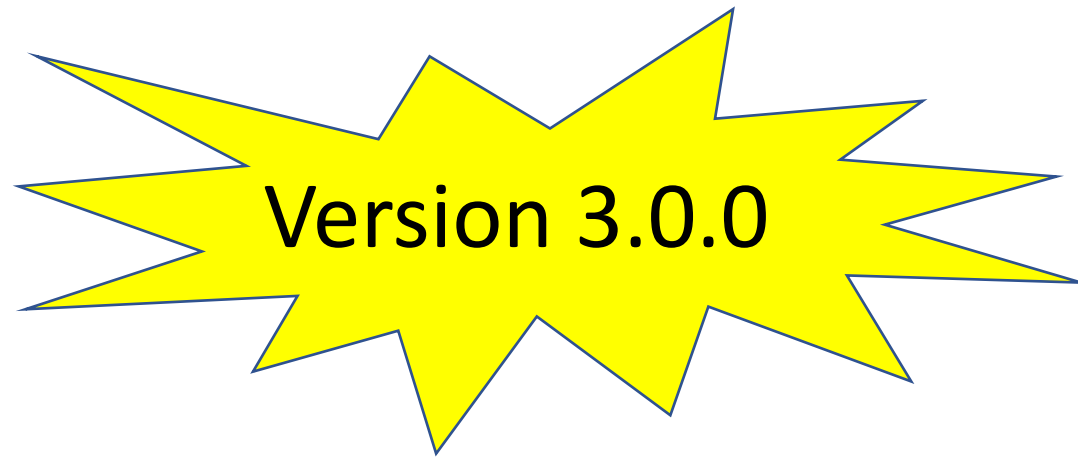


# New update of EuroForMix



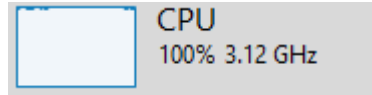
By Øyvind Bleka

Department of Forensic Sciences, Oslo University Hospital, Oslo, Norway

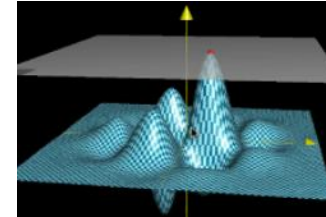
# Overview

- Better speed
- Progress bar
- Forward stutter model
- Automatic model selector
- Marker specific settings
- New 'MLE fit' layout
- Frequency normalization
- Reproducibility
- New setting recommendations

# Better speed

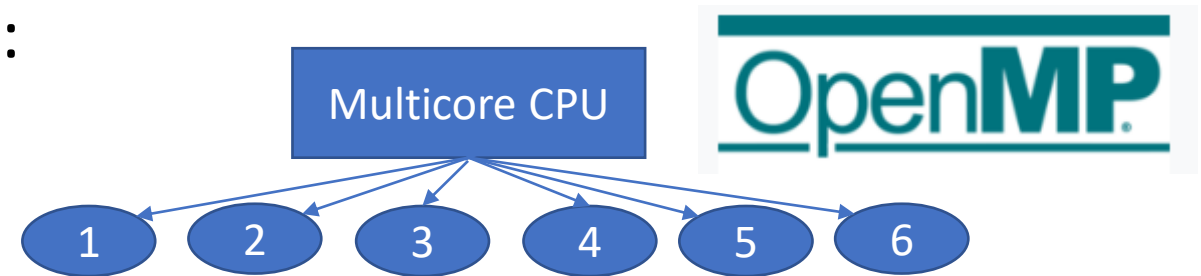


- The likelihood maximizer converges faster



- The likelihood function utilizes parallelization on the following sum:

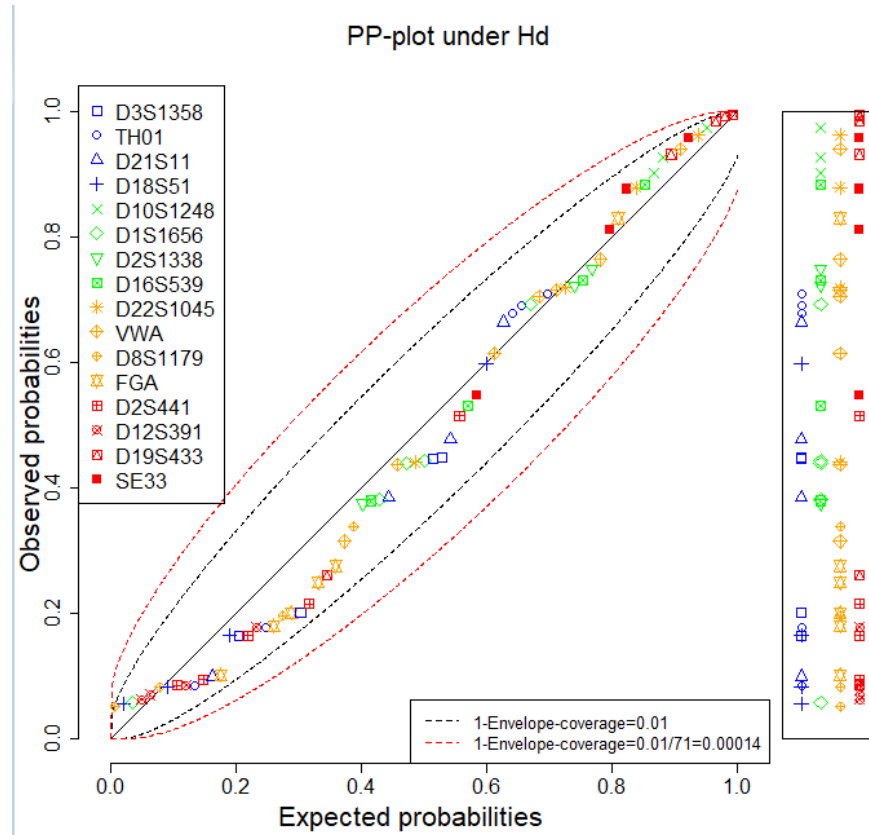
$$Pr(E|H) = \sum_{u_1 \in G} \dots \sum_{u_x \in G} Pr(u|H)Pr(E|g)$$



- Computation scales with “CPU speed” and “#CPU cores”

Base speed:	2.81 GHz
Sockets:	1
Cores:	2
Logical processors:	4
Virtualization:	Disabled

# Almost instant validation plots



Numerical integrals replaced by cumulative density function inside C++ code

# Progress bar (and time estimates)



```
Expected (upper) time is 00:02:05 (HH:MM:SS):  
|=====
```

```
| 10%
```

- User can follow progress in R-console
- Implemented for following functions:
  - Maximum likelihood method
    - An upper boundary of time is also estimated: “Expected (upper) time....”
  - MCMC simulations
  - Deconvolution
  - Integration based method

# Forward (FW) stutter model now supported

EuroForMix v3.0.0

File Frequencies Optimization MCMC Integration Deconvolution Database search Qual LR

Generate data Import data Model specification MLE fit Deconvolution Database search Qual. LR

## Model specification

Contributor(s) under Hp:

ref3

#unknowns (Hp): 2

Contributor(s) under Hd:

ref3

#unknowns (Hd): 3

1st unknown is

Unrelated

to

## Model options

Degradation:  YES  NO

BW Stutter:  YES  NO

FW Stutter:  YES  NO

## Data selection

Loci: stain5 ref3

D3S1358

TH01

D21S11

D18S51

D10S1248

D1S1656

D2S1338

D16S539

D22S1045

VWA

D8S1179

FGA

D2S441

D12S391

D19S433

SE33

## Show selected data

Evidence(s)

stain5

Show

## Calculations

Quantitative LR  
(Maximum Likelihood based)

Quantitative LR  
(Bayesian based)

Qualitative LR  
(semi-continuous)

## Estimates under Hd

Parameter estimates:

param	MLE	Std.Err.
Mix-prop. C1	4.844e-01	6.113e-02
Mix-prop. C2	3.407e-01	2.623e-02
Mix-prop. C3	1.749e-01	4.391e-02
P.H.expectation	1.002e+03	4.050e+01
P.H.variability	1.592e-01	2.222e-02
Degrad. slope	7.379e-01	3.199e-02
BWstutt-prop.	8.457e-02	1.664e-02
FWstutt-prop.	2.423e-08	7.186e-06

## Estimates under Hp

Parameter estimates:

param	MLE	Std.Err.
Mix-prop. C1	1.641e-01	1.873e-02
Mix-prop. C2	5.004e-01	2.442e-02
Mix-prop. C3	3.356e-01	2.208e-02
P.H.expectation	1.005e+03	3.875e+01
P.H.variability	1.520e-01	1.587e-02
Degrad. slope	7.402e-01	3.032e-02
BWstutt-prop.	9.076e-02	1.428e-02
FWstutt-prop.	2.198e-09	3.764e-07

# Automatic model selector (based on AIC)

The screenshot displays the EuroForMix v3.0.0 interface. The main window is titled "EuroForMix v3.0.0" and has a menu bar with "File", "Frequencies", "Optimization", "MCMC", "Integration", "Deconvolution", "Database search", and "Qual LR". Below the menu bar are tabs for "Generate data", "Import data", "Model specification", "MLE fit", "Deconvolution", "Database search", and "Qual LR".

The "Model specification" tab is active, showing several sections:

- Contributor(s) under Hp:** Includes checkboxes for "ref1" and "ref2", and a dropdown for "#unknowns (Hp):" set to 1.
- Contributor(s) under Hd:** Includes checkboxes for "ref1" and "ref2", and a dropdown for "#unknowns (Hd):" set to 2.
- 1st unknown is:** A dropdown menu set to "Unrelated".
- Model options:** Radio buttons for "Degradation:" (YES selected), "BW Stutter:" (NO selected), and "FW Stutter:" (NO selected).

The "Data selection" section shows a table of loci:

Loci	stain5	ref1	ref2
D3S1358	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
TH01	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D21S11	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D18S51	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D10S1248	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D1S1656	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D2S1338	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D16S539	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D22S1045	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
VWA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D8S1179	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
FGA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D2S441	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D12S391	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
D19S433	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SE33	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

The "Calculations" section has four buttons: "Quantitative LR (Maximum Likelihood based)", "Optimal quantitative LR (automatic model search)", "Quantitative LR (Bayesian based)", and "Qualitative LR (semi-continuous)". A red circle highlights the "Optimal quantitative LR (automatic model search)" button, with a red arrow pointing to the "Model comparison results" window.

The "Model comparison results" window shows a table with the following data:

NOC	DEG	BWstutt	FWstutt	logLik	adjLogLik	log10LR	MxPOI	SignifHp	SignifHd
2	FALSE	FALSE	FALSE	-581.04	-584.04	-5.71	0.43	0	0
2	TRUE	FALSE	FALSE	-573.75	-577.75	-2.67	0.42	0	0
2	FALSE	TRUE	FALSE	-520.72	-524.72	-0.63	0.42	0	0
2	TRUE	TRUE	FALSE	-512.45	-517.45	1.34	0.41	0	0
3	FALSE	FALSE	FALSE	-524.93	-528.93	9.87	0.35	0	0
3	TRUE	FALSE	FALSE	-513.27	-518.27	11.42	0.35	0	0
3	FALSE	TRUE	FALSE	-503.39	-508.39	11.76	0.35	0	0
3	TRUE	TRUE	FALSE	-486.78	-492.78	13.08	0.33	0	0

The last row of the table is highlighted in blue. An information box on the right of the "Select models to compare" window provides the following details:

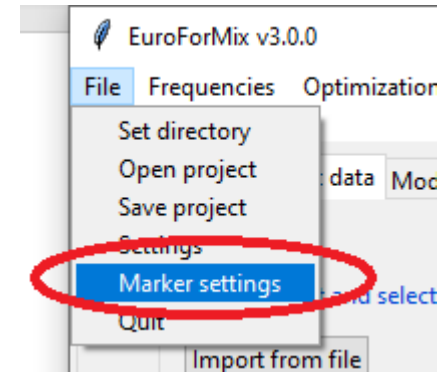
- Following setup to be evaluated: POI=ref2, Conditionals=ref1, Number of contributors={2,3}
- Model combinations: Degrad: YES/NO, BW stutter: YES/NO, FW stutter: NO
- Do you want to continue? (Yes/No buttons)

Best model given in "MLE fit" automatically  
Table will be stored when "create report"

Notice that degradation model must be selected (YES) in order to be an outcome

# Marker specific settings

- Analytical threshold
- Drop-in probability
- Hyperparameter lambda
- Theta-correction (Fst)



The 'Marker specific settings' dialog box contains a table with the following columns: Marker, Analyt. thresh (AT), Dropin prob. (pC), Hyperparam (lambda), Fst-correction (theta), and Dye (color). The table lists 17 markers with their respective settings.

Marker	Analyt. thresh (AT)	Dropin prob. (pC)	Hyperparam (lambda)	Fst-correction (theta)	Dye (color)
D3S1358	50	0	0.01	0	blue
TH01	50	0	0.01	0	blue
D21S11	50	0	0.01	0	blue
D18S51	50	0	0.01	0	blue
D10S1248	100	0	0.01	0	green
D1S1656	100	0	0.01	0	green
D2S1338	100	0	0.01	0	green
D16S539	100	0	0.01	0	green
D22S1045	50	0	0.01	0	yellow
D8S1179	50	0	0.01	0	yellow
FGA	50	0	0.01	0	yellow
D2S441	50	0	0.01	0	red
D12S391	50	0	0.01	0	red
D19S433	50	0	0.01	0	red
SE33	50	0	0.01	0	red

← Dyes provided when kit is selected

“Automatic dye fill-out”:  
Plugs in first dye value to all markers with same dye

Settings are remembered after closing program



# New “MLE fit” layout

**Estimates under Hd**

Parameter estimates:

param	MLE	Std.Err.
Mix-prop. C1	0.73817	0.03839
Mix-prop. C2	0.26183	0.03839
P.H.expectation	756.79942	59.08343
P.H.variability	0.16958	0.02873
Degrad. slope	0.77786	0.06526
BWstutt-prop.	0.15306	0.03431
FWstutt-prop.	0.03054	0.03472

Maximum Likelihood value

logLik= -208.3  
**adj.loglik= -214.3**  
 Lik= 3.455e-91

Further Action

MCMC simulation  
 Deconvolution  
 Model validation  
 Model fitted P.H.

**Estimates under Hp**

Parameter estimates:

param	MLE	Std.Err.
Mix-prop. C1	0.74293	0.03703
Mix-prop. C2	0.25707	0.03703
P.H.expectation	757.99426	58.29054
P.H.variability	0.16735	0.02660
Degrad. slope	0.77807	0.06433
BWstutt-prop.	0.15021	0.03323
FWstutt-prop.	0.03328	0.03229

Maximum Likelihood value

logLik= -183.6  
 adj.loglik= -189.6  
 Lik= 1.908e-80

Further Action

MCMC simulation  
 Deconvolution  
 Model validation  
 Model fitted P.H.

**Joint LR**

LR= 5.524e+10  
 log10LR= 10.74  
 Upper boundary= 10.77

**Non-contributor analysis**

Select reference to replace with non-contributor:

Sample maximum based  
 Sample integrated based

**Further**

LR sensitivity  
 Create report

**LR for each locus**

VWA	9.252
D16S539	23.43
D2S1338	158.8
D19S433	276
TH 01	22.38
FGA	19
D3S1358	13.68

Upper boundary of LR  
 1/Random match prob.  
 (Fst and conditional contributors are included)



LR should not exceed this value

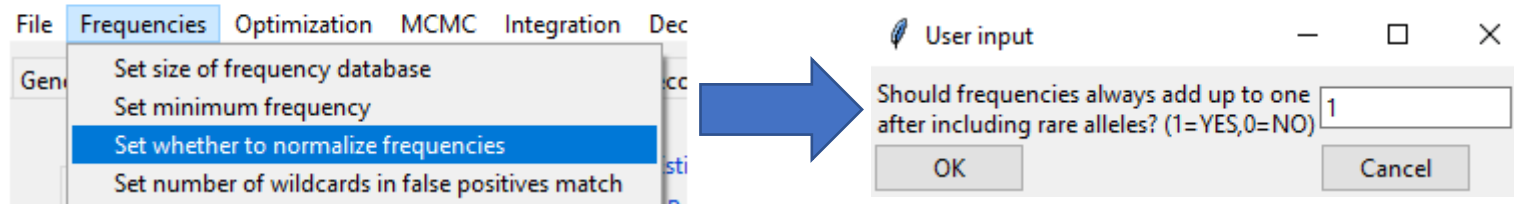
adj.loglik=logLik - #param (only effective\*)

\*since the *NOC* mix-prop parameters sum to 1 we only count ‘*NOC-1*’  
 Here #param = *NOC* - 1 + 2 (P.H.exp/var) + 3 (deg+BW+FW) = 6

# Frequency normalization

Frequencies summed to 1?

- Concerns whether allele frequencies are **normalized** after inserting rare alleles (with minimum frequency), i.e. alleles not in the frequency database (tiny effect on LR values)

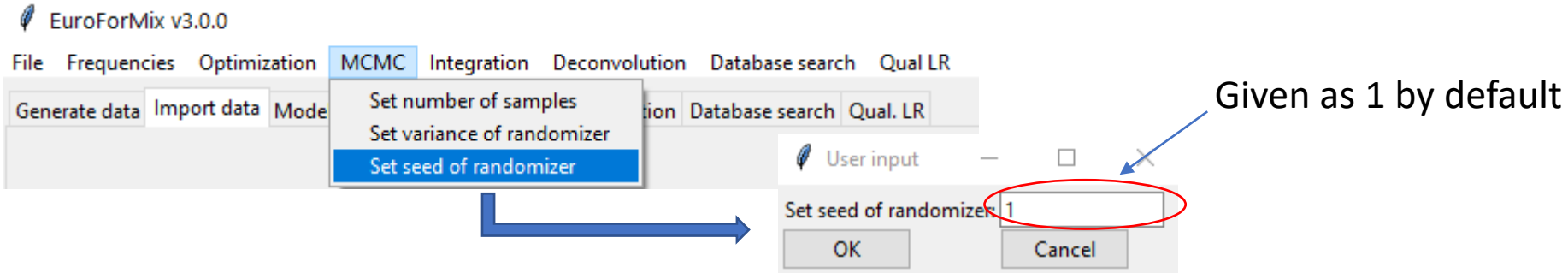


- Choice stored in report: 

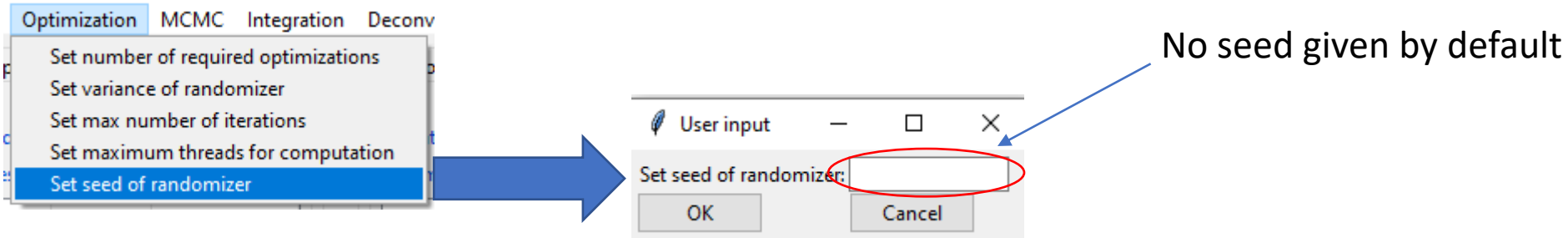
```
Rare allele frequency (minFreq):0.001  
Normalized after impute: Yes
```
- Implementations:
  - EuroForMix v1: **Yes**
  - LRmix Studio: **No**
  - EuroForMix v2: **No**
  - DNASstatistix: **Yes**
  - EuroForMix v3: **Yes** (by default, can be changed to **No**)

# Reproducibility

- Set seed for reproducibility for “Conservative LR” (MCMC)



- Set (optional) seed for reproducibility of startpoints used for “Optimization” to obtain *maximum likelihood based LR*



- Seed information included in report:

```
-----Optimalisation setting-----  
Number of required optimisations: 2  
Seed for optimizations: NONE
```

```
---Conservative LR (5% lower log10LR quantile)---  
log10LR=9.032  
Number of MCMC samples: 2000  
Seed: 1
```

# New settings recommendations

- Set “Optimization -> variance of randomizer” to 1 (**not 10**) – now default
  - This is important to get robust and fast MLE maximizer
- Set “Optimization -> number of req. optimizations ” to x=2 (**not 4**) – now default
  - The new implementation require x independent optimizations obtaining same maximum likelihood value.
- **Remember to optimize sampler settings for ‘LR sensitivity’:**
  - Set “MCMC-> variance of randomizer” such that the acceptance **rate** is around 0.2 (good MCMC sampling properties)
    - Calibrate this number:
      - Lower the number of samples (about 200-500) to check this
      - Trial and error using **MCMC simulation** under *Further action* (use Hp for fastest computation)
      - Tradeoff: Small variance -> low **rate** vs High variance -> high **rate**