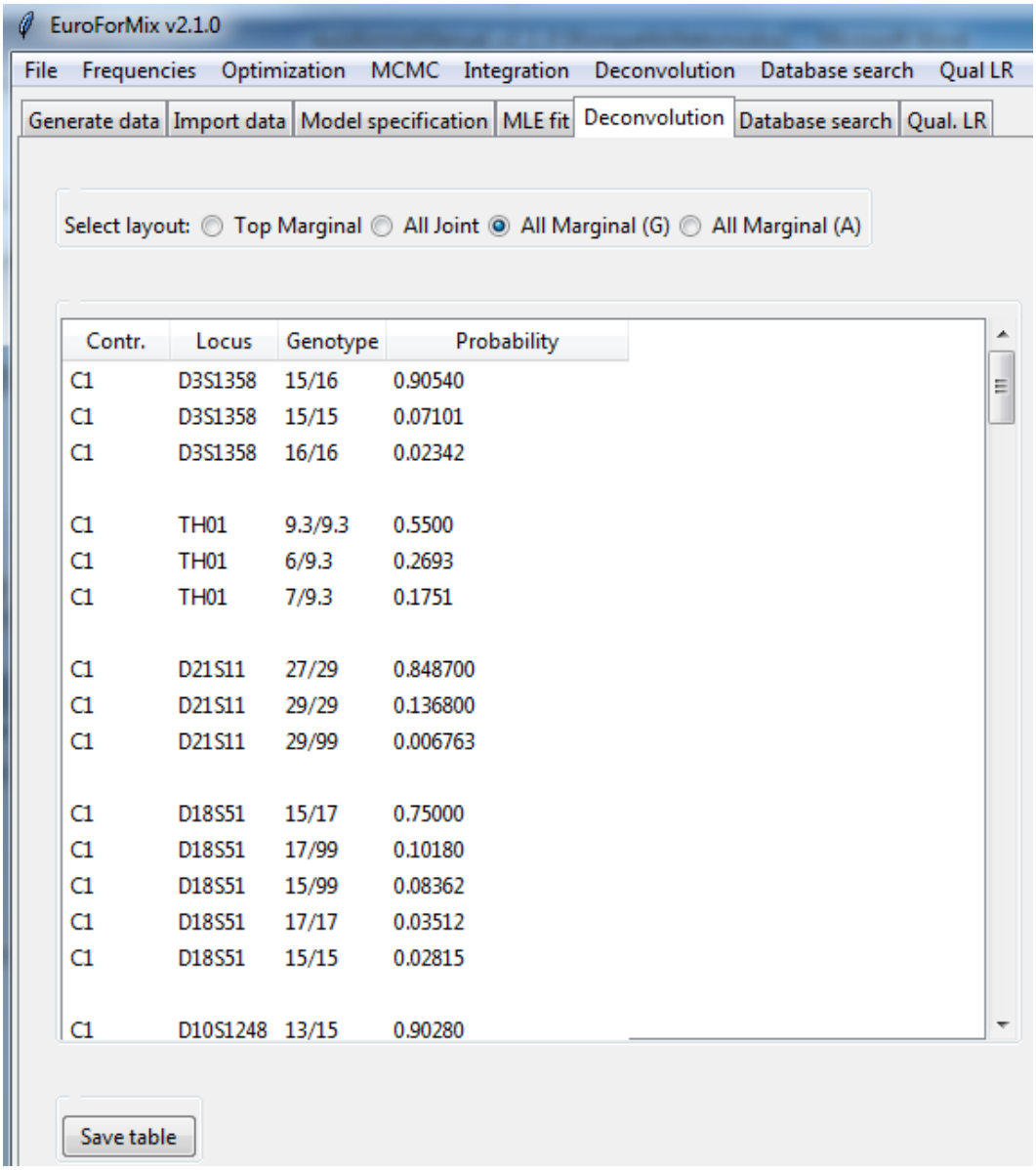


Figure 33: The figure shows **All Marginal (G),** the ranked table of the genotype profiles for each of the contributors for each locus.

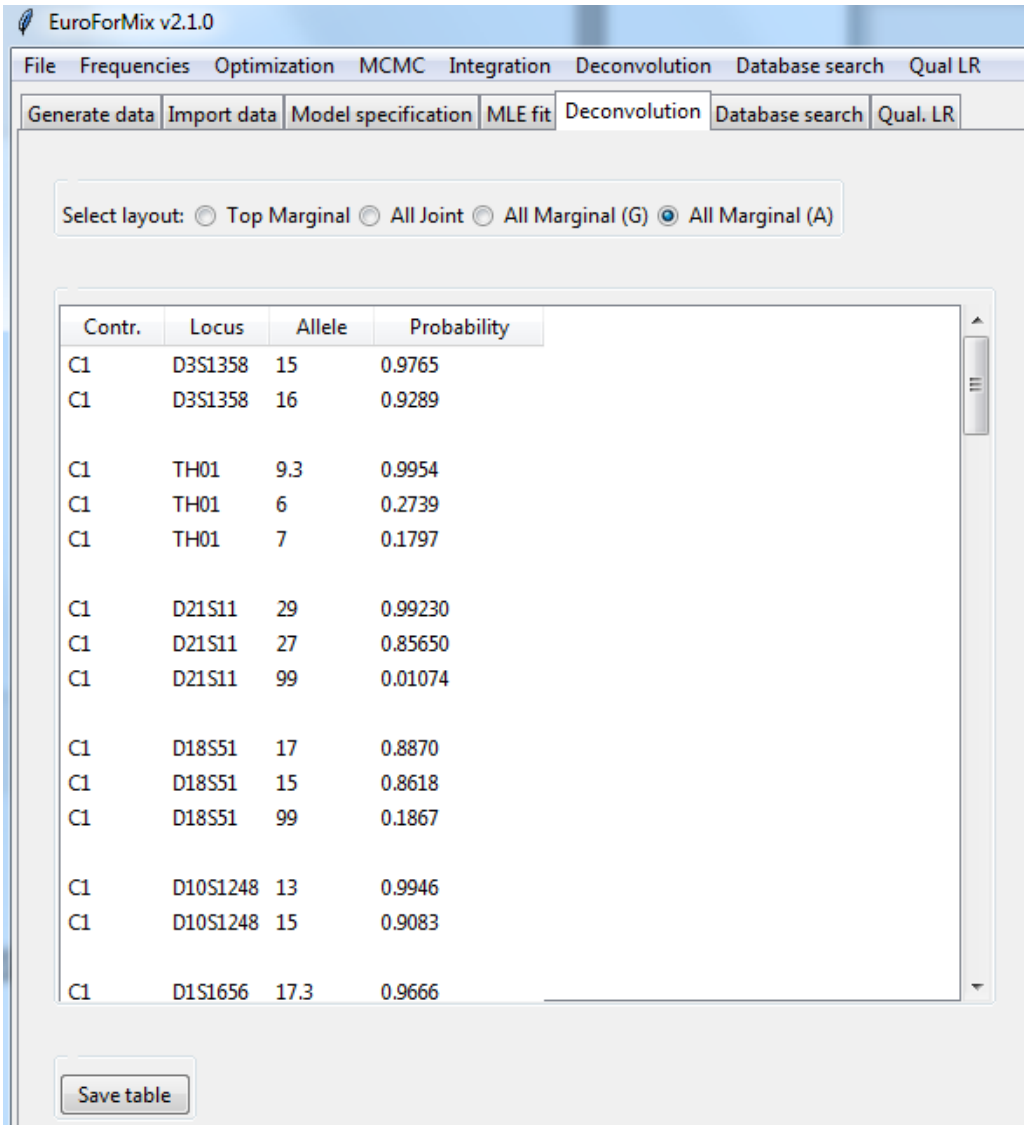


Figure 34: The figure shows **All Marginal (A),** the ranked table of single alleles for each of the contributors for each locus.

6- Database search:

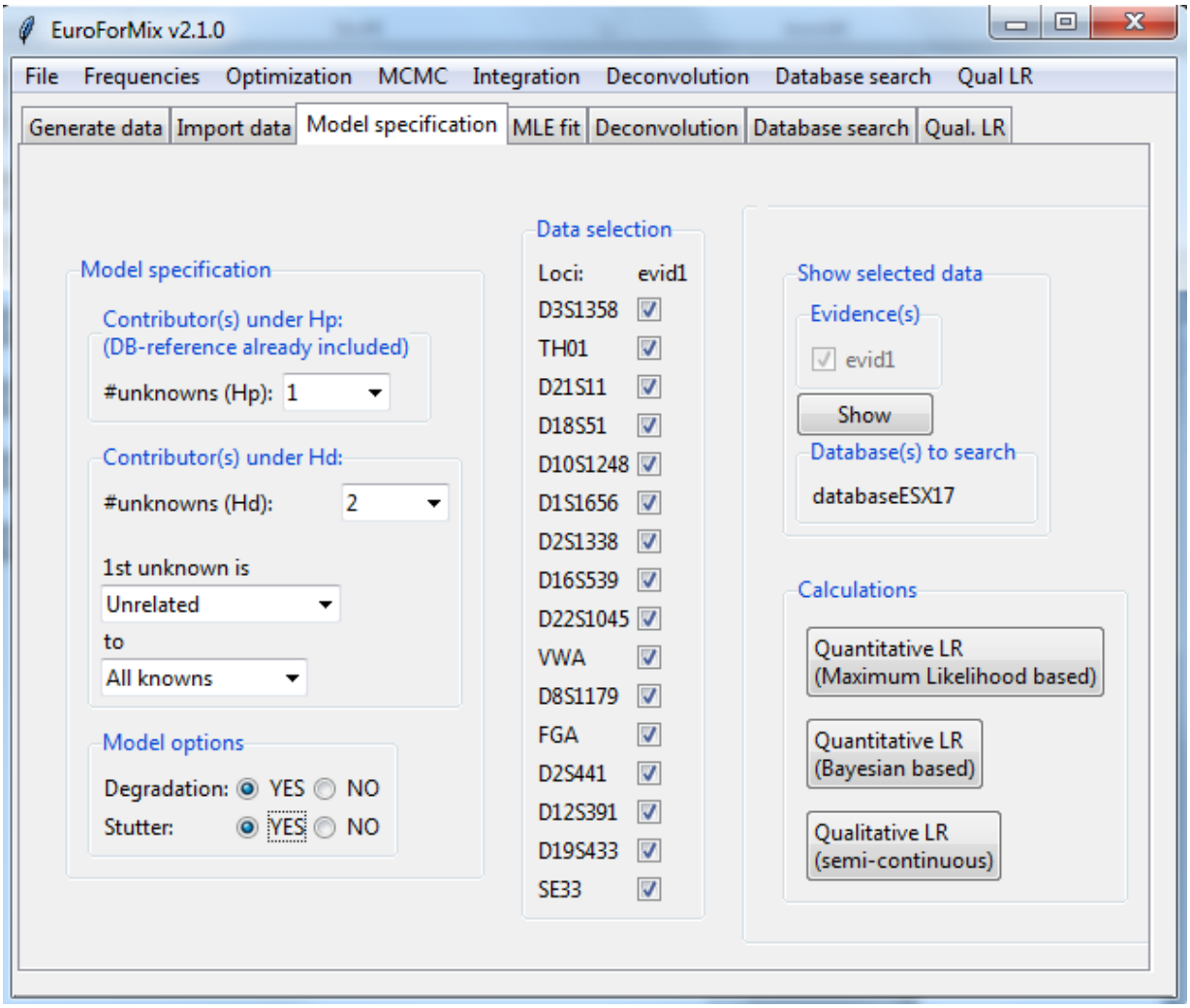


Figure 35: The figure shows the page of the model specification for doing database search on the database file "databaseESX17".

- Description:
  - The database to search must be loaded first from the Import data page.
  - Click the database search button from the Import data page which takes you to the Model specification page
  - The 'Database search' is very similar as the Weight-of-Evidence (**Figure 35**) with the only difference in that each individual in the reference-database is assumed to be 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 **'Quantitative LR' (Maximum Likelihood based** or **Bayesian based**)' calculation or ignoring the peak heights in a 'Qualitative LR' calculation.

- When selecting 'Quantitative LR': (Leads to the MLE fit page as seen in **Figure 36**)
  - 'Qualitative LR' is always calculated along with the 'Quantitative LR' values (see **Figure 37**).
  - The qualitative model assumes an allele drop-out parameter as 0.1 fixed and fst=0.
  - 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.
  - If "Quantitative LR (Maximum Likelihood based)" calculation is used, the optimized parameters under the Hd -hypothesis are first shown (see **Figure 36** where we have used threshold=150, fst=0 and no allele drop-in).

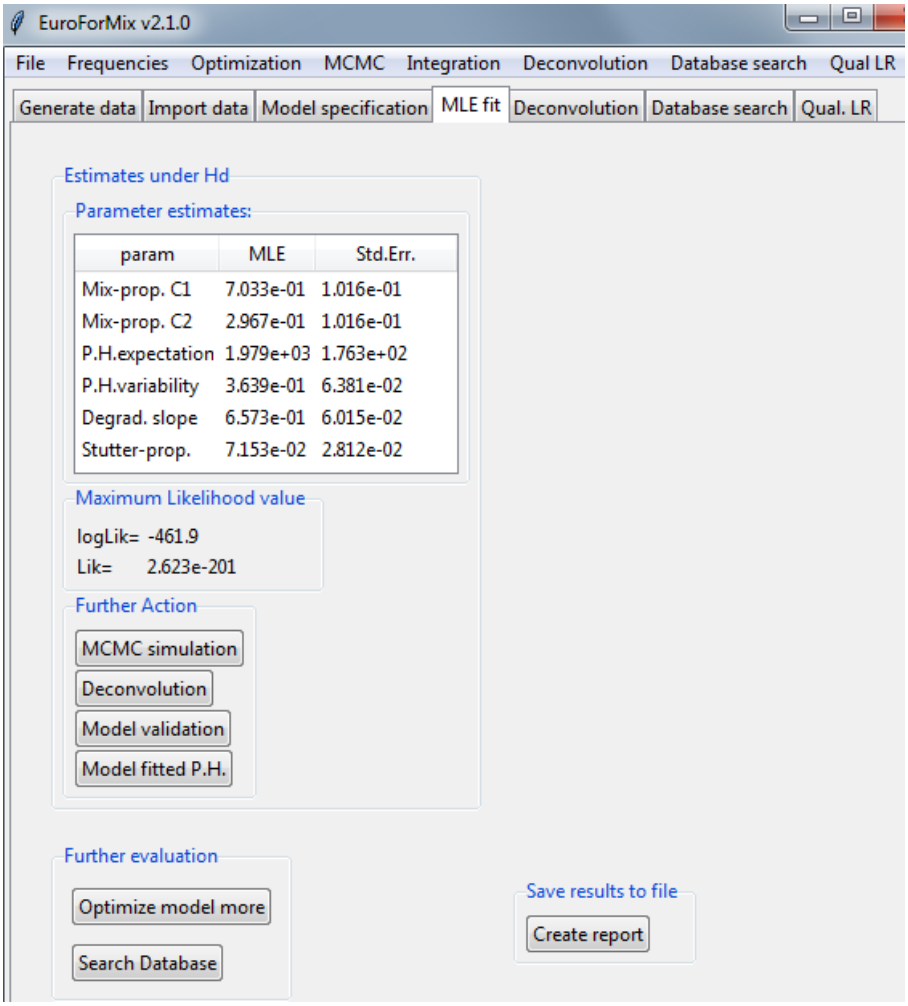


Figure 36: The figure shows the optimized parameters (i.e. the MLE fit) for the quantitative model under Hd (with specifications as given in **Figure 35**). The fitted model has the same "Further Action" possibilities as for "Weight-of-Evidence" and "Deconvolution". The user must push the "**Database search**" button to carry out the actual database searching.

- When selecting 'Qualitative LR' from the 'database search page:
  - 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 using median of the '*allele drop-out probability given number of observed alleles*' distribution.
  - The 'Quantitative LR' calculation is ignored.

- Note:
  - The 'Quantitative LR' calculation is based on the quantitative **model** as given in the euroformix paper and can handle allele drop-in, drop-out, degradation, backward-stutter and relatedness under Hd (if applied).
  - Quantitative LR (Bayesian based) is not possible to use for replicates.
  - The reason for showing the MLE fitted parameters under Hd (see **Figure 36**) for "Quantitative 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/she can go back and change the model specification.

- **Database search table** (**Figure 37**):

  - '**Reference name'** is name of individuals given in the reference-database.

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

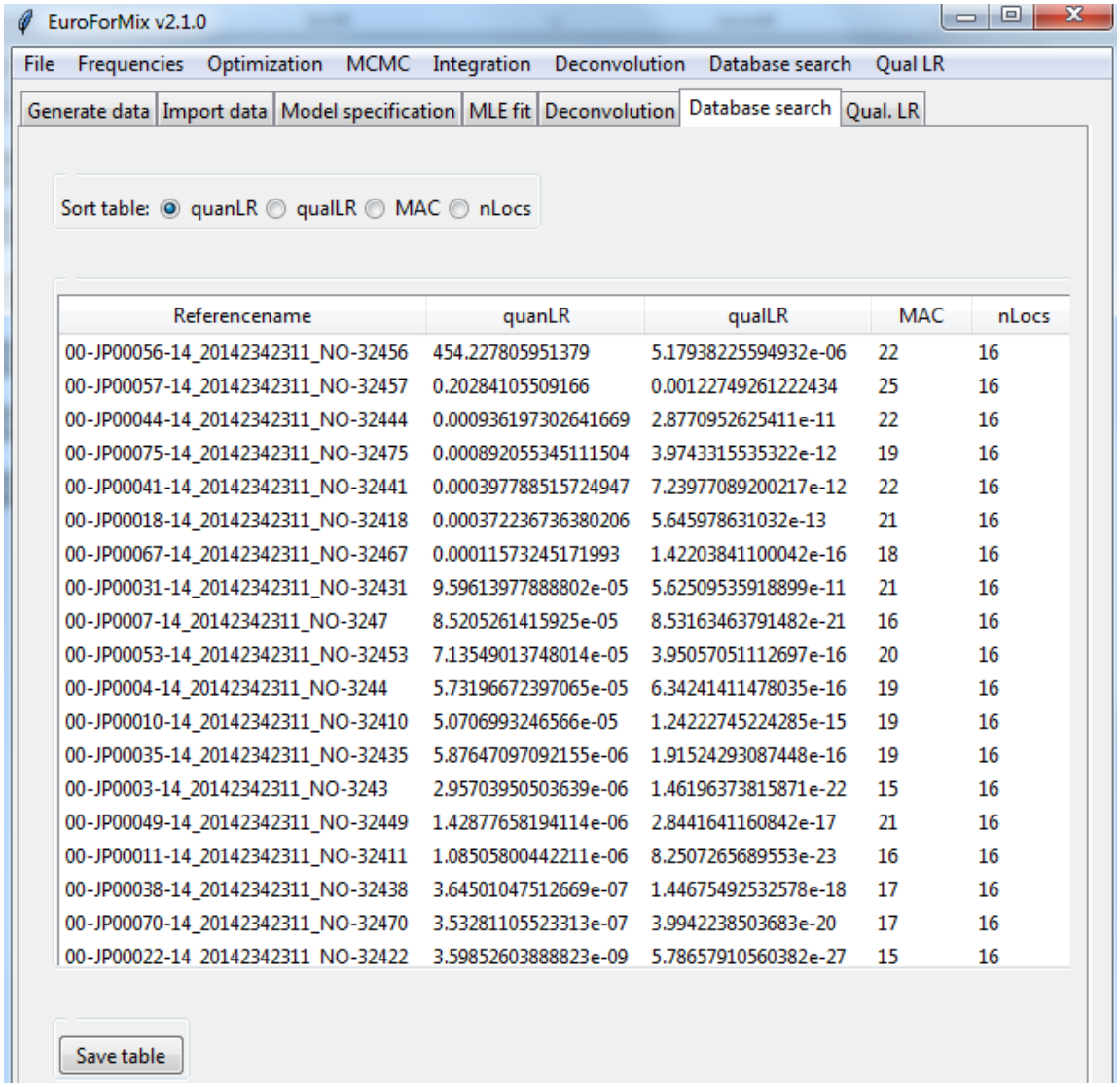
  - **qualLR** (Qualitative LR (semi-continuous model))
    - Parameter for dropout probability is based on the median of 2000 samples from the 'distribution of dropout-probability'.
      - Number of required samples may be changed under 'Qual LR' in toolbar.
      - Dropout probability is fixed to 0.1 when searched with "Quantitative LR".
      - 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.

  - **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 in toolbar 'Database search->Set maximum view-elements'.
  - Setting fst>0 may be very time-consuming since we require that individual 'j' is a known non-contributor under Hd, and hence the likelihood for Hd is calculated for each individual in database.

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



Referencename	quanLR	qualLR	MAC	nLocs
00-JP00056-14_20142342311_NO-32456	454.227805951379	5.17938225594932e-06	22	16
00-JP00057-14_20142342311_NO-32457	0.20284105509166	0.00122749261222434	25	16
00-JP00044-14_20142342311_NO-32444	0.000936197302641669	2.8770952625411e-11	22	16
00-JP00075-14_20142342311_NO-32475	0.000892055345111504	3.9743315535322e-12	19	16
00-JP00041-14_20142342311_NO-32441	0.000397788515724947	7.23977089200217e-12	22	16
00-JP00018-14_20142342311_NO-32418	0.000372236736380206	5.645978631032e-13	21	16
00-JP00067-14_20142342311_NO-32467	0.00011573245171993	1.42203841100042e-16	18	16
00-JP00031-14_20142342311_NO-32431	9.59613977888802e-05	5.62509535918899e-11	21	16
00-JP0007-14_20142342311_NO-3247	8.5205261415925e-05	8.53163463791482e-21	16	16
00-JP00053-14_20142342311_NO-32453	7.13549013748014e-05	3.95057051112697e-16	20	16
00-JP0004-14_20142342311_NO-3244	5.73196672397065e-05	6.34241411478035e-16	19	16
00-JP00010-14_20142342311_NO-32410	5.0706993246566e-05	1.24222745224285e-15	19	16
00-JP00035-14_20142342311_NO-32435	5.87647097092155e-06	1.91524293087448e-16	19	16
00-JP0003-14_20142342311_NO-3243	2.95703950503639e-06	1.46196373815871e-22	15	16
00-JP00049-14_20142342311_NO-32449	1.42877658194114e-06	2.8441641160842e-17	21	16
00-JP00011-14_20142342311_NO-32411	1.08505800442211e-06	8.2507265689553e-23	16	16
00-JP00038-14_20142342311_NO-32438	3.64501047512669e-07	1.44675492532578e-18	17	16
00-JP00070-14_20142342311_NO-32470	3.53281105523313e-07	3.9942238503683e-20	17	16
00-JP00022-14_20142342311_NO-32422	3.59852603888823e-09	5.78657910560382e-27	15	16

Figure 37: The figure shows the table from the database search with specifications as given in **Figure 35** based on **'Quantitative LR' (Maximum Likelihood based)"** calculations. The references are sorted due to the contLR's.

7- Qual. LR: 'Qualitative model'

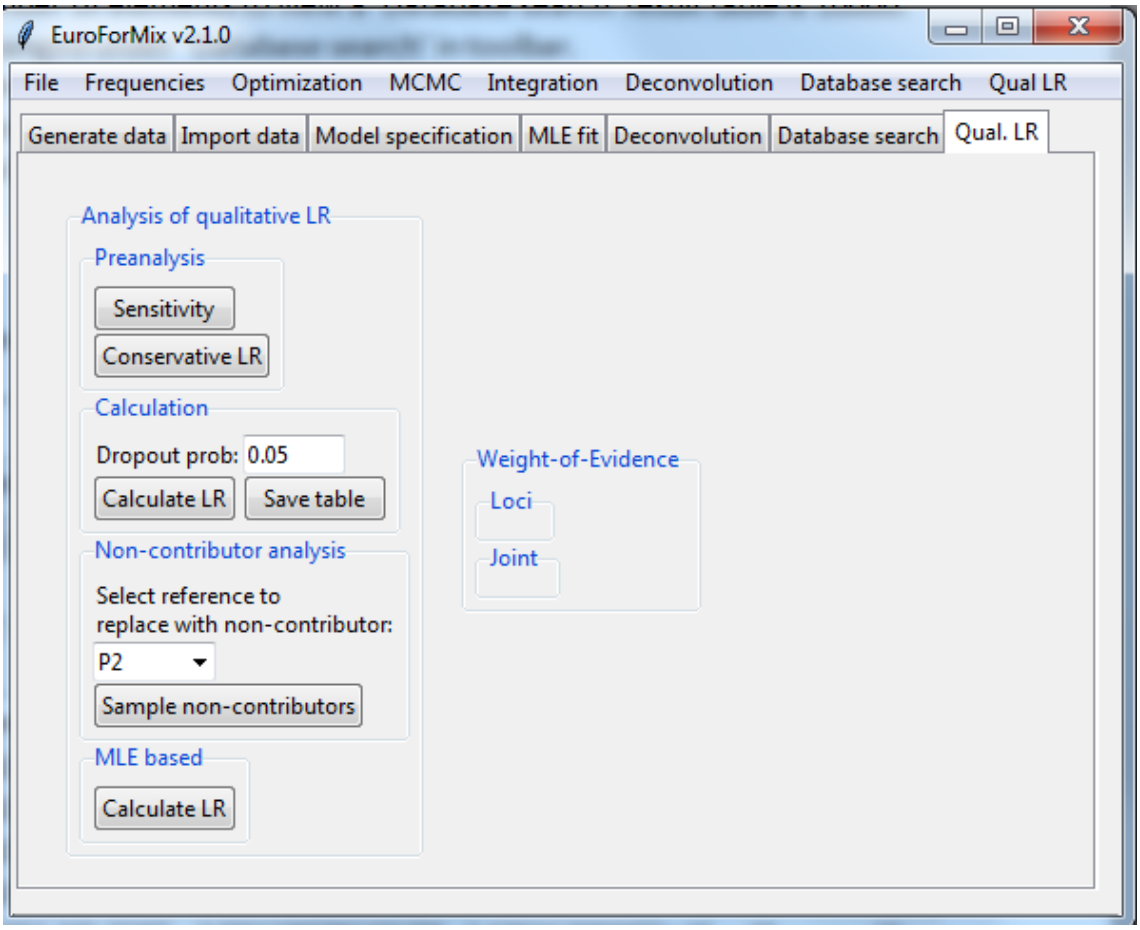


Figure 38: The figure shows the page where the weight-of-evidence evaluation based on the qualitative model is carried out.

- Description:

  - From 'Import data' page, check evidence evid1 and reference P2, and press 'Weight-of-Evidence' button which leads to the 'Model specification' page. Under model specifications, construct Hp: P2+2 unknown contributors vs Hd: 3 unknown contributors. Then select the 'Qualitative LR' button which leads to the 'Qual. LR' page shown in **Figure 38**.

  - This module samples from the distribution of the '*allele drop-out probability given number of observed alleles*' to evaluate the qualitative LR automatically.
    - Note: the model will crash if there are too many alleles compared to the number of contributors – always check that the model specification is reasonable

- Also a sensitivity plot as a function of allele-dropout probability and a non-contributor sampling analysis is implemented (see **Figure 42)**.

PREANALYSIS

- **Sensitivity:**

  - Plots the log10LR as a function of allele-dropout probability (see **Figure 39**).
    - This is the same function as in LRmix Studio.
    - 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).

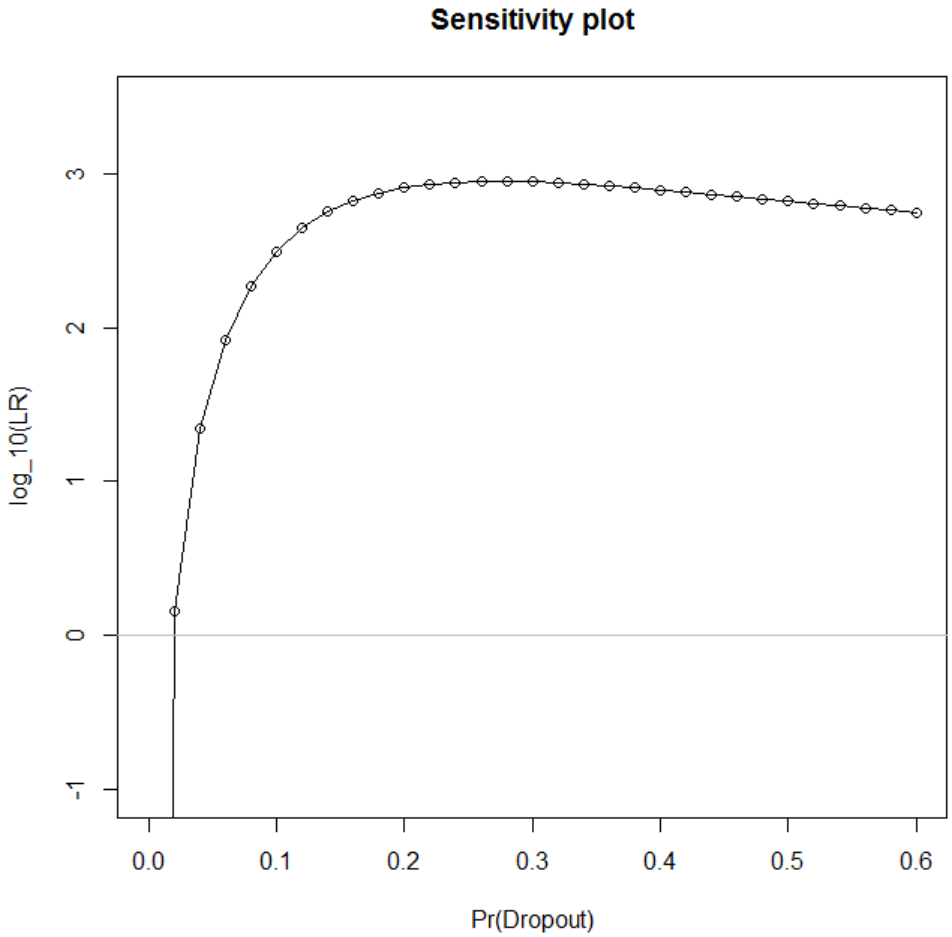


Figure 39: The figure shows the plot of Weight-of-evidence (Likelihood Ratio) as a function of allele drop-out probability.

- **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 40** and **Figure 41**).
    - 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 will give similar results as in LRmix Studio.
        - 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->Set required samples in dropout distr.'

    - 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] "Total number of observed alleles for sample(s):"
      x
 52
[1] "For evidence evid1:"
[1] "Estimating quantiles from allele dropout distribution under Hp..."
           x
 0.1122676
 0.2290957
 0.3524354
[1] "Estimating quantiles from allele dropout distribution under Hd..."
           x
 0.1445842
 0.2602897
 0.3769952
      5%  95%
[1,] 0.11 0.35
[2,] 0.14 0.38
```

Figure 40: The plot shows the sampled 5%, 50% and 95% quantiles of the distribution of the '*allele drop-out probability given number of observed alleles*' for each of the hypotheses.

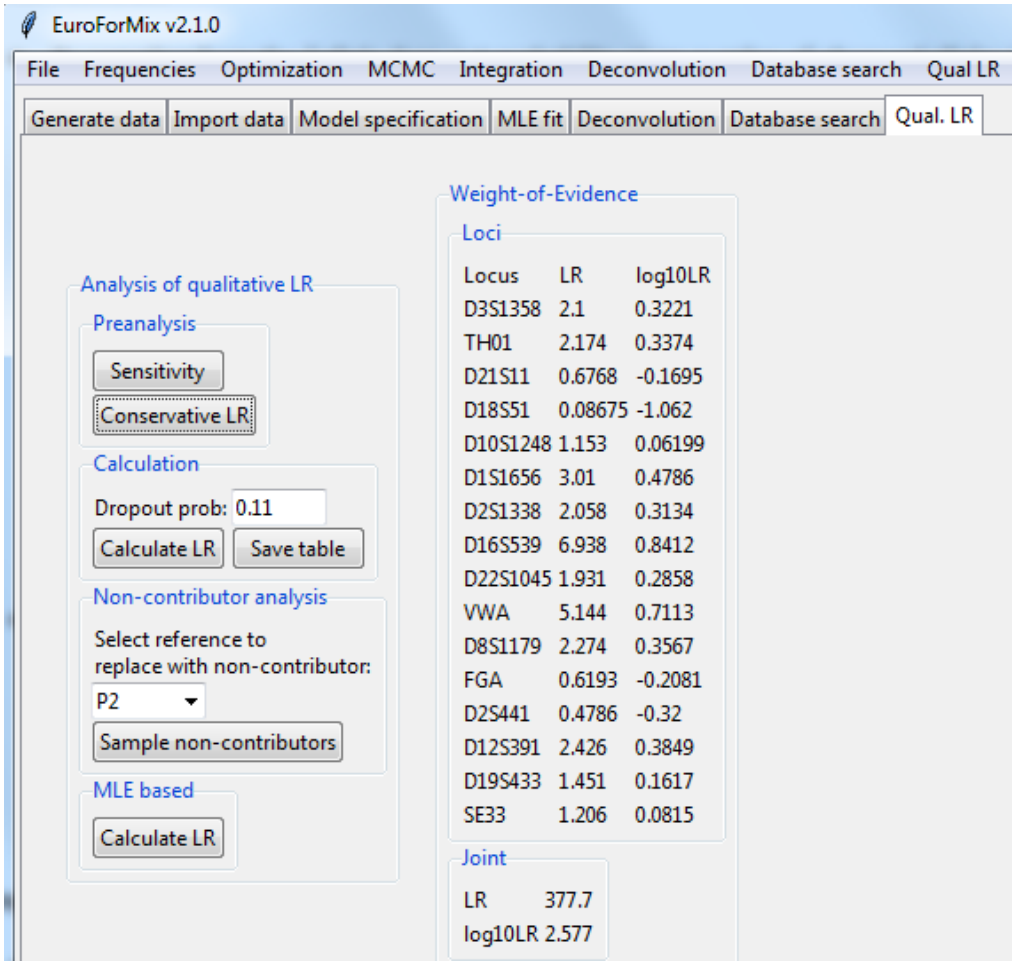


Figure 41: The plot shows the conservative Weight-of-Evidence values (Likelihood Ratios) after pushing the "**Conservative LR**" button. The most conservative estimated allele drop-out probability-quantile from **Figure 40** was the 5% quantile under Hd which gave 0.11. Hence the table in this plot shows the LR inserted for this value.

CALCULATION

- **Dropout prob:**
  - The user may specify the assumed value of the 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.

NON-CONTRIBUTOR ANALYSIS (Postanalysis)

- **Select reference to replace with non-contributor:**

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

- **Sample non-contributors:**

    - Random non-contributor 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.
        - The mean, standard errors of LR and log10LR-quantiles (50%, 95%, 99%, max) are printed out to R-console.
        - A plot of the cumulative distribution of log10LR will be shown (see **Figure 39**).
        - Number of non-contributors can be changed under 'Database search->Set number of non-contributors' in the toolbar.
    - If weight-of-evidence has been calculated:
        - The reporting LR for the "replaced reference" is superimposed as a blue line to the plot (see **Figure 39**).
        - The discriminatory metric (log10LR-q99%) is printed out to R-console.
    - Note: Precalculations are always carried out previous to the non-contributor sampling, therefore the number of non-contributors are only limited to make the plot.

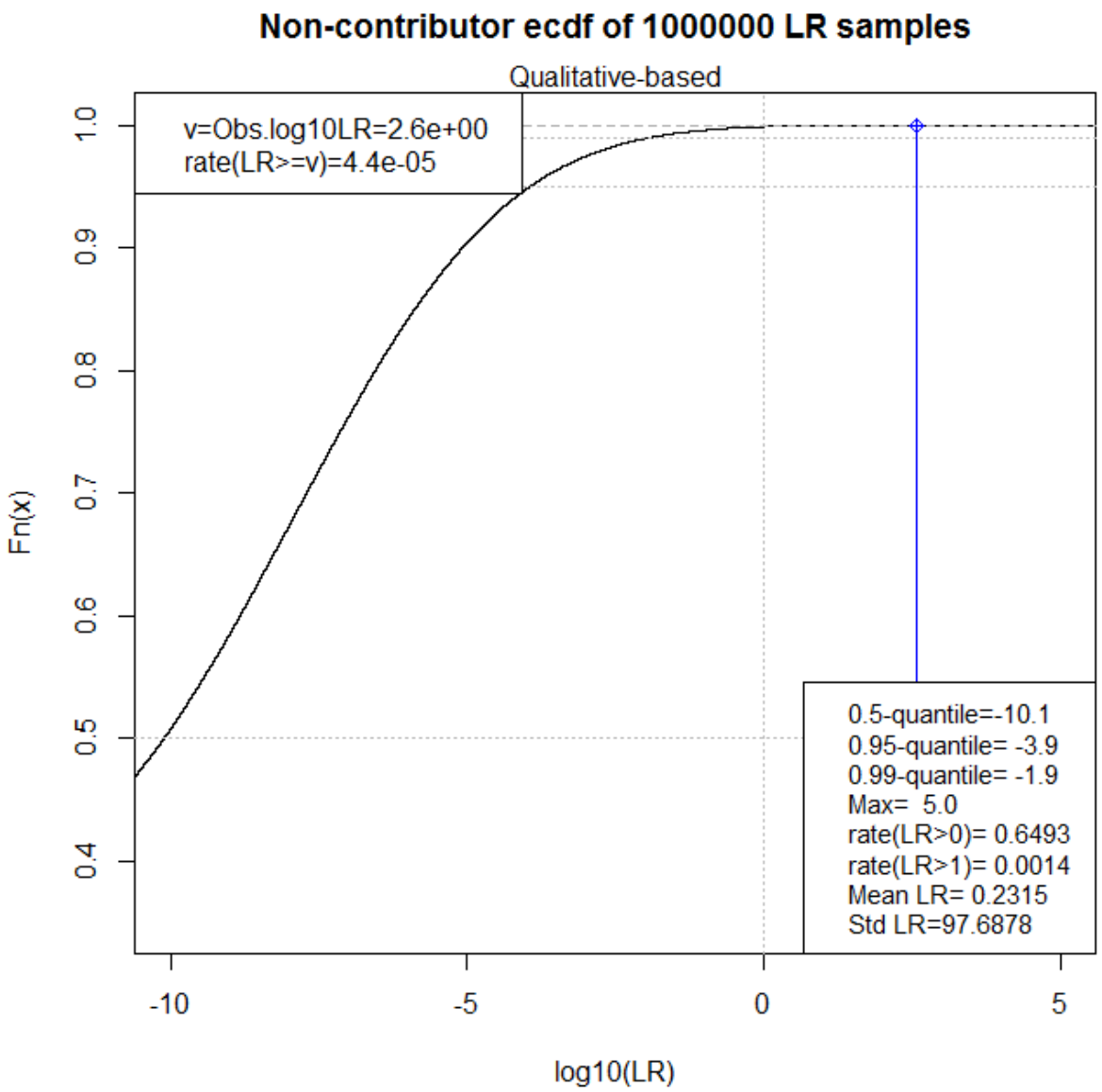


Figure 42: The figure shows a cumulative distribution of 1e6 log10LR of non-contributors, where each sample is based on replacing the "Suspect" in hypothesis Hp with a random man from the population. The reporting LR for the replaced reference (i.e. "Suspect in this case) is superimposed as a blue line to the plot. The mean and standard errors of LR, proportion of LR greater than zero and one, and log10LR-quantiles (50%, 95%, 99%, max) based on the simulated non-contributors are given in the plot as well. In upper left box, the proportion of non-contributors LR exceeding the reported LR (v) is given.

Qualitative MLE-based approach (alternative analysis)

- Description
  - This functionality will follow the maximum likelihood approach for estimating the dropout probability for each of the hypotheses.

- **Calculate LR**
  - The LR value based on the maximum likelihood estimates, with the corresponding estimated dropout probabilities for each of the hypotheses, is given in a pop-up window (see **Figure 43**).

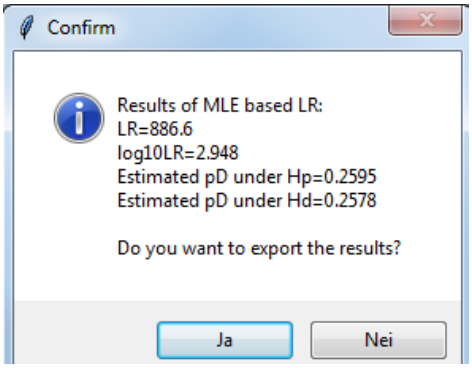


Figure 43: The calculated LR value based on the maximum likelihood estimates, with the corresponding estimated dropout probabilities for each of the hypotheses indicated.

8- Generate data: 'from the quantitative model'

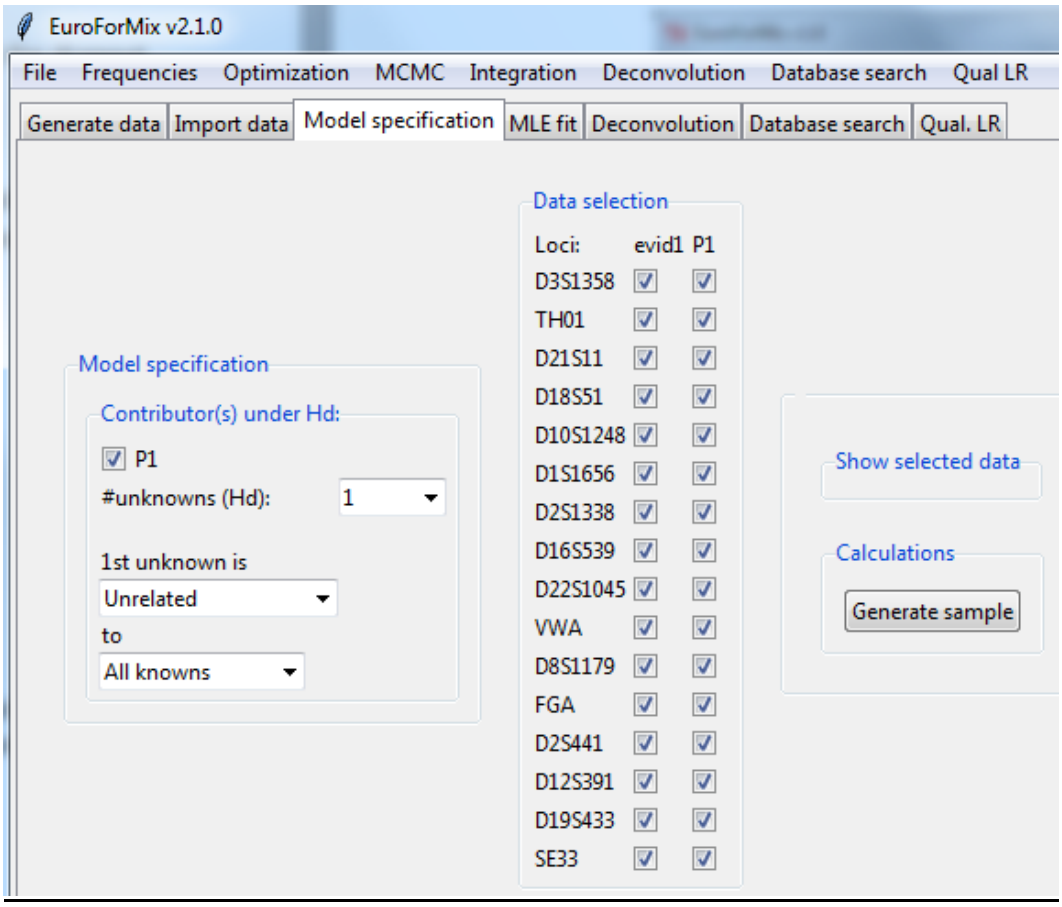


Figure 44: The figure shows the Model specification page for generating allele with corresponding peak heights from the quantitative model for a given specified model. From here we will generate data which are contributed P1 and an unknown individual. We assume a detection threshold (LOD) of 150 RFU and no allele drop-in is considered.

- Description:

  - To generate data, the user must first specify the assumptions (hypothesis and known parameters) in the quantitative model. **NB: The relatedness module will be ignored!**
  - The module will generate alleles using the population frequencies and simulates peak heights for a specified hypothesis (see **Figure 44**) using the quantitative model.
  - The generation may simulate allele-dropout, drop-in (with a peak height model), degradation and backward-stutter (see **Figure 45**).
    - Allele-dropout is indirectly simulated if the peak height is below the defined threshold.

- **Parameters**:
  - P.H.expectation (mu): Expectation of the peak height for a single heterozygote (Mix-prop=1) allele without degradation
  - P.H.variability (sigma): Coefficient of variation of the peak height for a single heterozygote (Mix-prop=1) allele without degradation
  - Stutter-prop (xi): A global parameter related to backward-Stutter proportion. The expected fraction of peak height that are stutter.
  - Degrad.slope (beta): A global parameter related to the degree of degradation as a function of fragment length (**kit must be selected**). Value 1 is no degradation, and lower values as for instance 0.6 is much degradation. Default is 1.
  - Mix-prop. Ci (mxi): Mixture-proportion for contributor 'i'. Default is (K:1)/K, with K number of contributors.

- 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 45**).

- **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.

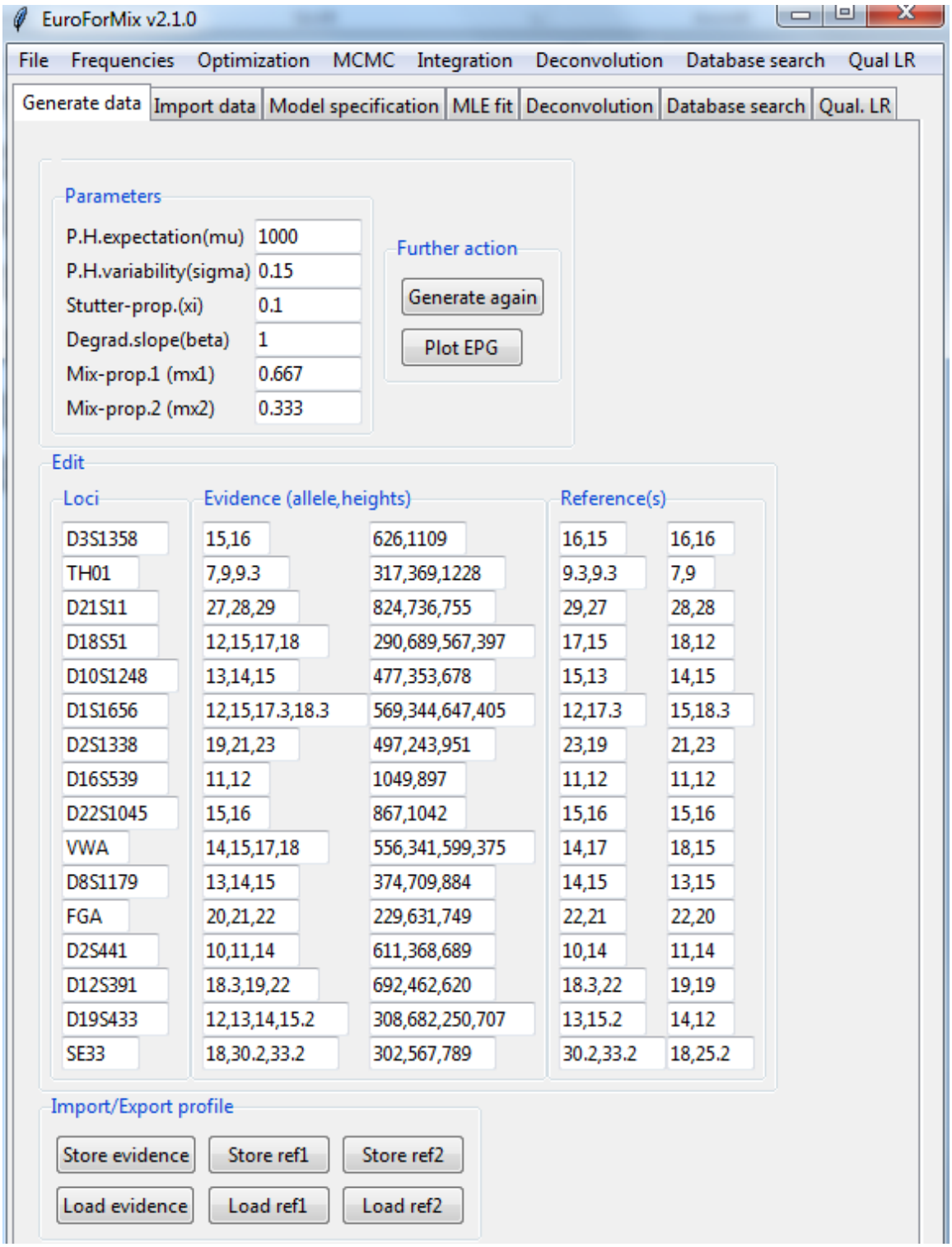


Figure 45: The figure shows the Generate data page which shows the generated alleles and corresponding peak heights (under **Evidence**) for the given selected set of parameters under **Parameters**. The true contributors are given under **Reference(s)**.

(A) Mathematical details

Exact random allele sharing with a evidence profile

Consider the observed allele vector at marker *i* in the evidence as $M_i = (A_{i1}, \dots, A_{iI})$ with corresponding allele frequencies $p_{i1}, \dots, p_{iI}$. The number of alleles the defendant shares with the evidence for this marker is denoted Z_i. Let $S_i = p_{i1} + \cdots + p_{iI}$ be the sum of the allele frequencies at marker *i*. Then a direct argument gives (calculations assume Hardy Weinberg and Linkage Equilibrium)

$$P(Z_i = 0) = (1 - s_i)^2$$
$$P(Z_i = 1) = 2s_i(1 - s_i)$$
$$P(Z_i = 2) = s_i^2$$

Let $Z = Z_1 + \cdots + Z_I$ be the total number of alleles shared and $\boldsymbol{w} = (w_1, \dots, w_I)$ where $w_i = \{0,1,2\}$ one of these values. Then a for a $k = 0, \dots, I, \dots, 2I$,

$$P(Z = k) = \sum_{\text{all permutations in } \boldsymbol{w}: \sum w_i = k} \prod_{i=1}^{I} P(Z_i = w_i)$$

Here "all permutations" means all possible ordered combinations of the elements in the vector $\boldsymbol{w}$. Note here that RMNE simplifies to $P(Z = 2I) = \prod_{i=1}^{I} P(Z_i = 2)$.