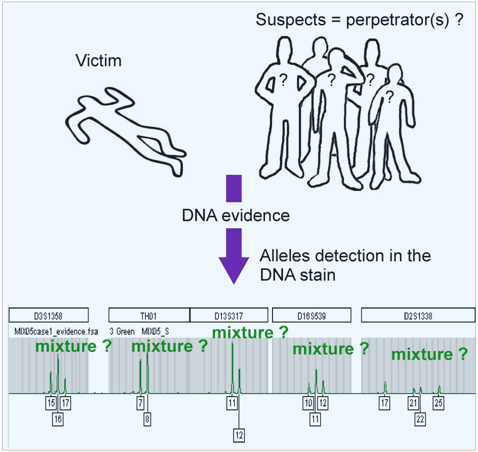
forensim: a free software for DNA evidence interpretation and forensic genetics simulations

Hinda Haned

Biometry and Evolutionary Biology, University of Lyon, France

European DNA Profiling Group; Edinburgh October 2009

Forensic DNA mixtures : A challenging task



Suspects = perpetrator(s) ?

Victim

DNA evidence

Alleles detection in the DNA stain

Interpretation issues

- Is it a mixture ?
- How many people involved ?
- Weight of the stain as an evidence ?

Available methods

- Several methods dedicated to mixtures interpretation are available :

LR in case of population substructure	Curran *et al* 1999
Number of contributors	Egeland *et al* 2003
Unknown related contributors	Fung and Hu 2003
Genotyping errors	Thompson *et al* 2003

Available methods

- Several methods dedicated to mixtures interpretation are available :

LR in case of population substructure	Curran *et al* 1999
Number of contributors	Egeland *et al* 2003
Unknown related contributors	Fung and Hu 2003
Genotyping errors	Thompson *et al* 2003

None of these methods are in widespread use ⇒

Lack of evaluation of methods' efficiency and robustness

How to evaluate these methods ?

On simulated DNA stains where the circumstances of the hypothetical crime are known by the experimenter.

The experimenter would evaluate method's efficiency :

1. While varying accurate parameters :
    - type of markers analyzed
    - number of markers analyzed
    - number of contributors to the DNA evidence
2. In critical situations :
    - population subdivision (co-ancestry)
    - partial profiles
    - relatedness between contributors to the DNA stain
    - allele dropout

How to evaluate these methods ?

- **Laboratory simulated DNA stains :**
  - Some scenarios are hard to test in laboratory (ex. population substructure)
  - Cost issues : new experiments are to be conducted for each tested scenario

- Computer simulated DNA stains :
  - Complex scenarios can be simulated
  - No cost issues

Currently, there is no free software providing simulation tools specific to forensic genetics.

How to evaluate these methods ?

- Laboratory simulated DNA stains :
  - Some scenarios are hard to test in laboratory (ex. population substructure)
  - Cost issues : new experiments are to be conducted for each tested scenario

- **Computer simulated DNA stains :**
  - Complex scenarios can be simulated
  - No cost issues

Currently, there is no free software providing simulation tools specific to forensic genetics.

The *forensim* software

Main features

- *forensim* is a package for the R statistical software
- *forensim* is freely available
- Relies on object oriented programming
- Sources freely available on R-Forge
- Compiles and runs on a wide variety of UNIX platforms, Windows and MacOS

forensim's structure

Simulation tools

Simulation of data commonly encountered in forensic casework

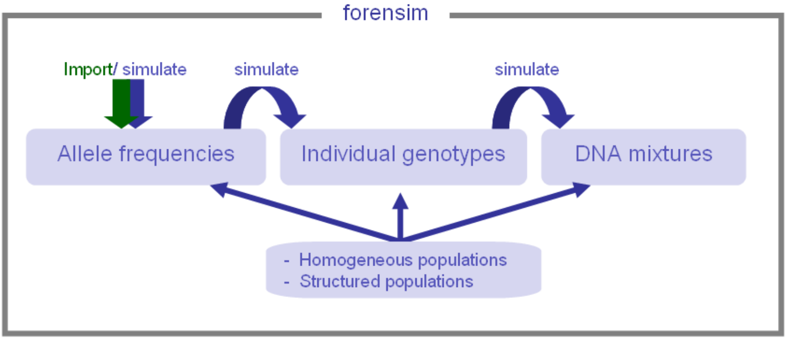
Statistical tools

Main statistical methods for forensic DNA evidence interpretation

Simulation tools

Object oriented programming

Structured code, can easily be modified/ enriched ⇒ allows a wide variety of scenarios



Statistical tools

Statistical methods usually used to report the weight of a DNA stain are implemented :

Random man exclusion probability

- θ correction for allele dependencies Weir, In Buckelton *et al*, 2005
- Correction for allele dropouts Van Nieuwerburgh *et al*, 2009

Likelihood ratios

- General formula for likelihood ratios Curran *et al*, 1999

Random match probabilities

- Accounts for :
  - relatedness
  - allele dependencies

Balding & Nichols, 1994

Original methods are also implemented

As an R package, *forensim* is constantly evolving. New methods are regularly implemented in the package.

Recently implemented

A maximum-likelihood estimator of the number of contributors to a forensic DNA mixture.

Haned *et al* (2009) Estimating the number of contributors to forensic DNA mixtures : Does maximum likelihood perform better than maximum allele count ? *J. Forensic. Sci., In revision*

Soon available

A modification of the RMNE probability allowing allelic dropout.

Van Nieuwerburgh *et al* (2009) Impact of allelic dropout on evidential value of forensic DNA profiles using RMNE. *Bioinformatics*, **25**, 225-229

Simulation tools : Focus on DNA mixtures

Two kinds of information stored :

Usual information

- Alleles present in the stain
- Marker names
- Allele frequencies of the putative population

Simulation-related information

- Number of individuals involved
- Contributors’ genotypes
- Contributors’ populations

Simulating forensic DNA mixtures

Simulating a 3-person mixture, using the African American allele frequencies (Butler *et al*, 2003) :

Step1 : load the package

```
> library(forensim)
   ### forensim 1.1.1 is loaded ###
```

Simulating forensic DNA mixtures

Simulating a 3-person mixture, using the African American allele frequencies (Butler *et al*, 2003) :

Step1 : load the package

```
> library(forensim)

   ### forensim 1.1.1 is loaded ###
```

Step2 : generate the data

```
> data(strusa)
> geno <- simugeno(strusa, n = c(100, 0, 0))
> mix3 <- simumix(geno, ncontri = c(3, 0, 0))
```

Simulating forensic DNA mixtures

Mixture representation in forensim

```
> mix3

   # Simumix object: simulated mixture #


@which.loc: vector of  15  locus names
@ncontri: 3
@mix.prof:  3 x 15 data frame of the contributors genotypes
@mix.all: list of the alleles found in the mixture
@popinfo: populations of the contributors
```

Simulating forensic DNA mixtures

Mixture representation in forensim

```
> mix3

   # Simumix object: simulated mixture #


@which.loc: vector of  15  locus names
@ncontri: 3
@mix.prof:  3 x 15 data frame of the contributors genotypes
@mix.all: list of the alleles found in the mixture
@popinfo: populations of the contributors
```

Display stain profiles at locus FGA

```
> mix3$mix.all$FGA

[1] "20" "21" "24" "25"
```

Simulating forensic DNA mixtures

Mixture representation in forensim

```
> mix3

   # Simumix object: simulated mixture #


@which.loc: vector of  15  locus names
@ncontri: 3
@mix.prof:  3 x 15 data frame of the contributors genotypes
@mix.all: list of the alleles found in the mixture
@popinfo: populations of the contributors
```

Display stain profiles at locus FGA

```
> mix3$mix.all$FGA

[1] "20" "21" "24" "25"
```

Display contributors profiles at locus FGA

```
> mix3$mix.prof[, "FGA"]

  ind70   ind58    ind1
"21/24" "24/25" "21/20"
```

Reporting the weight of the evidence

What is the exclusion probability of the DNA stain ?

```
>PE(mix3, freq = strusa, refpop = "Afri", theta = 0, byloc =FALSE)

      PE
0.999989
```

Reporting the weight of the evidence

What is the exclusion probability of the DNA stain ?

```
>PE(mix3, freq = strusa, refpop = "Afri", theta = 0, byloc =FALSE)

      PE
0.999989
```

Help page

mix : the DNA mixture

freq : the allele frequencies to use

refpop : the reference population, used only if freq contains allele frequencies for multiple populations

theta : θ correction for allele dependencies

byloc : logical indicating whether the PE is computed by/overall loci

Reporting the weight of the evidence

By locus exclusion probability

```
> PE(mix3, freq = strusa, refpop = "Afri", byloc = TRUE)

          PE_l
CSF1PO  0.6315
FGA     0.6320
TH01    0.4140
TPOX    0.2629
VWA     0.1739
D3S1358 0.2893
D5S818  0.2018
D7S820  0.2259
D8S1179 0.6082
D13S317 0.1739
D16S539 0.4404
D18S51  0.5828
D21S11  0.5426
D2S1338 0.6339
D19S433 0.8437
```

More statistical methods via other packages

Methods from other R packages can be used in *forensim*

- Basic statistical inference
- Bayesian inference
- Familial analysis
- Population genetics

“Work on the cutting edge of statistical computing”

-R-core team-

More Info/Help

You are not familiar with R :

Do not worry ! A detailed tutorial with practical and reproducible examples is available online :
`http://forensim.r-forge.r-project.org/`

You are encountering problems using *forensim* :

- Post a message on *forensim* mailing list :
  `forensim-help@lists.r-forge.r-project.org.`
- Contact me :
  `haned@biomserv.univ-lyon1.fr`

Contributions are greatly encouraged !

forensim is evolving, and you can participate !

- Suggestions ?
- Particular needs ?
- Contributions to the package : data, methods... are welcome !

`http://forensim.r-forge.r-project.org/`