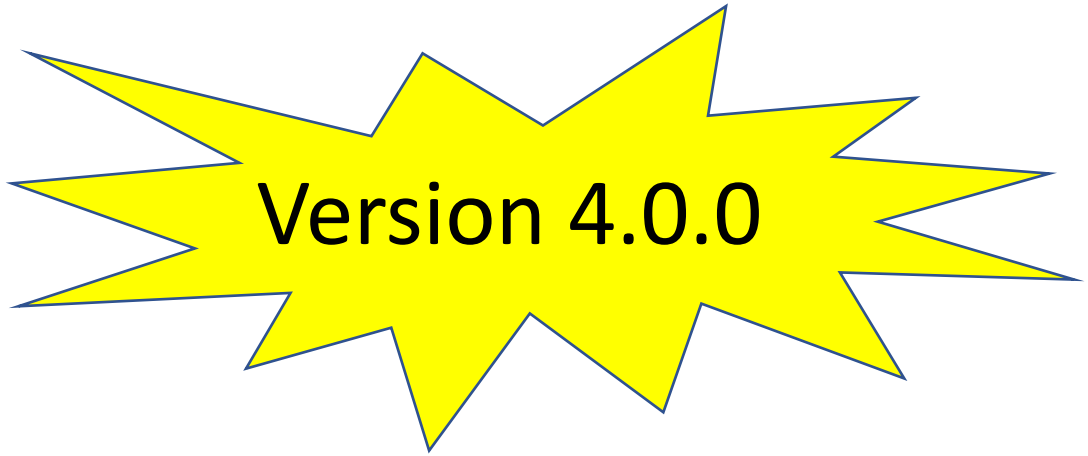
New update of EuroForMix



By Øyvind Bleka
Department of Forensic Sciences, Oslo University Hospital, Oslo, Norway

Overview

- **BETTER speed**: Possibly 4-6x faster for 4 unknowns
  - For MLE, MCMC, INTEGRATION
  - Also for Qualitative model (forensim R-package not used)

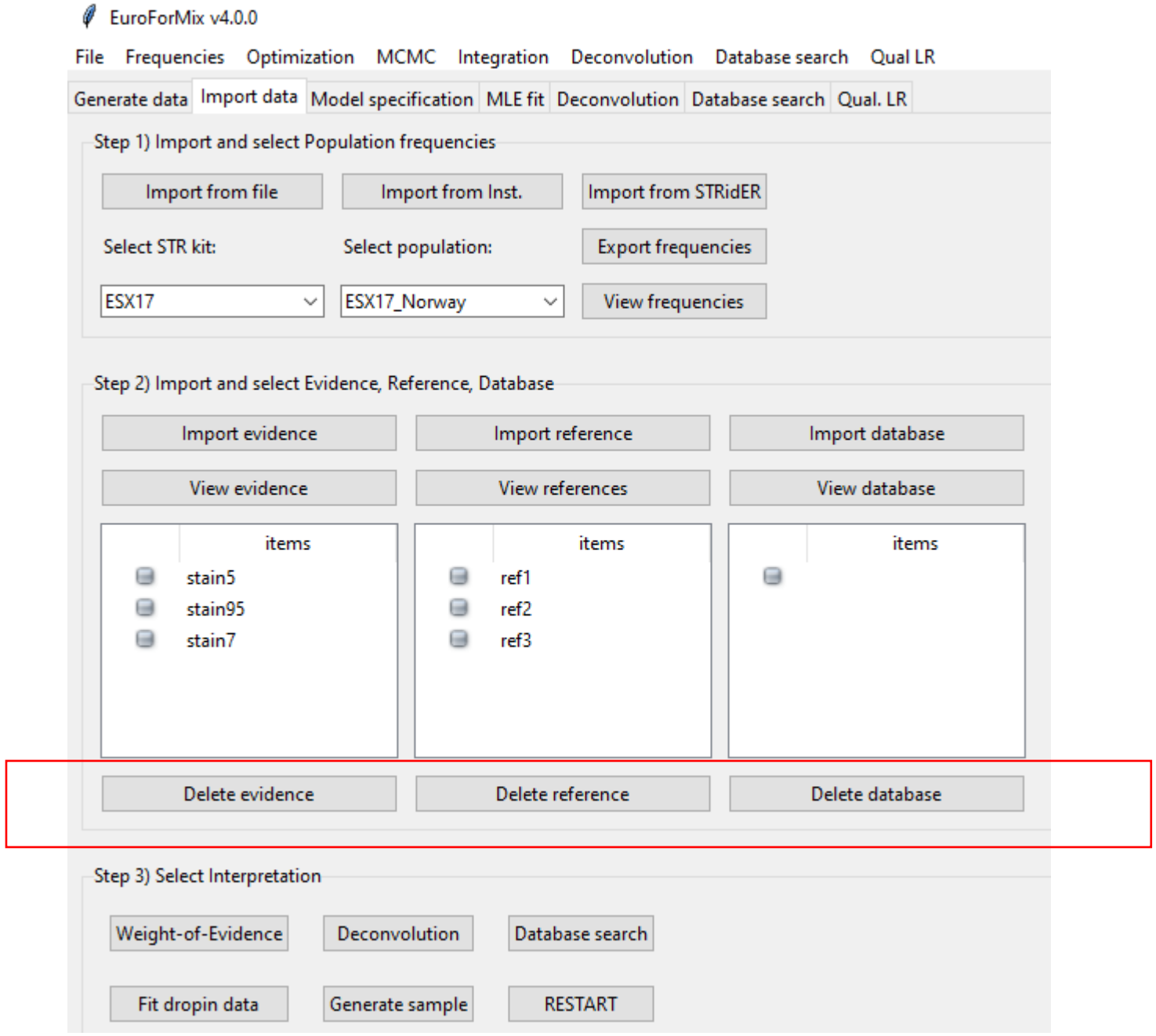
- Better functionalities for Bayesian based LR (BayesFactor/quantile):
  - Bayesian integral is more robust and must be done after the MLE calculation
  - MCMC trace plot of BayesFactor and Conservative LR (quantile)

- New GUI layouts (next slides)

- New functions which simplify terminal interface
  - A Rmarkdown tutorial is also available.

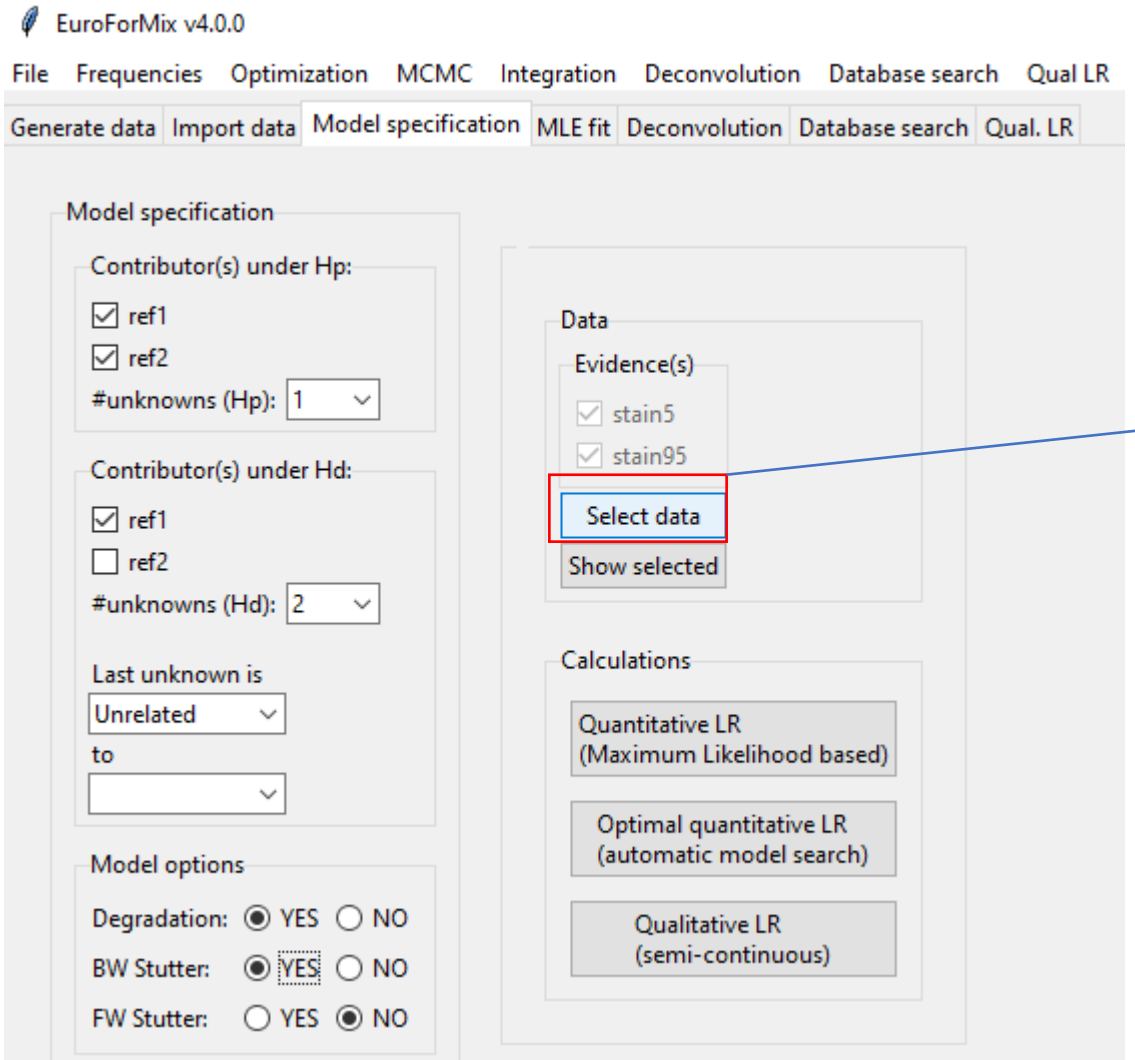
++ more

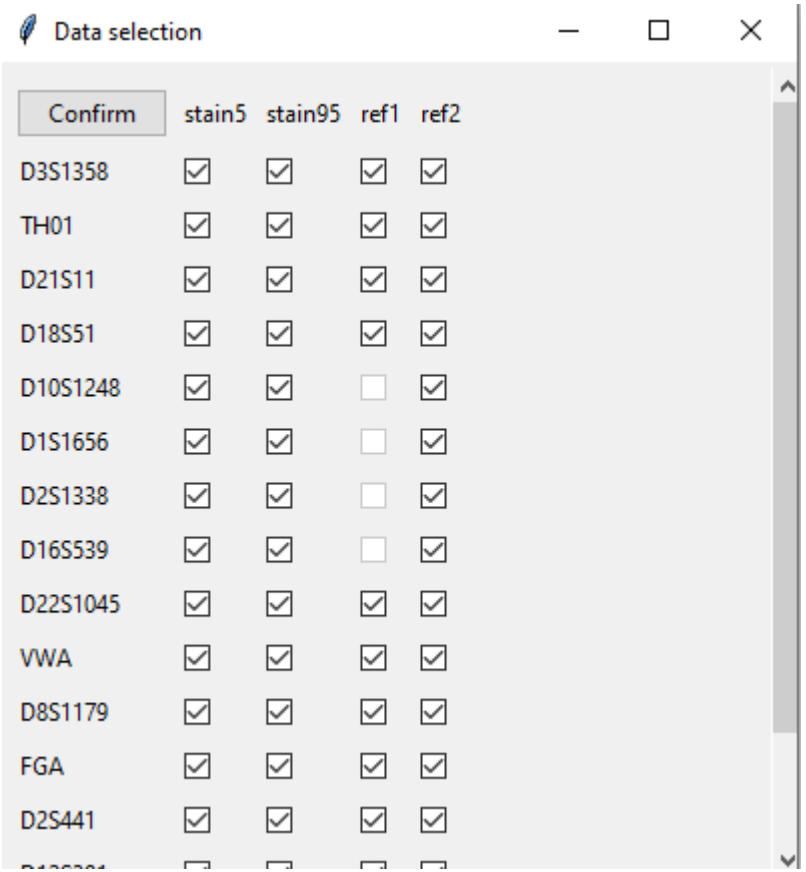
Important/useful changes made to the GUI

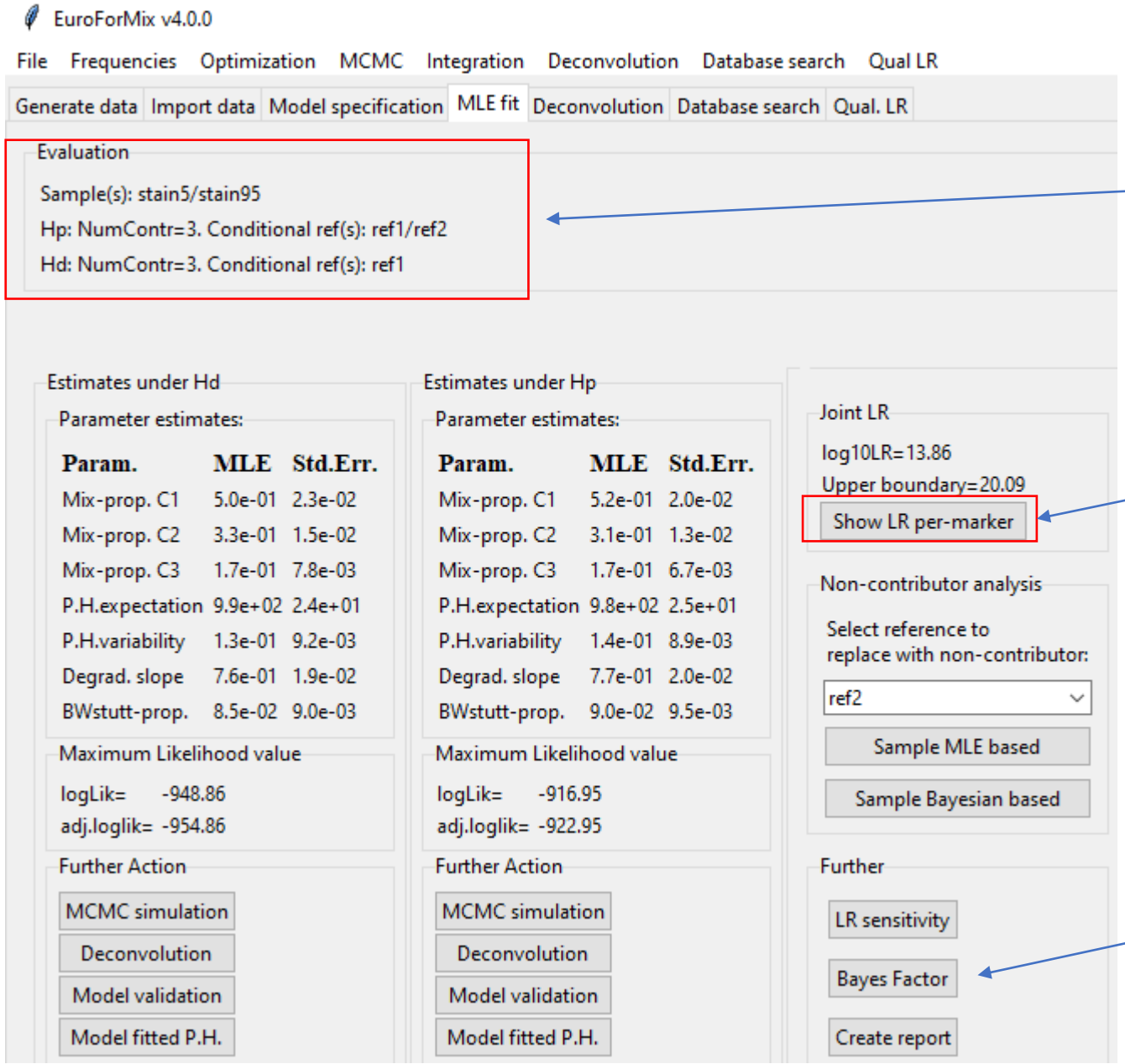


Delete profiles from GUI

Select data in a separate window





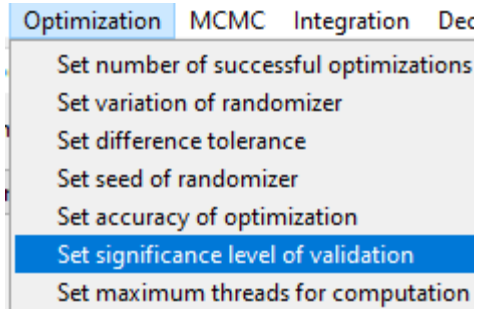


New header showing evaluated Evidence and hypotheses

Show LR per marker in separate window

Bayes Factor (Bayesian integral) moved here: Utilizes MLE information

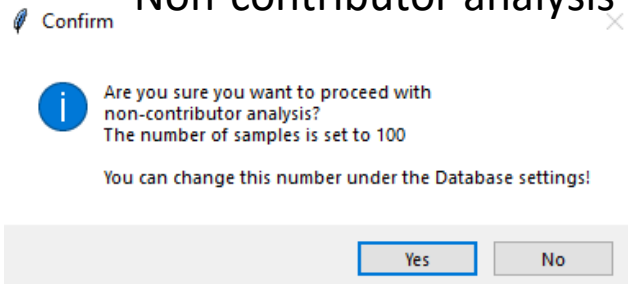
User-friendly changes

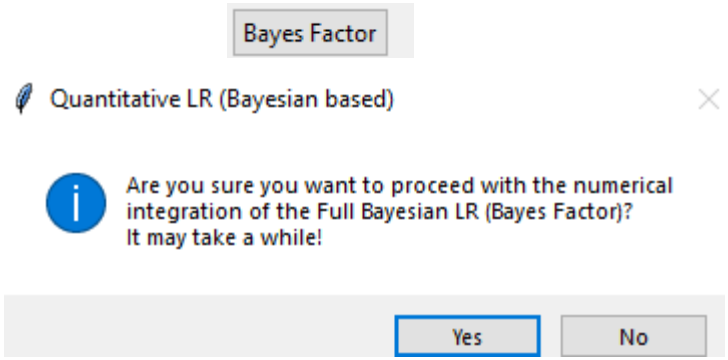


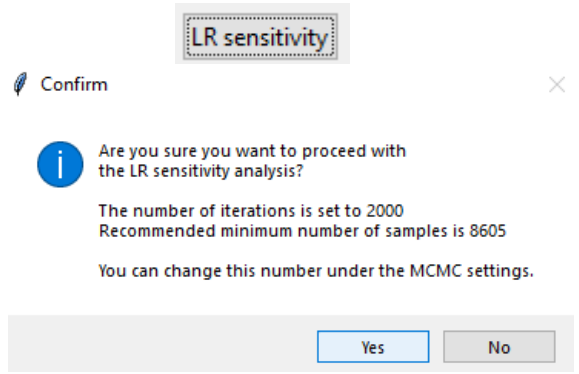
- Model validation:
  - Entering significance level **no longer needed**.
    - Can be changed under “Optimization toolbar”
  - Number of significant points outside envelope is now indicated in the report, <u>but only if user has conducted the analysis.</u>
    - Avoids timeconsuming report creation.

- Pop-up question for user (in case of accidental clicks):

Non-contributor analysis

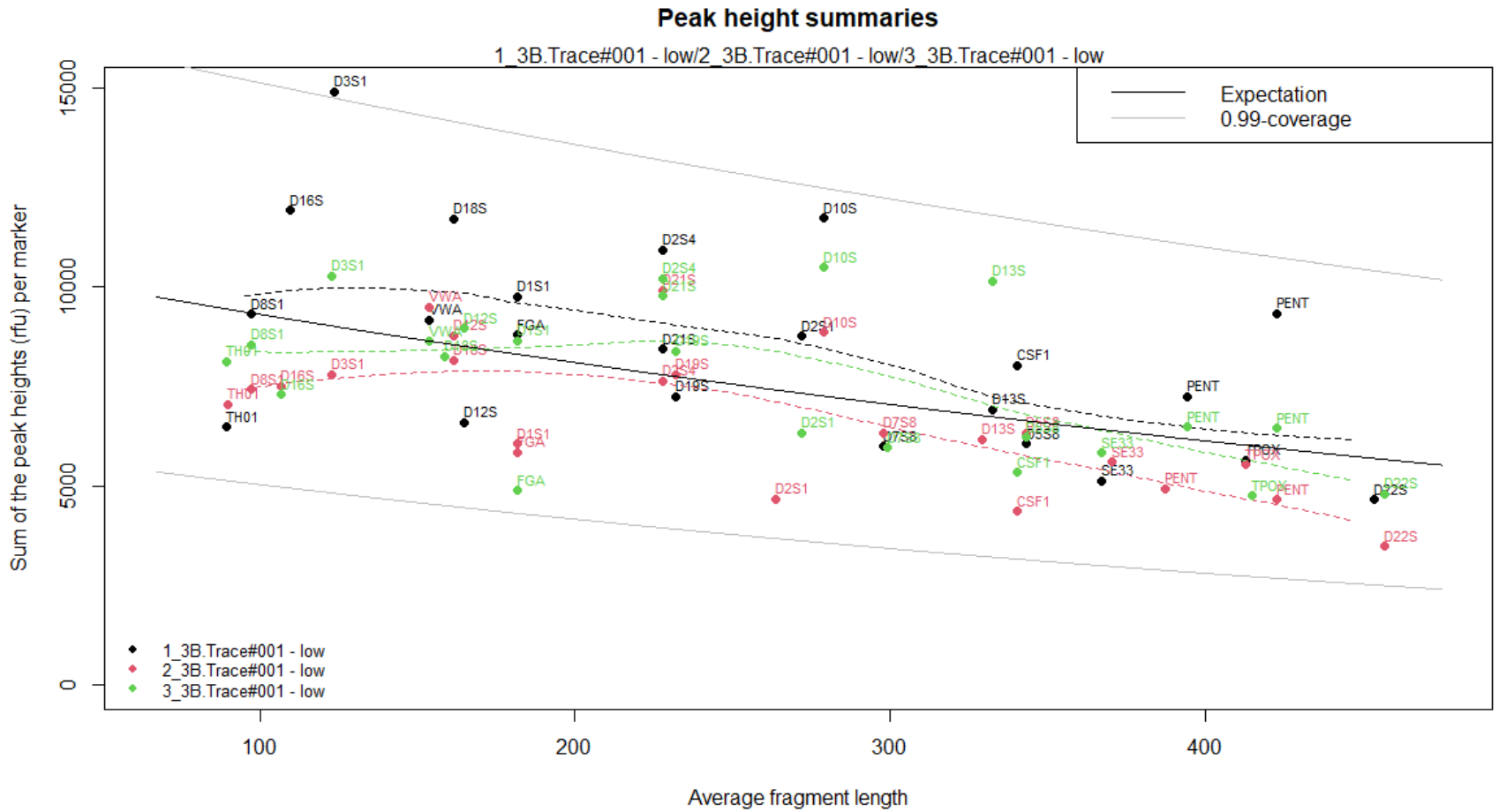






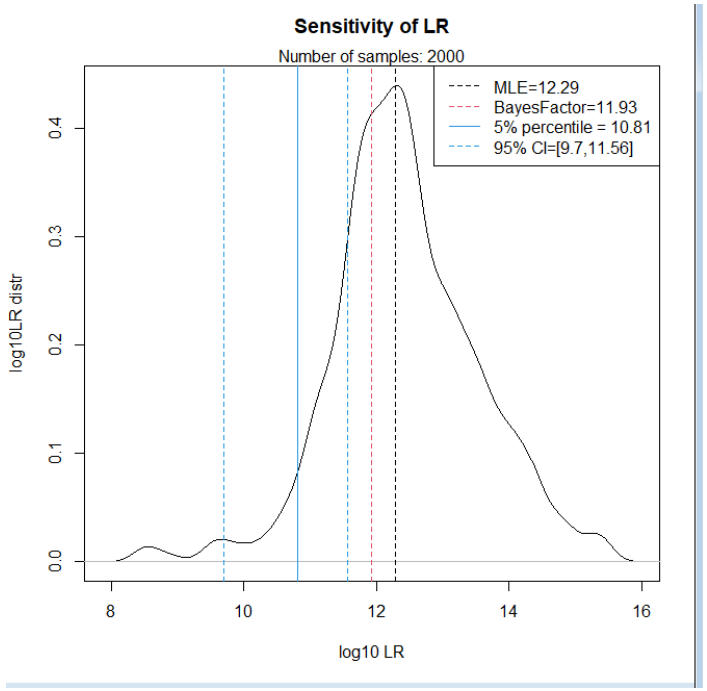
Updated Peak height summary plot (view Evidence)

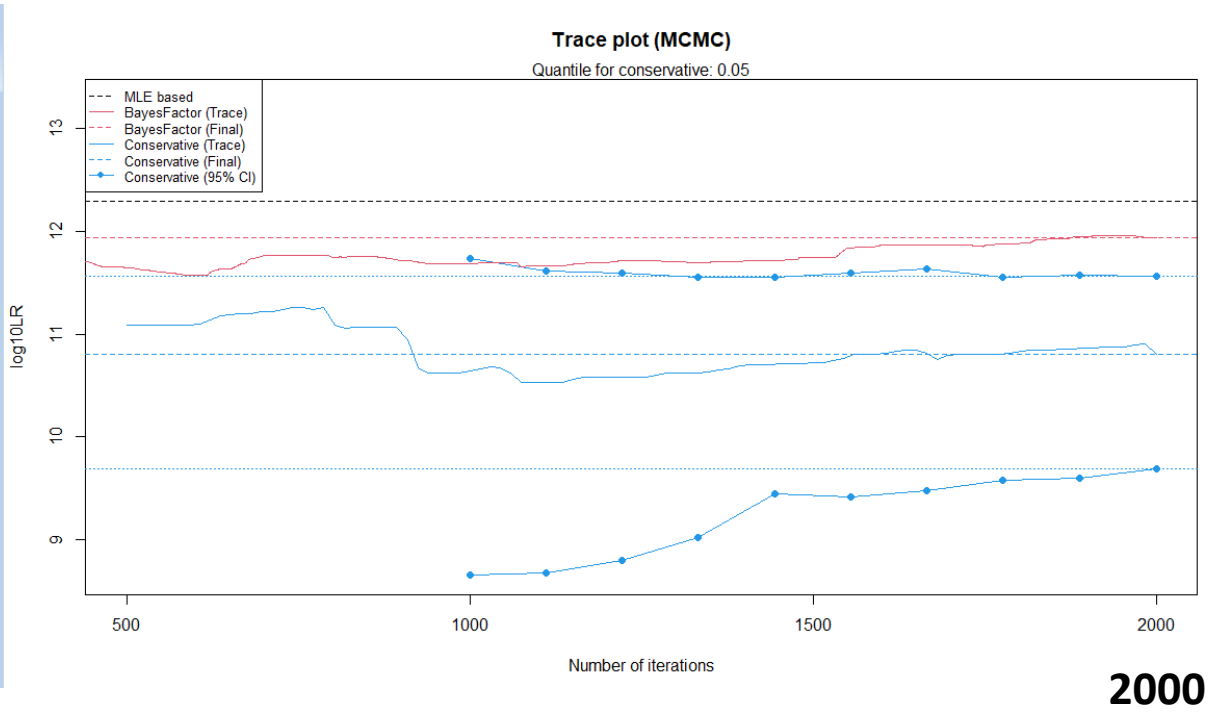
- Easier to recognize differences:
  - Highlighting each sample with different colors.
  - A dashed smoothing curve is provided.



Trace plots for LR-sensitivity

- Possible to visualize a “trace plot” for checking convergence of
  - Conservative LR (chosen quantile)
  - Bayes Factor estimate

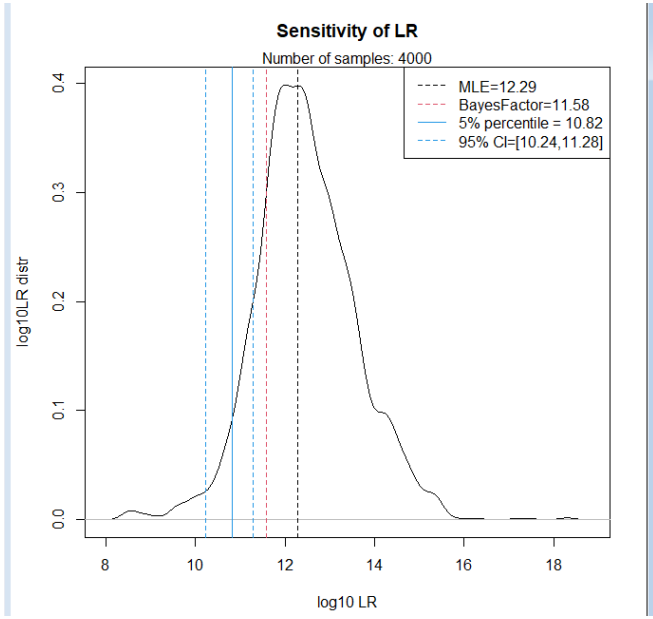


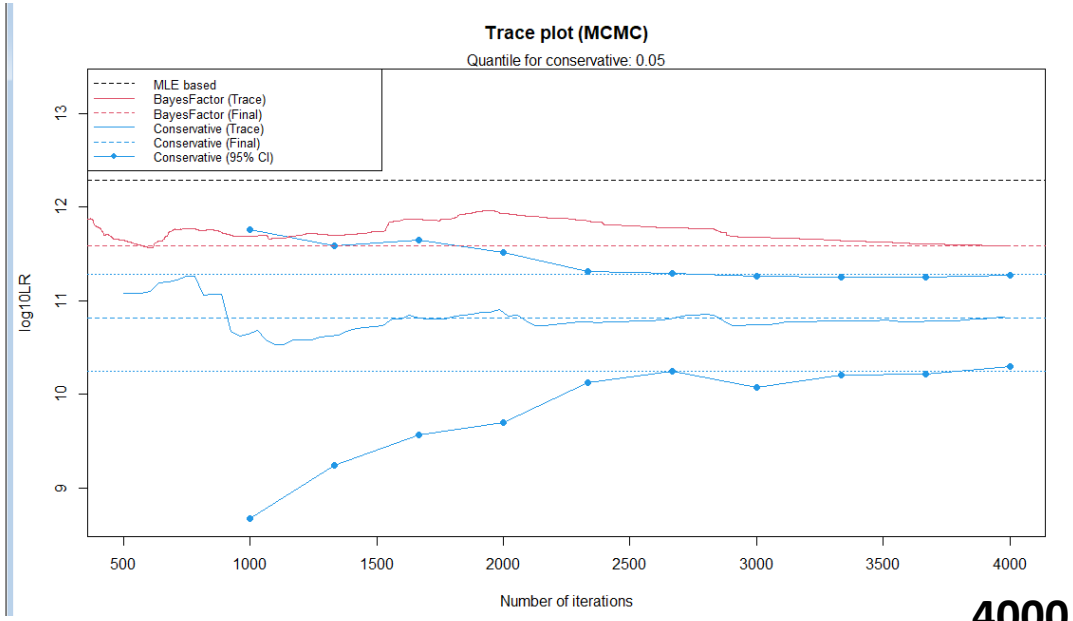


2000

Click “LR sensitivity” again for extending number of samples

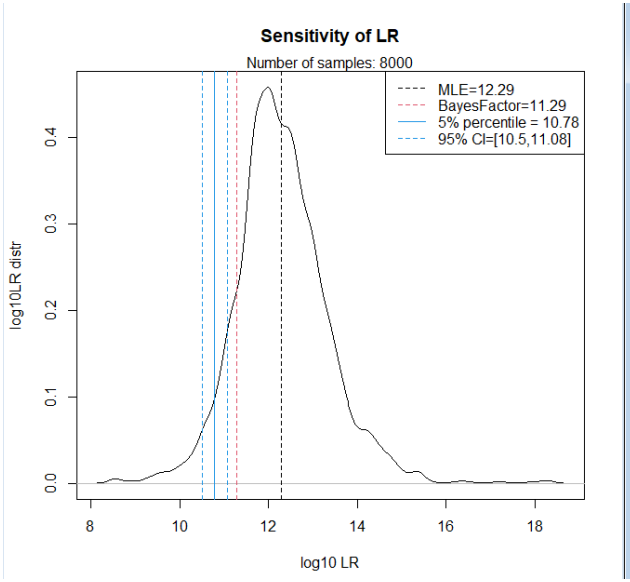
1x more

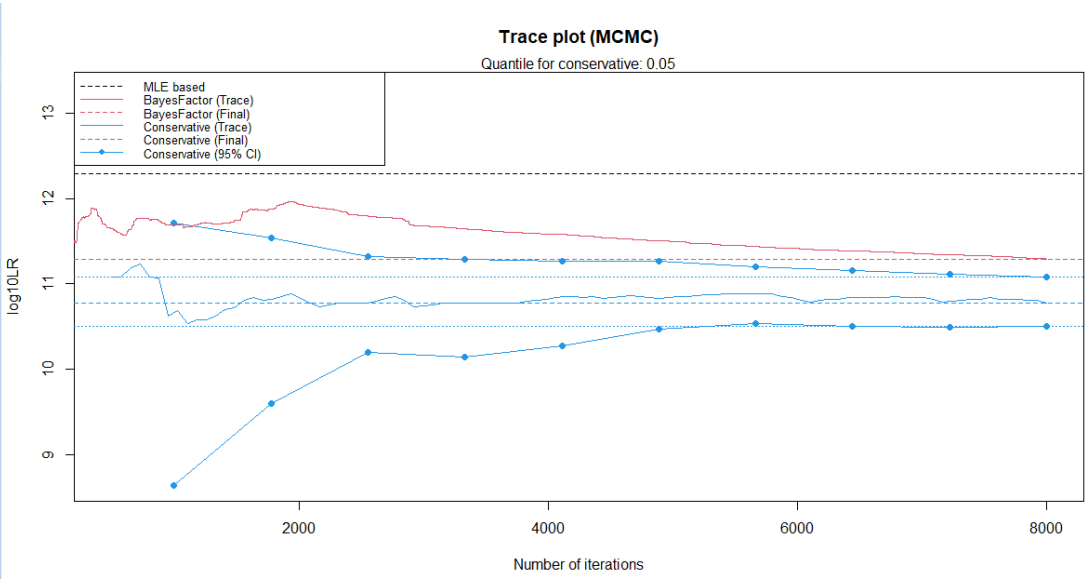




4000

3x more





Including more data details to the Report

Data used for the evaluation is included last in the report

“x” is the allele of reference

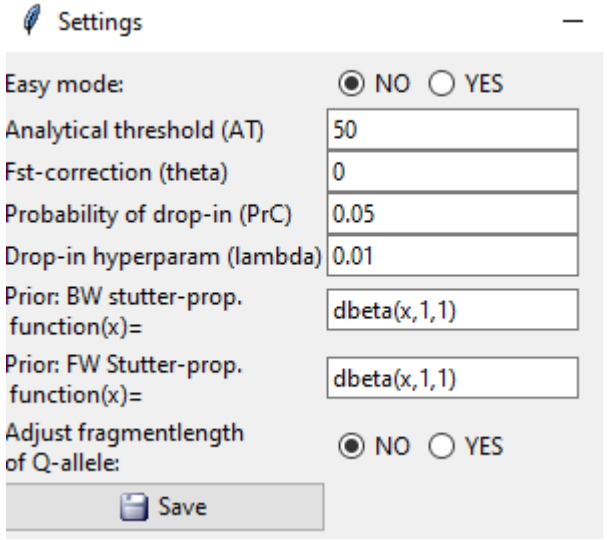
```
-------Evaluating data---------
        stain5,stain95 | ref1,ref2 | Freqs.

D3S1358
14      | 417 437      |     x    |0.124113829256116
15      | 314 388      |     x    |0.270993509735397
16      | 909 983      | x        |0.23155267099351
17      | 175 182      |          |0.202695956065901
99      |              |          |0.170644033949076

TH01
6       | 100 120      |          |0.209086894771432
7       | 94 219       |          |0.212282109973287
9       | 350 415      | x        |0.140789169831766
9.3     | 1060 934     | x    x   |0.343984886574776
8.3     | 118 104      |          |0.000896146569750075
99      |              |          |0.0929607922789887
```

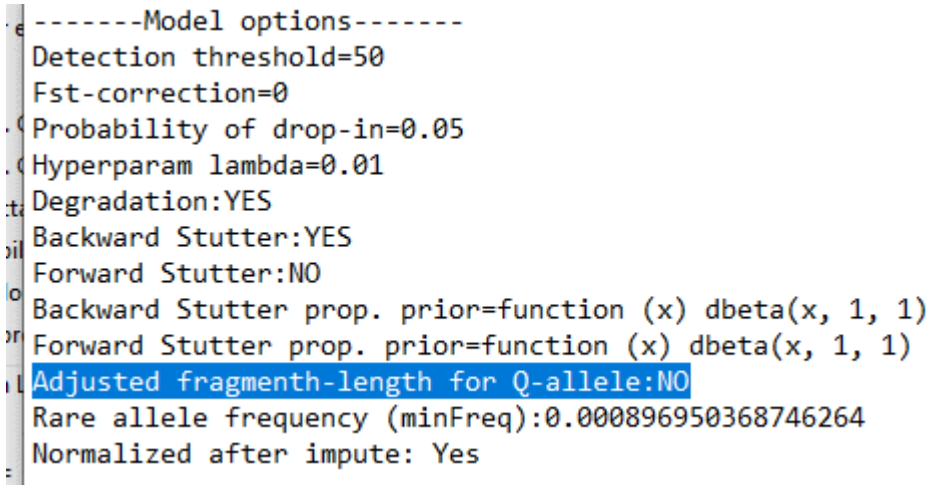
Last result from LR-sensitivity included to report when created:

```
---RESULTS BASED ON MCMC SAMPLING---
Conservative LR (5%): log10LR=10.78
95% CI of conservative LR (5%): log10LR=[10.50,11.08]
Bayes Factor (MCMC): log10LR=11.29
Number of MCMC samples (setting): 8000
Variation of randomizer (setting): 2
Tuned variation of randomizer (estimated): 2
Seed of randomizer (setting): 1
```

Defined fragment length of Q-allele

- Note: This is only relevant when applying a degradation model.
- The user can now use an adjusted version of the fragment length for the Q-allele (YES in Settings): The weighted average of frequencies of the non-observed alleles. This will shift the fragment length of the Q-allele towards alleles that are more common in the population.
- Default (NO): The «maximum defined» fragment length at considered marker (taken from kit.txt):
- Impact: The adjusted version (YES) may have a slight impact on the likelihood values and LR.





Command line

Part 1: Calculating MLE based LR

Step 1: Import and visualize profiles

```
library(euroformix) #Load package
pkg = path.package("euroformix") #get package install folder

kit = "ESX17" #defining kit to use (must be defined in getKit())
AT = 50 #analytical threshold used (global for all markers)

#Importing allele frequencies
freqFile = paste0(pkg,"/FreqDatabases/",kit,"_Norway.csv") #frequency file to use
popFreq =  freqImport(freqFile)[[1]] #need to select 1st population


#Importing evidence and reference profiles:
evidfn = paste0(pkg,"/examples/",kit,"_3p.csv")
reffn = paste0(pkg,"/examples/",kit,"_refs.csv")
evidData = sample_tableToList(tableReader(evidfn))
refData = sample_tableToList(tableReader(reffn))
```

```
plotEPG2(evidData,kit,refData) #Show in graphical interface
```

Step 2: Specify hypotheses for interpretation

Hypothesis sets: Ref3 as person of interest (POI)
Hp: Ref1 + ref3 + 1 unknown
Hd: Ref2 + 2 unknowns (all unrelated)

```
#Set up hypothesis (contributors)
POIidx = 3 #index of POI (in refData)
#Must construct a 'contribution vector' for each hypothesis:
condHp = c(1,0,2) #C1=Ref1, C2=Ref3
condHd = c(1,0,0) #C1=Ref1
knownRefhp = NULL #No known non-contributor reference under Hp
knownRefhd = POIidx #known non-contributor reference under Hd
NOC = 3 #assumed number of contributors
```

Step 3: Model fit of Hp and Hd

```
#We keep degradation and back-stutter models on (default), but turns off forward stutter model:
mleHp = calcMLE(NOC,evidData,popFreq,refData, condHp, knownRefhp, kit, FWS=FALSE)
mleHd = calcMLE(NOC,evidData,popFreq,refData, condHd, knownRefhd, kit, FWS=FALSE)
```

Step 4: Obtain calculated LR based on MLE

```
MLEresult = calcLRmle(mleHp,mleHd)
LRmle = MLEresult$log10LR #get LR on log10 scale
LRmleMarkers = MLEresult$log10LRmarker #get LR per markers
upperLR = MLEresult$log10LRupper #get theoretical upper LR
```

Step 5: Perform model validation

```
validhp = validMLEmodel(mleHp,"Hp")
validhd = validMLEmodel(mleHd,"Hd")
nSignifHp = sum(validhp$Significant) #numbers outside envelope
nSignifHd = sum(validhp$Significant) #numbers outside envelope
```

Step 6: Provide deconvolution

```
DCtableHp = deconvolve(mleHp)$table2 #top ranked genotypes (Hp)
DCtableHd = deconvolve(mleHd)$table2 #top ranked genotypes (Hd)
```

Step 7: Show model fit (expectation vs observations)

```
plotTopEPG2(mleHp)
```

Step 8: Provide non-contributor simulations:

```
nTippets = 10 #this typically takes a while (depending on the number)
tippets = calcTippet(POIidx,mleHp,mleHd,nTippets,seed = 1234)
```

Part 2: Calculating Bayesian based LR

Step 8: Calculate conservative LR (and estimate Bayes Factor)

```
#obtain 10% quantile as 'conservative LR'
mcmc = calcLRmcmc(mleHp,mleHd, 5000,quantile = 0.10,seed=1234)
LRcons_mcmc = mcmc$log10LRcons #estimated conservative LR
LRbayes_mcmc = mcmc$log10LRbayes #estimated Bayes factor (LR) using MCMC
```

Step 9: Calculate Bayes Factor with numerical integration

```
#The following calculation would take some time
int = calcLRint(mleHp,mleHd, reltol = 0.1, maxEval=20000)

#Obtaining calculated LR with relative errors:
LRbayes_int = int$log10LR
LRbayes_intError = int$log10LRerror
```