



Netherlands Forensic Institute  
*Ministry of Security and Justice*

# Introduction to the software for Forensim users

Hinda Haned

[h.haned@nfi.minvenj.nl](mailto:h.haned@nfi.minvenj.nl)



**Forensic Genomics**  
Consortium Netherlands



Part 1: general introduction to R


Part 2: introduction to the LRmix module



# Part1



# What is ?

 is a freely available language and environment for statistical computing and graphics

Open source means that the **source code** is available to **all potential users** and they are **free** to use, modify, and re-distribute the source code.



# What is open source software?

Open source means that the **source code** is available to **all potential users** and they are **free** to use, modify, and re-distribute the source code.



# What is source code?

- ▶ A source code file is the original computer program written by a programmer
- ▶ It shows the logic behind a program, the strategies used for solving various problems, and all the details needed to make the software run



# What is source code?

PCR simulation module in Forensim:

simPCR2TK: a graphical simulation interface the PCR

**simPCR2TK: a graphical interface for PCR simulation**

pre-PCR parameters	PCR parameters	Cells ploidy
number of cells <input type="text" value="5"/>	Prob. PCR efficiency <input type="text" value="0.8"/>	<input checked="" type="checkbox"/> Diploid
Prob. extraction <input type="text" value="0.6"/>	#PCR cycles <input type="text" value="28"/>	prob. allele A in haploid cells <input type="text" value="0.5"/>
Prob. surviving aliquot <input type="text" value="0.3"/>	Allele detection threshold <input type="text" value="20000000"/>	

# Replicate simulations

**Dismiss** **Simulate!**



# What is source code?

Source code of the module:

```
1 "simPCR2TK" <- function()  
2 {  
3   if(!require(tcltk)) stop("package tcltk is required")  
4   if(!require(tkrplot)) stop("package tkrplot is required")  
5   tclRequire("Tktable")  
6   tclRequire("Tktable")  
7   font0 <- tkfont.create(family="courrier",size=35,weight="bold",slant="italic")  
8   font1<-tkfont.create(family="times",size=14,weight="bold")#,slant="italic")  
9   font2<-tkfont.create(family="times",size=16,weight="bold",slant="italic")  
10  font3<-tkfont.create(family="times",size=12)#,slant="italic")  
11  font4<-tkfont.create(family="courrier",size=14)#,slant="italic")  
12  font5<-tkfont.create(family="courrier",size=13,weight="bold")#,slant="italic")  
13  font6<-tkfont.create(family="times",size=13)#tkframe entries labels  
14  tf <- tktoplevel()  
15  tkwm.title(tf,"simPCR2TK: a graphical simulation interface the PCR")  
16  
17  done <- tclVar(0)  
18  ...
```





# Open source software: Definitions

- ▶ Refers exclusively to the source code and it is possible to have support, services, documentation, and even binary versions which are not monetarily free.
- ▶ Open source = transparent
- ▶ Open source ← “black box”
- ▶ Open source ← free of charge, but most of the time it is!



# Open source software: Definitions

Four essential freedoms:

1. *Redistribute* software without restriction
2. *Access* the source code
3. *Modify* the source code
4. *Distribute* the modified version of the software : for free...or not!

Open source evolves through community cooperation:

- ✓ Community of users
- ✓ Community of developers



# Open source $\neq$ Freeware

- ▶ Freeware : A software which can be downloaded, used, and copied without restrictions, but, no access to the source code.
- ▶ There is no community and no development infrastructure around freeware" as there is around open source software

Open source  $\neq$  Freeware



## Open source vs. Freeware: example



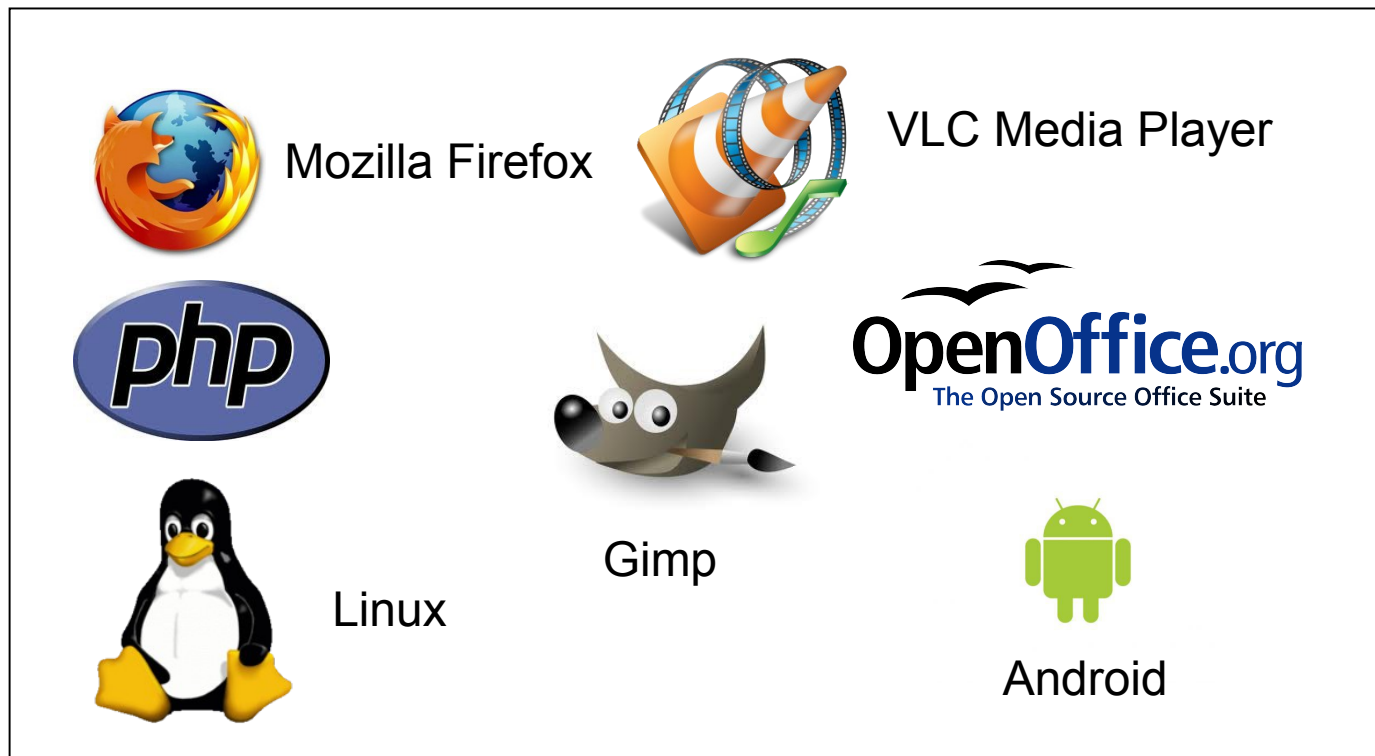
- ▶ Open source software
- ▶ Community of developers
- ▶ Community of users
- ▶ Users report bugs to developers



- ▶ Freeware  $\neq$  open source
- ▶ Community of users
- ▶ Users report bugs to Microsoft



## Widely used open source software





## Is open source secure?


*“Since everyone can contribute to open source software, isn't less reliable than commercial software?”*

- ▶ No ! Any change to an open source project is submitted to the filter of a group of maintainers first
- ▶ Open source software can be more secure than commercial software:
  - active communities tracking bugs and inconsistencies
  - users = testers that report back to the project (mailing lists, forums)



# Open source = Free software

Free as in free speech, not as in free beer!

**Example:** Commercial version of the R software:  **REVOLUTION**  
ANALYTICS

- ▶ Core program is free, but the company charges for extra products
  - Enterprise deployment
  - Technical support
  - Consulting
  - Training
  - ...



# Summary

- ▶ Open source software : freedom to run, copy, distribute, study, change and improve the software.
- ▶ Open source  $\neq$  non-commercial

*“Open source promotes software reliability and quality by supporting independent peer review and rapid evolution of source code”*  
- The Free Software Foundation-





# Why should you use ?

## Advantages

- Fast and free
- Work on the cutting edge of statistical research
- Very active user community
- Excellent for simulation, programming, computer, intensive analyses...
- Script language: forces you to think about your analysis!

## Disadvantages

- Not user-friendly
- Data preparation and cleaning might be difficult
- R-Help list : famous to be hostile !



# fundamentals

- R is like other programming languages : C, Perl and Python
- R is particularly useful because it contains built-in mechanisms for organizing data, running calculations, creating graphical representations of data sets
- Researchers and engineers can improve the existing code for a specific task, for example, the calculation of the mean, and make these new functions available in a package



# fundamentals

A package is a collection of small programs dedicated to a specific task. You can find packages for almost anything !

- Statistical Genetics
- Forensic genetics
- Bayesian inference
- Computational physics
- Clinical trials
- Probability distributions
- Analysis of ecological data
- Finance
- Graphics
- Medical image analysis
- Multivariate statistics
- Statistics for the social sciences
- Analysis of spatial data
- Survival analysis
- Time series analysis

For an overview of available packages per topic:

<http://cran.r-project.org/web/views/>



## fundamentals

- The Comprehensive R Archive Network (CRAN) : is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R.
- Currently, the CRAN package repository features 2343 available packages!



# The Forensim package

- Forensim is a package for the R statistical software
- Forensim is freely available
- Sources are freely available on the web
- Compiles and runs on a wide variety of UNIX platforms, Windows and MacOS



# Overview of the Forensim package

Forensim combines different features to answer some of the identified needs in forensic genetics

Simulation tools: simulation of data commonly encountered in forensic casework

Statistical tools: main statistical methods for forensic DNA evidence interpretation



# Forensim package: simulation tools

## R script

- Allele frequencies- population substructure
- DNA profiles (qualitative and quantitative data)
- DNA mixtures

## User-friendly modules

- Polymorphism chain reaction: simPCR2
- Heterozygote balance (demonstration later today): Hbsimu



# Forensim package: statistical tools

- Exclusion probability (RMNE)
- Likelihood ratios: General formula for likelihood ratios – LRmix user-friendly module
- Curran et al, 1999, Balding & Buckleton, FSIG, 2009
- Random match probabilities





## Forensim package: documentation

- Manual: all functions and data sets are described, examples are given
- Detailed tutorials with practical and reproducible examples are available online
- LRmix tutorial is distributed during the course

<http://forensim.r-forge.r-project.org/>



## Forensim package: how to get help

- Post a message on Forensim mailing list [forensim-help](#)
- Contact me: [h.haned@nfi.minvenj.nl](mailto:h.haned@nfi.minvenj.nl)/[hi.haned@gmail.com](mailto:hi.haned@gmail.com)
- Subscribe to [forensimnews@gmail.com](mailto:forensimnews@gmail.com)



# Part 2: introduction to LRmix



# (1) Install the R software

www.cran.r-project.org



## CRAN

[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

## About R

[R Homepage](#)  
[The R Journal](#)

## Software

[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

## Documentation

[Manuals](#)  
[FAQs](#)  
[Contributed](#)

## The Comprehensive R Archive Network

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want

- [Download R for Linux](#)
- [Download R for MacOS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the li

### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2012-06-22, Roasted Marshmallows): [R-2.15.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and](#) corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

### Questions About R



# (1) Install the R software



## CRAN

[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

## About R

[R Homepage](#)  
[The R Journal](#)

## Software

[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

## Documentation

[Manuals](#)  
[FAQs](#)  
[Contributed](#)

Subdirectories:

[base](#)

[contrib](#)

[Rtools](#)

Binaries for base distribution (managed by Duncan Murdoch). This is what you want to [install R for the first time](#).

Binaries of contributed packages (managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN services and corresponding environment and make variables.

Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or itself.

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to this.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

R for Windows



R-2.15.1 for Windows (32/64 bit)

[Download R 2.15.1 for Windows](#) (47 megabytes, 32/64 bit)

[Installation and other instructions](#)

[New features in this version](#)

#### CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

#### About R

[R Homepage](#)

[The R Journal](#)

#### Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

#### Documentation

[Manuals](#)

[FAQs](#)

[Contributed](#)

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

#### Frequently asked questions

- [How do I install R when using Windows Vista?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

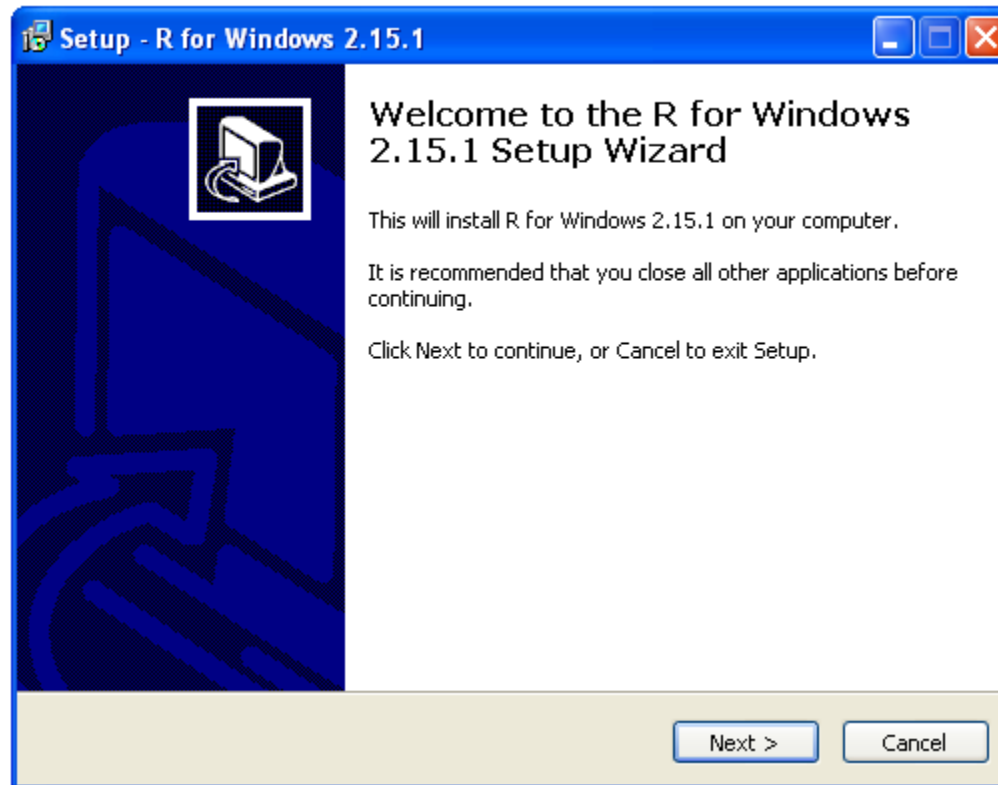
#### Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel](#) build.
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is <http://<CRAN MIRROR>/bin/windows/base/release.htm>.

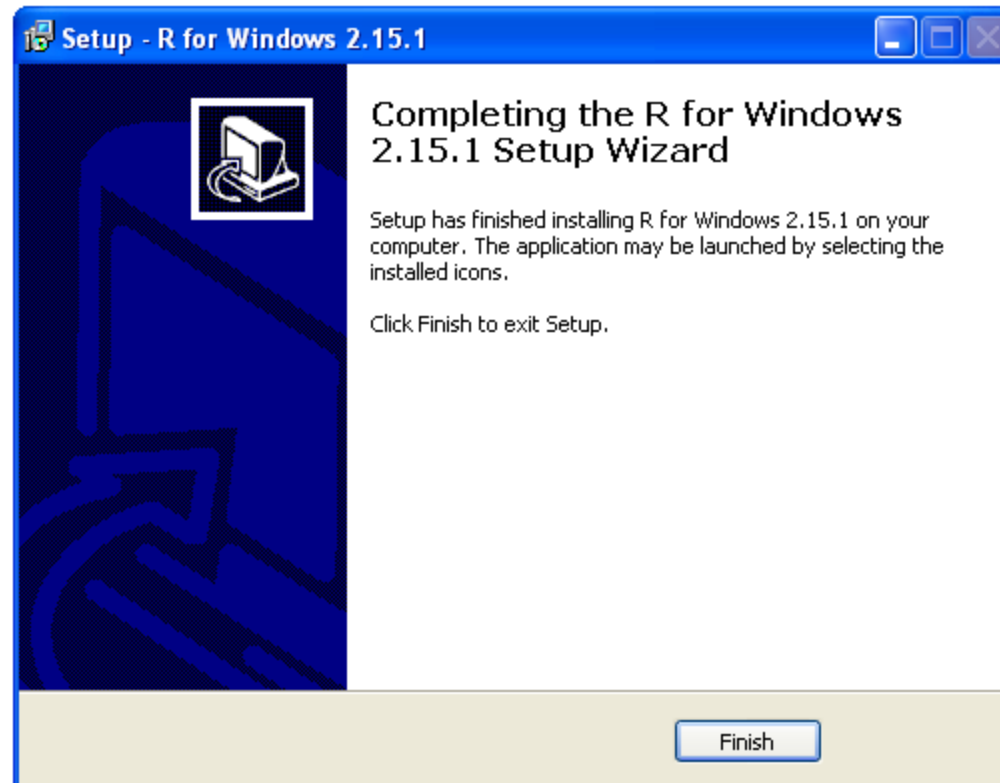


- ❑ An executable file will be downloaded automatically.
- ❑ R.2.15.1.exe
- ❑ Simply click and follow the instructions!
- ❑ Es possible elegir Español



Press 'next' until...

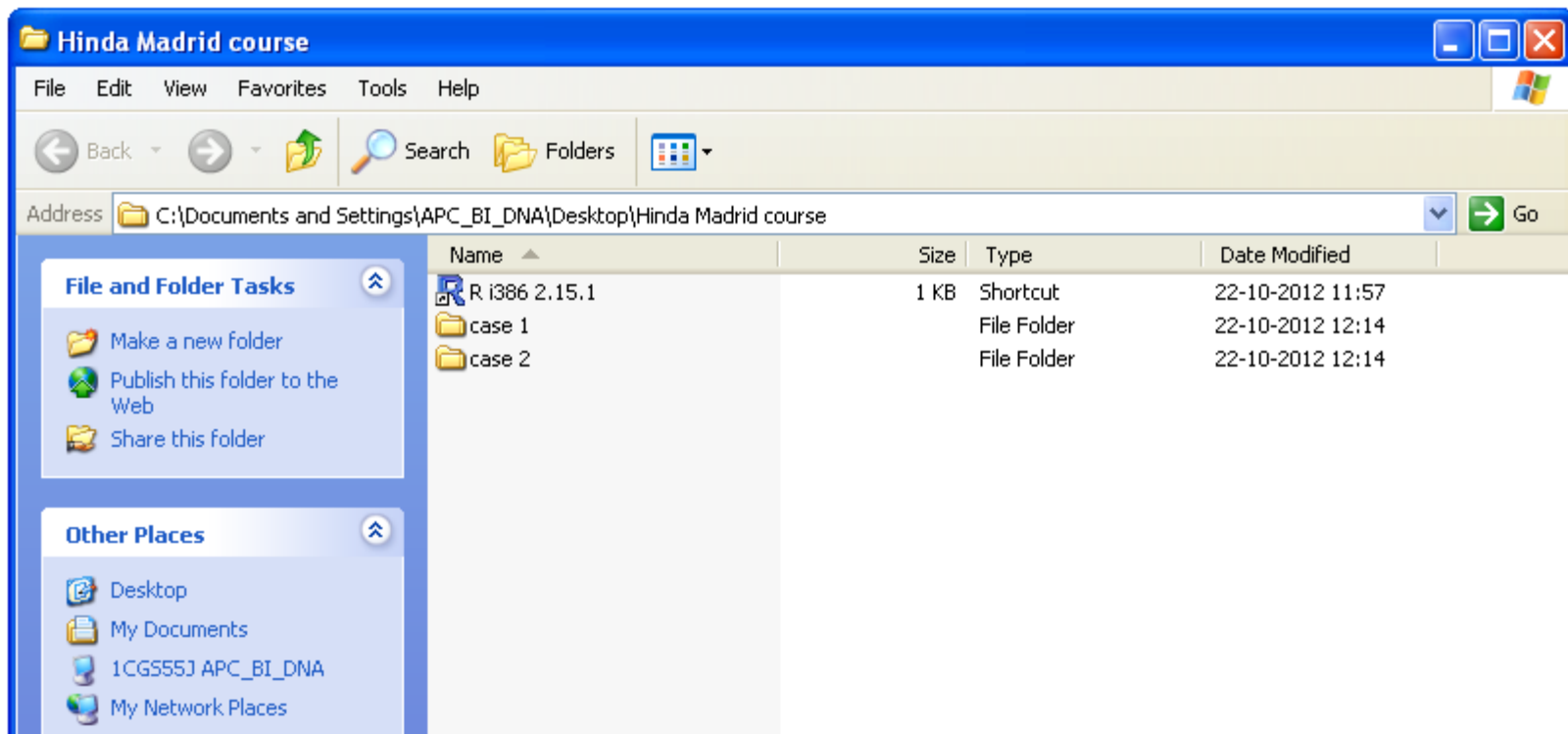






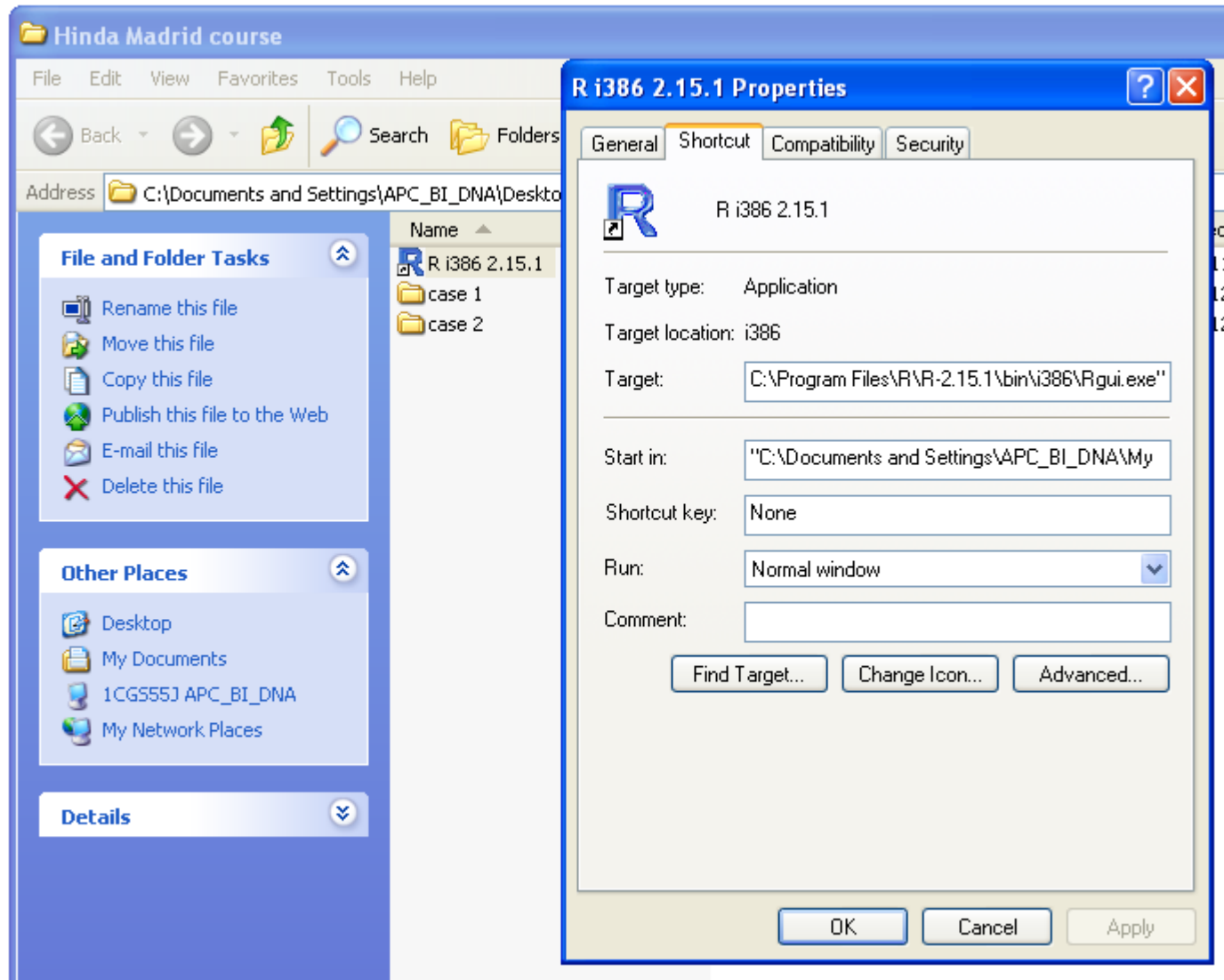
# Prepare your working folder first!

- Create a folder where you will put your cases, notes, lectures, etc
- Copy the blue R Icon in the folder



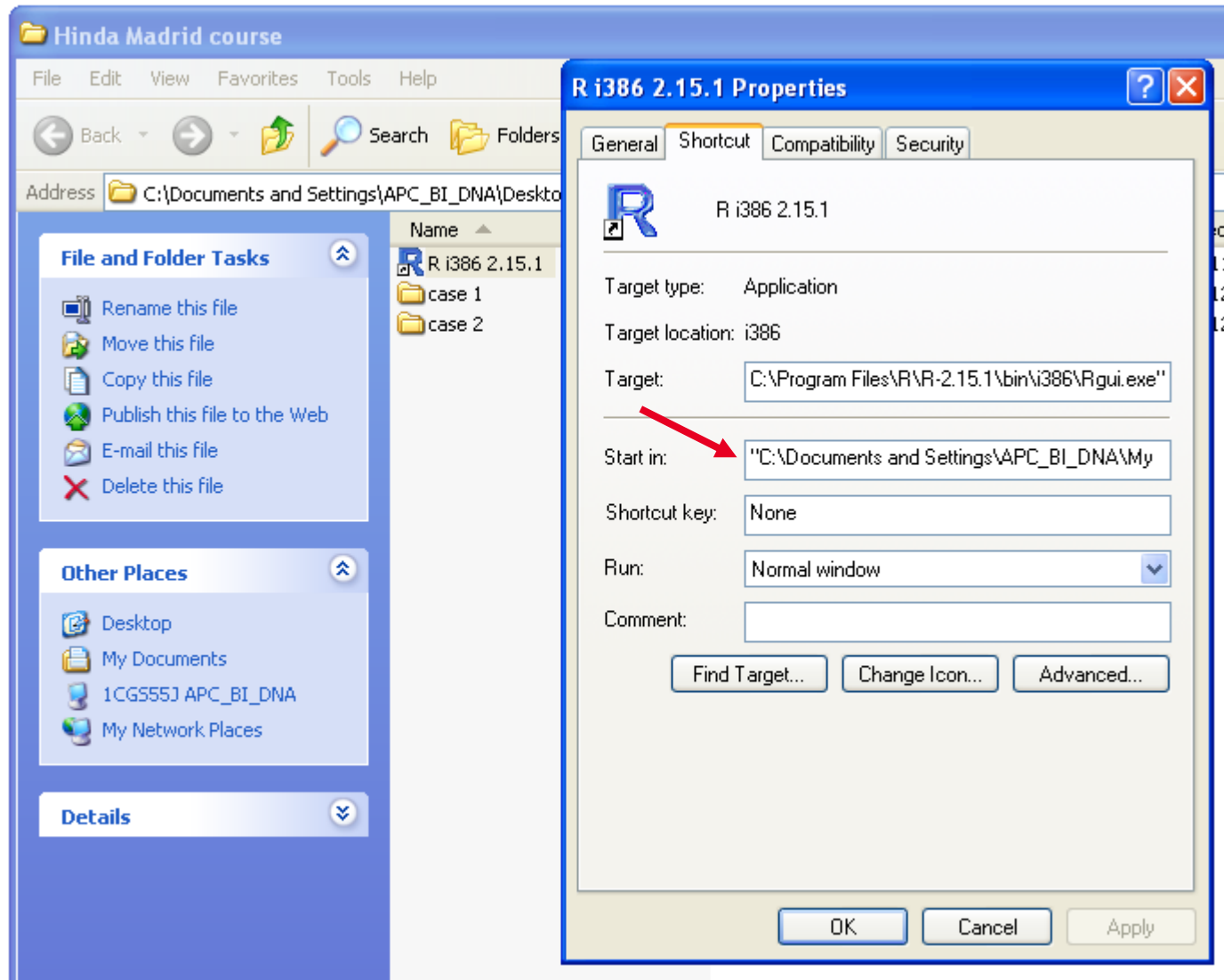


# A little trick to make your life easier...





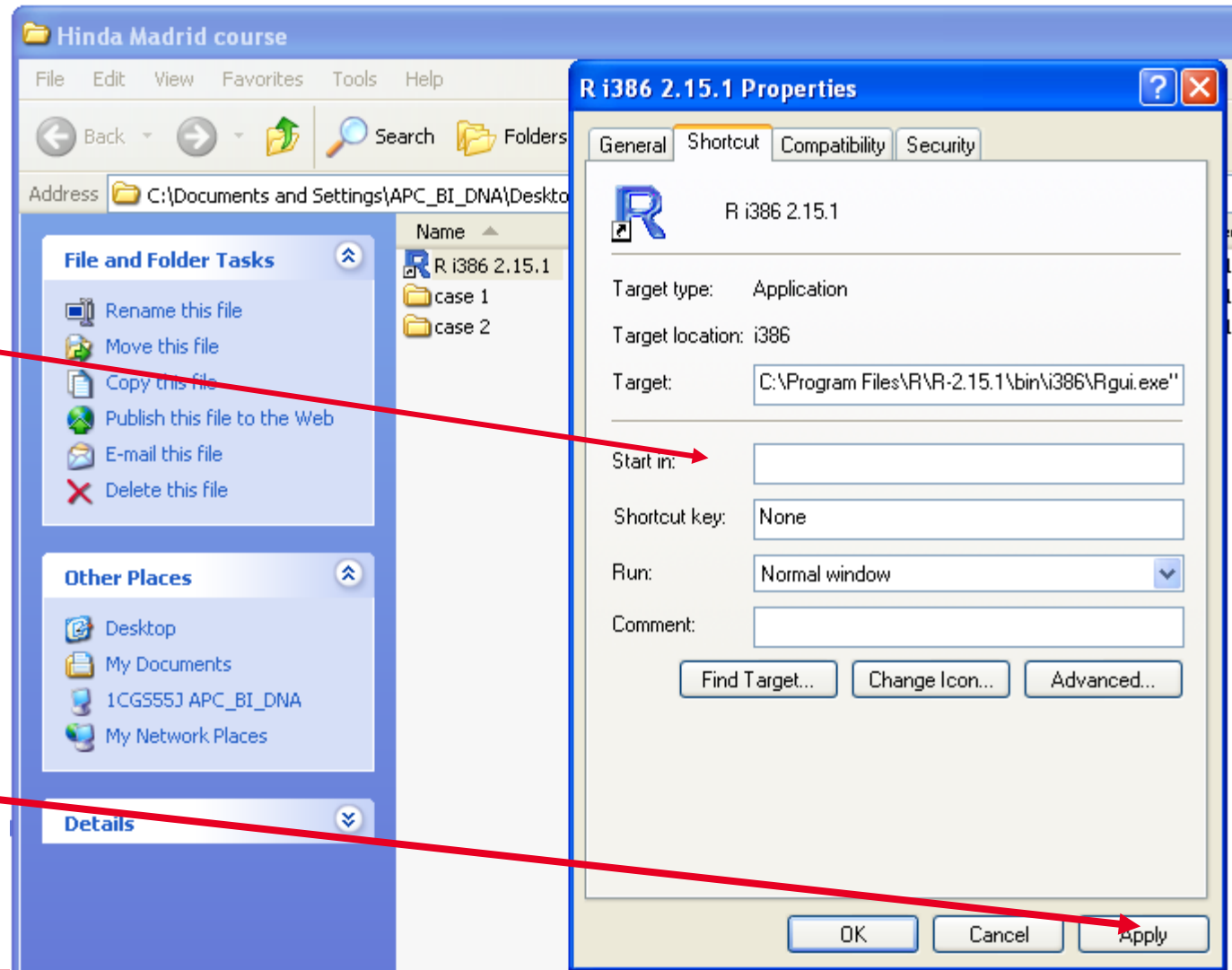
# A little trick to make your life easier...





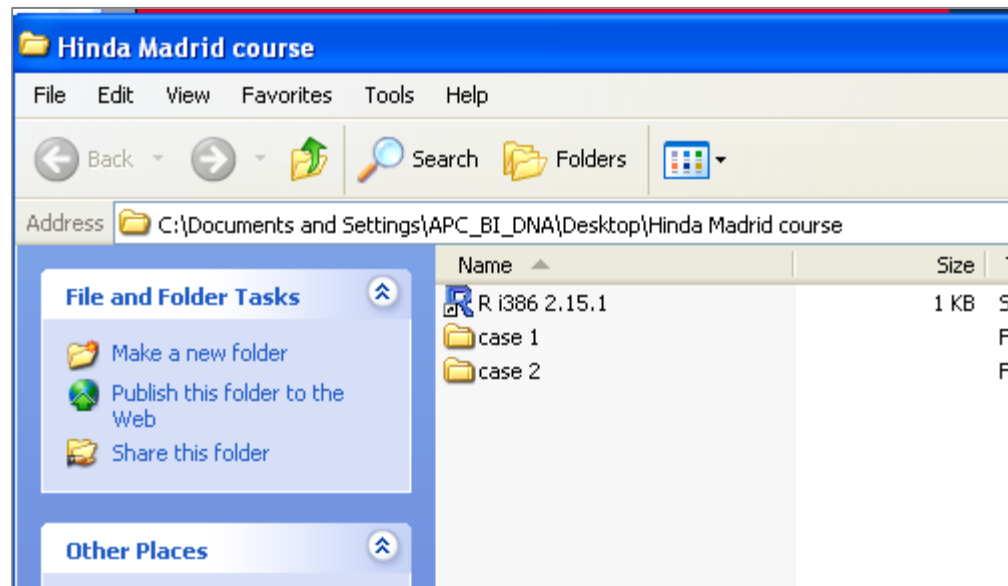
Leave it  
blank

Press Apply





# You are now ready to launch R



## Simply click the blue Icon!

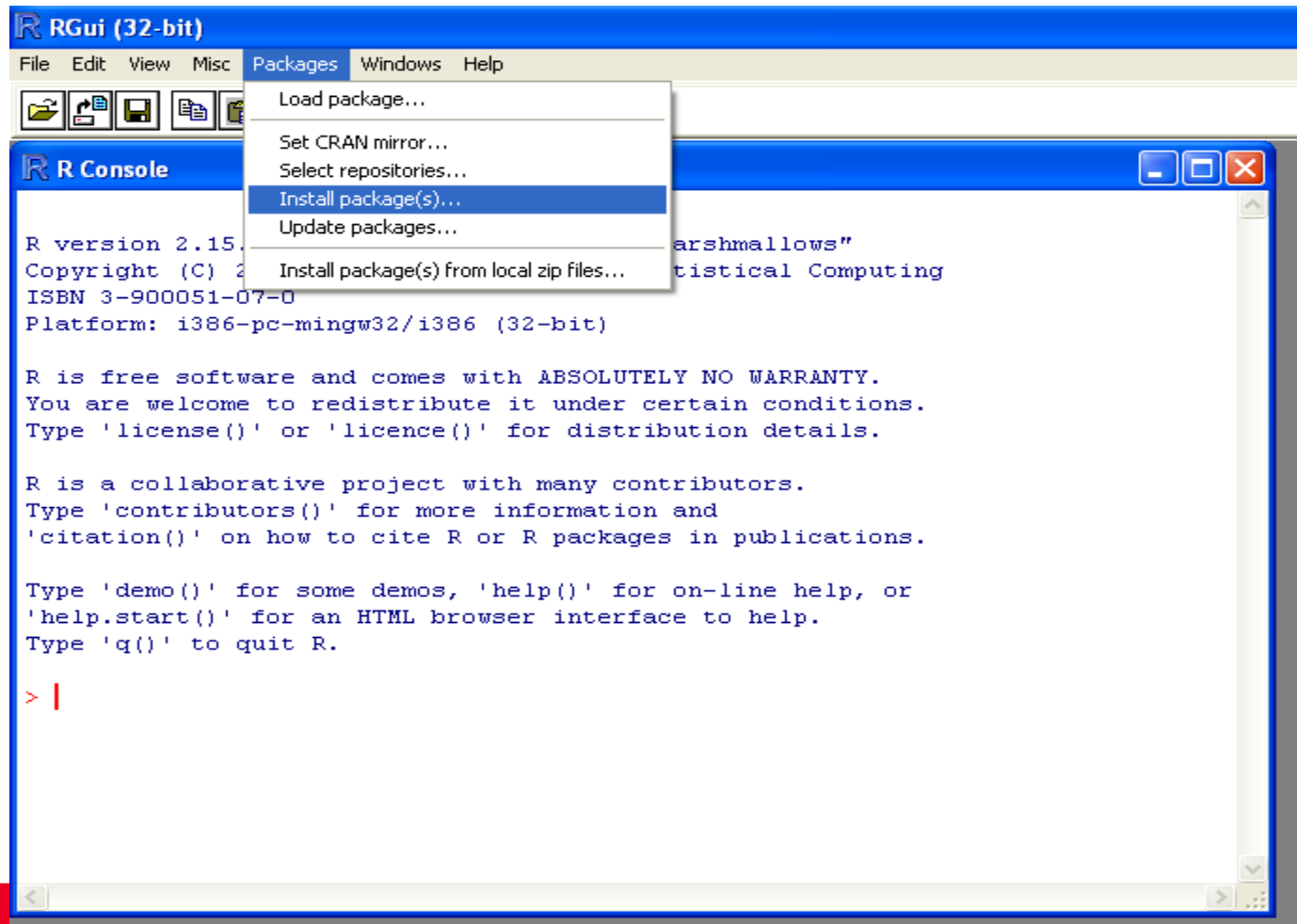


## (2) Install the Forensim package

- ❑ Option 1: install the package directly from the R environment (Internet connexion)
- ❑ Option 2: Install the package manually (no Internet connexion)
  - ➡ Refer to LRmix tutorial online



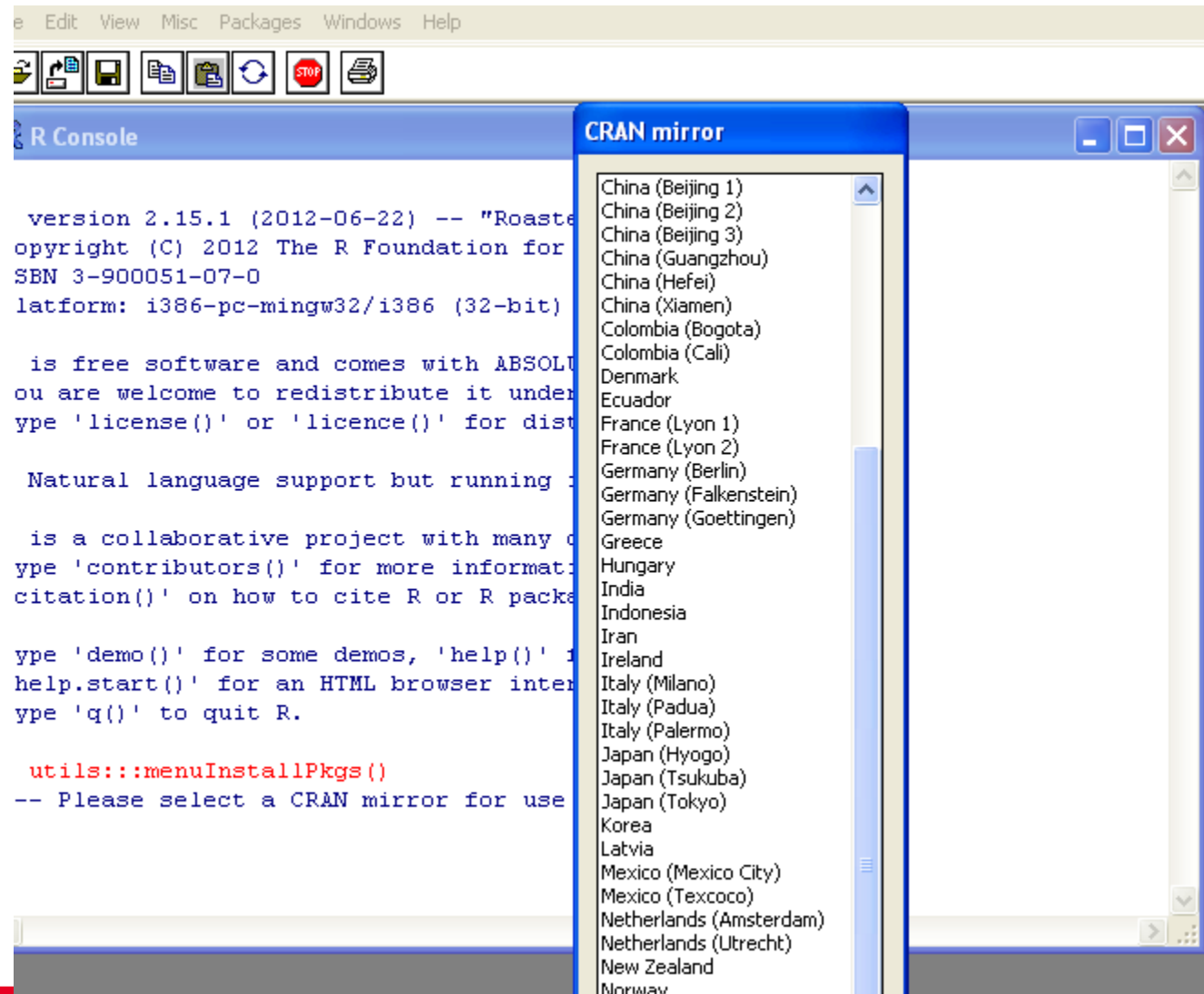
## (2) Install the Forensim package





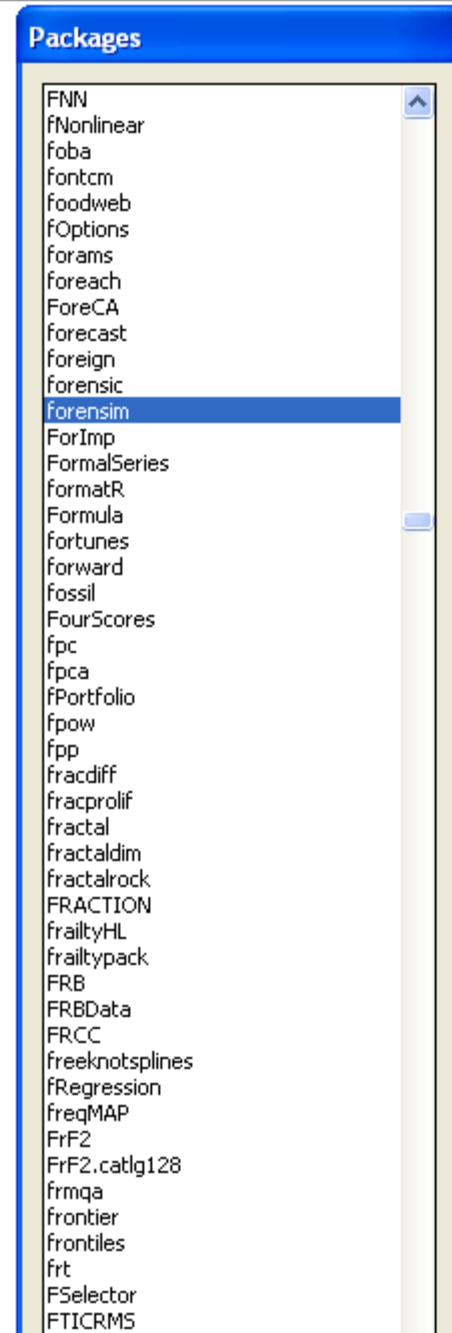


## Choose mirror (in Spain)



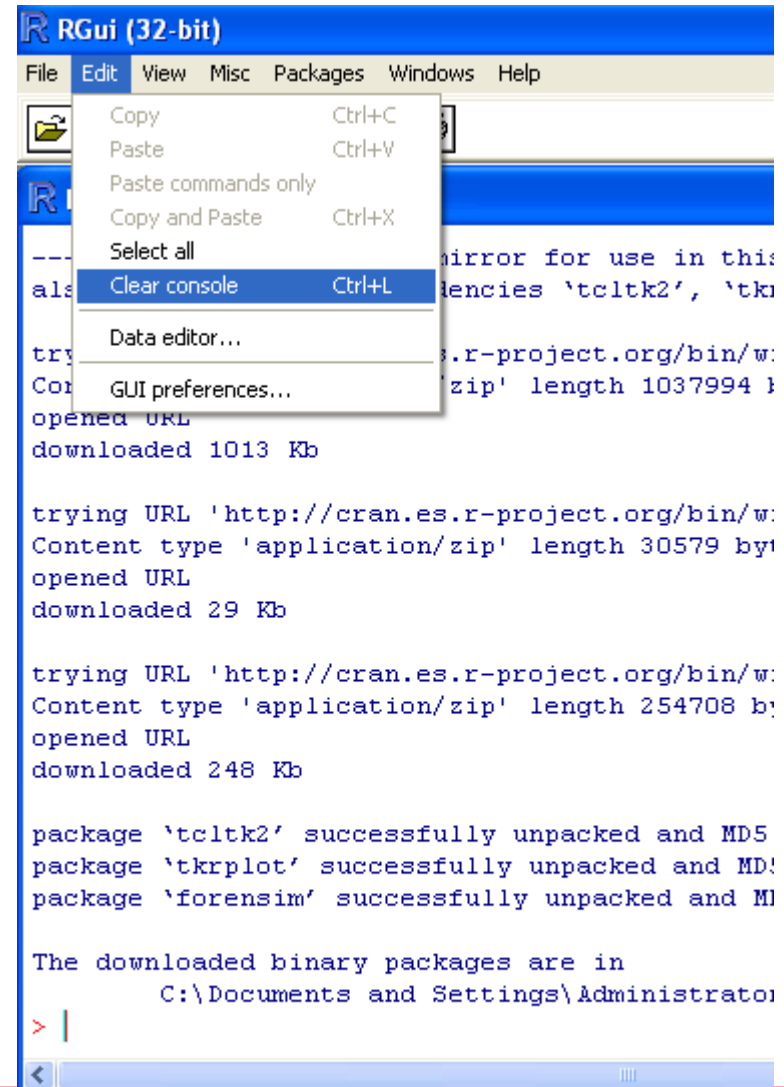


Choose package forensim





Clear console for better  
visibility  
(nothing will be deleted)

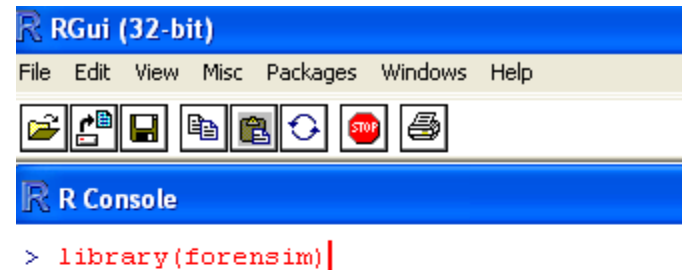




### (3) Load the Forensim library

Type the following code in the R console:

```
library(forensim)
```






### (3) Load the Forensim library

Type the following code in the R console:

```
library(forensim)
```



```
RGui (32-bit) - [R Console]
File Edit View Misc Packages Windows Help

> library(forensim)
Loading required package: tcltk
Loading Tcl/Tk interface ... done
Loading required package: tcltk2
Loading required package: tkrplot
> |
```



## (4) Start LRmix

Type the following code in the R console:

```
library(forensim)  
LRmixTK()
```

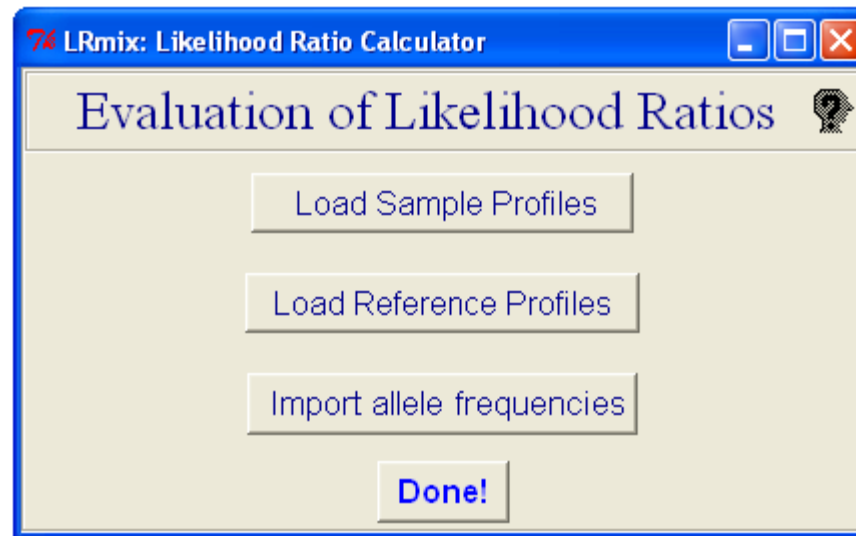
```
RGui (32-bit) - [R Console]  
File Edit View Misc Packages Windows Help  
[Icons]  
  
> library(forensim)  
Loading required package: tcltk  
Loading Tcl/Tk interface ... done  
Loading required package: tcltk2  
Loading required package: tkrplot  
> LRmixTK()  
<Tcl>  
> |
```



Illustration of the LRmix module using the Hammer case, Published in Gill et al, FSIG, 2007



# Main LRmix interface







## Input files in LRmix

Type 1: CSV files, they are comma separated files (`,`), and the decimal separator is the dot (`.')

Type 2: tab separated files, they are tab separated (`\t`, e.g. Excel), and the dot (`.')

Never use spaces in your column-names, or in the sample-names (epg, or references)



# Installing OpenOffice could greatly assist!!

Open office is the equivalent of Microsoft Office, except:

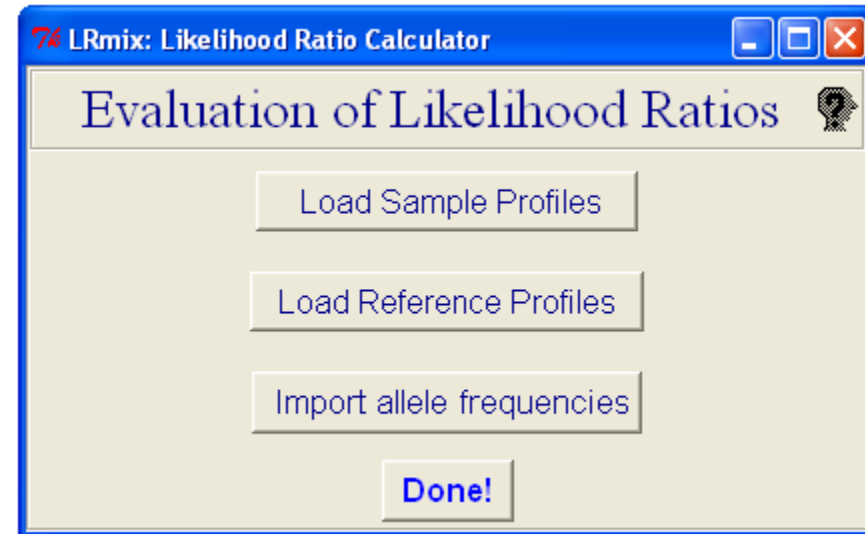
- It is open source
- Free of charge
- Much more efficient when it comes to visualising data
- You can visualise data much more easily than with Excel or notepad

[www.openoffice.org](http://www.openoffice.org)



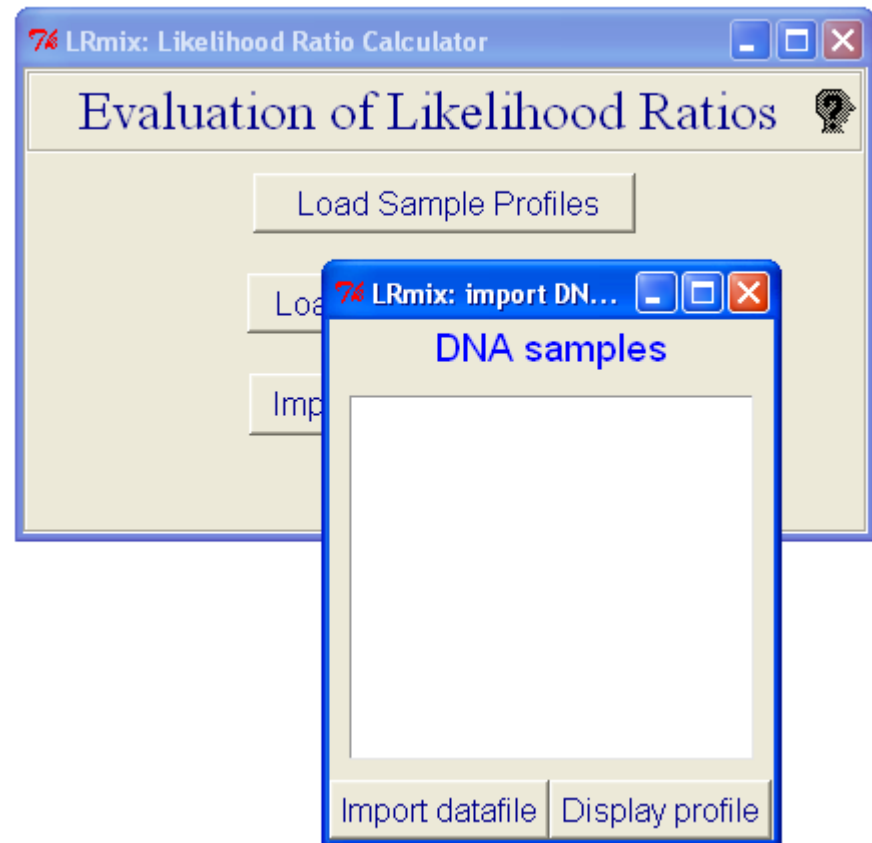
## Main LRmix interface

- (1) Load the crime-sample profile
- (2) Load the references  
(suspect/victim)
- (3) Load your allele frequencies





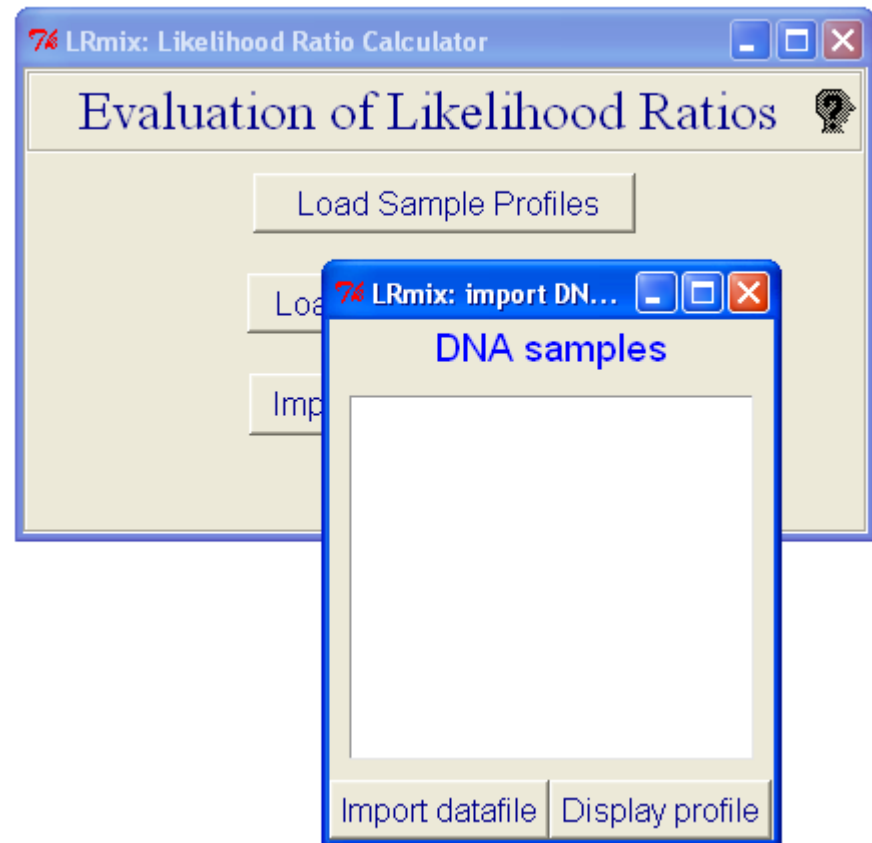
## (1) Load the crime-sample profiles





## (1) Load the crime-sample profiles

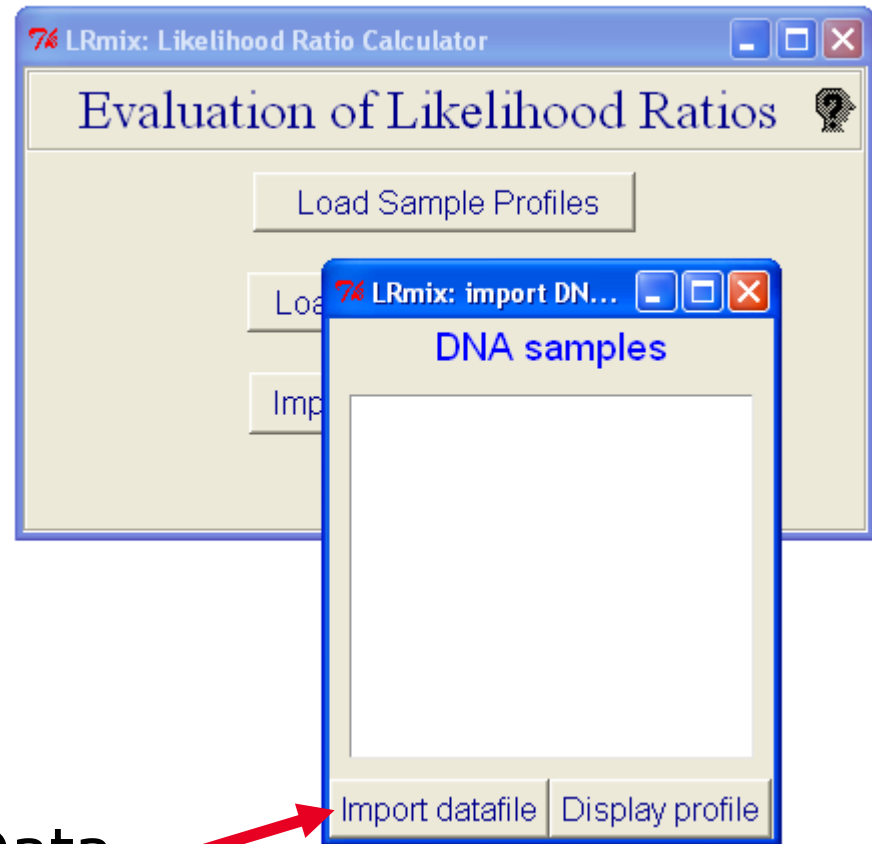
sampleHammer.txt





## (1) Load the crime-sample profiles

sampleHammer.txt



Import Data

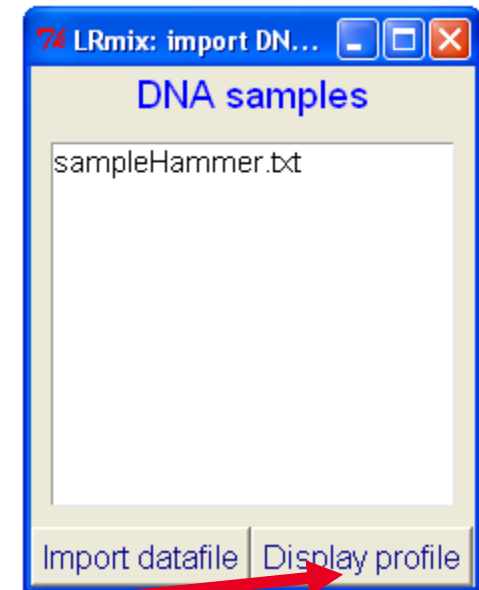


## (1) Load the crime-sample profiles

sampleHammer.txt

Display profiles,

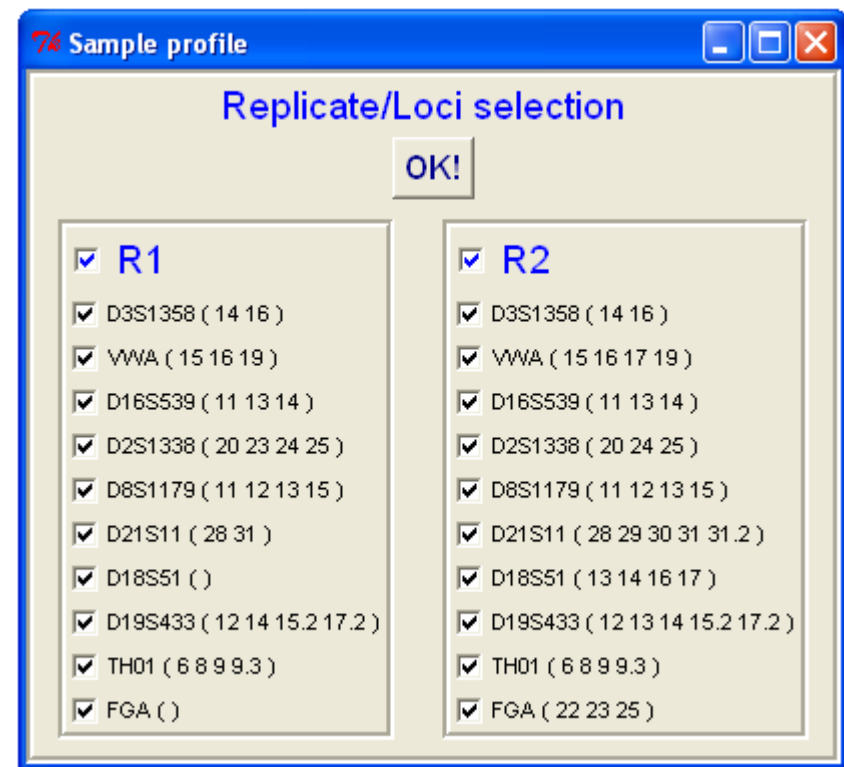
To make sure the data are OK





If everything looks good, press OK!

- You can select loci
- You can select replicates



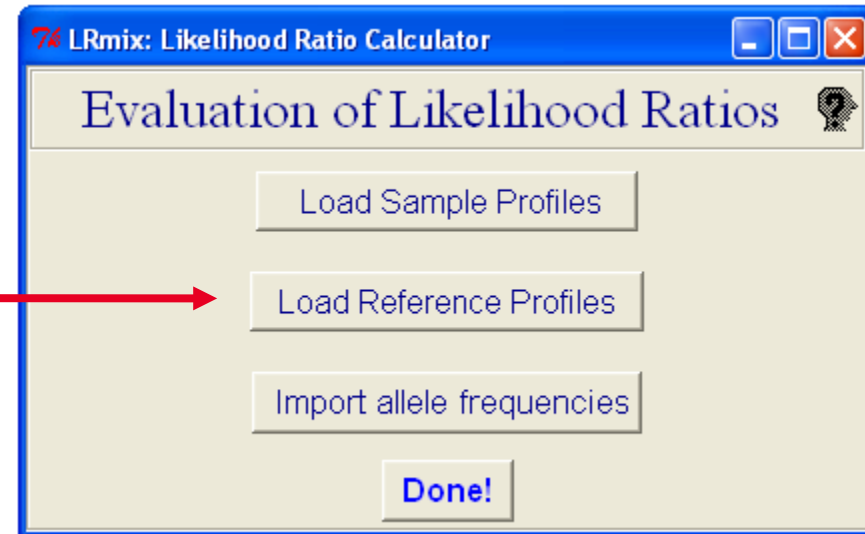


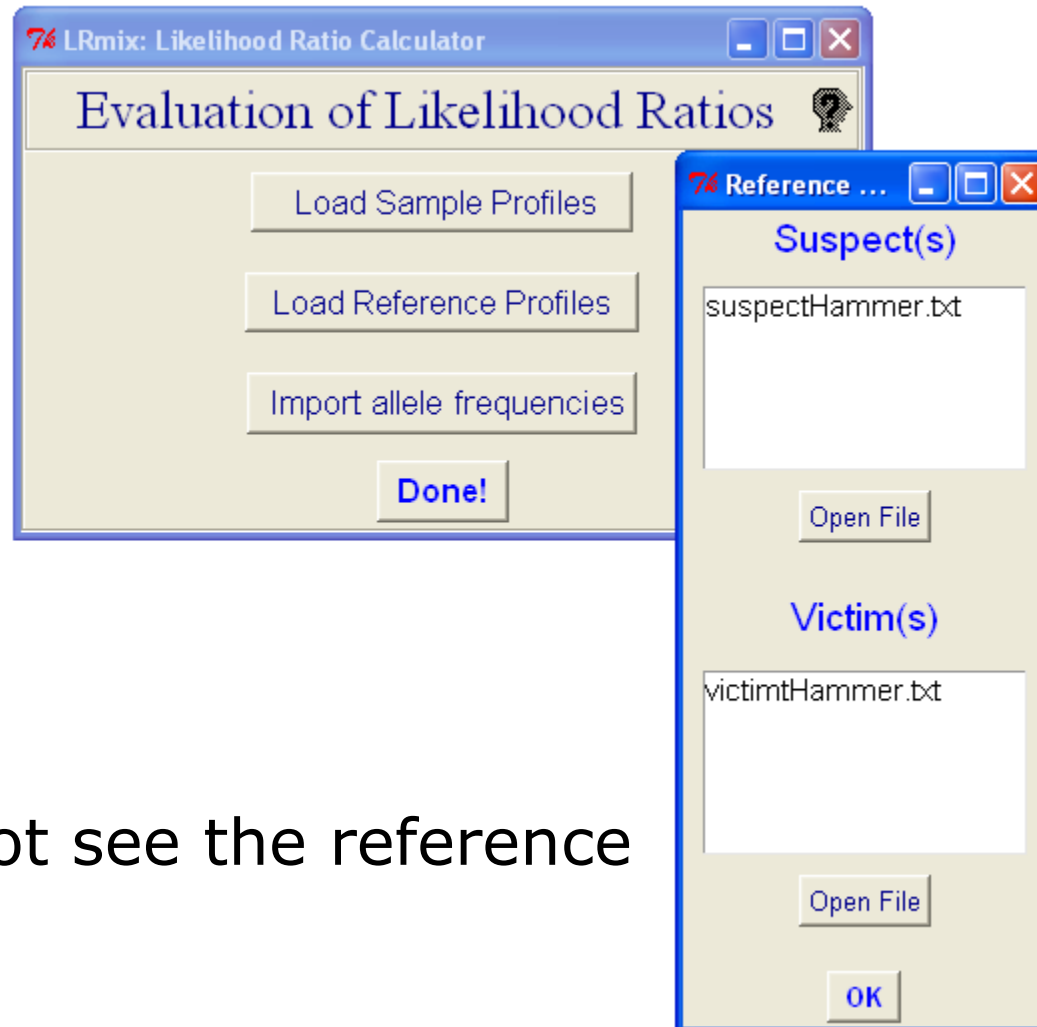


## (2) Load reference profiles

suspectHammer.txt

victimHammer.txt





You cannot see the reference profiles

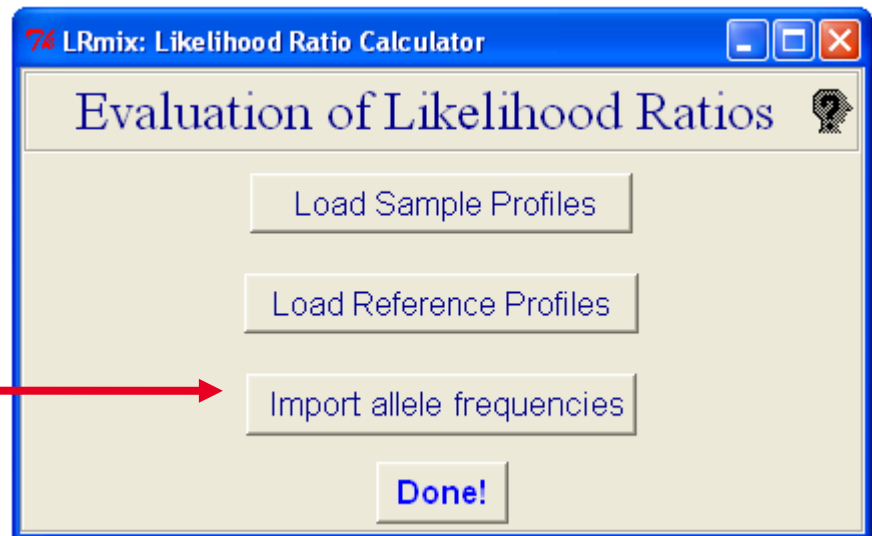


- ❑ You cannot see the reference profiles
- ❑ The program will automatically select the loci you chose in step (1)
- ❑ If there are loci in the epg that are not given in the reference profile, the program will give an error message



### (3) Import the allele frequencies

identifilerUSA.txt





You may need help building frequencies file for the first time!

Allele	CSF1PO	FGA	TH01	TPOX	VWA	D3S1358
5			0.002	0.002		
6			0.232	0.002		
7			0.190			
8	0.005		0.084	0.535		
8.1						
...	...	...	...	...	...	...
16.2						
17					0.281	0.215
17.2						
18		0.026			0.200	0.152
18.2						
19		0.053			0.104	0.012
19.2						
20		0.127			0.005	0.002
21		0.185			0.002	
21.2		0.005				
22		0.219				
22.2		0.012				
...	...	...	...	...	...	...



**Analyse the profiles**

### Hypotheses

**Contributors under  $H_p$**

suspect

☒ victim1

☒ victim2

**Contributors under  $H_d$**

☒ victim1

☒ victim2

### Parameters

**Unknown contributors**

Under  $H_p$

Under  $H_d$

**$\Pr(D)$ ,  $\Pr(C)$ ,  $\theta$**

Probability of Dropout  $\Pr(D)$

Probability of Contamination  $\Pr(C)$

Theta Correction (Fst)

### Performance test

**Choose suspect**

suspect ☐

number of iterations



- Don't worry about the drop-out parameter for now, you will be able to carry out a sensitivity plot in the following steps
- We will address the performance tests later on today



Hp: Victim 1 + victim 2+ Suspect  
Hd: Victim 1+ Victim 2+ Unknown

Drop-in=0.05

Drop-out=0.10

Theta=0



The screenshot shows the 'LRmix: Results' window. It features a table of results for various loci, with columns for the locus name and the LR value. To the right of the table is a box for the 'Overall LR' value. At the bottom of the window are two buttons: 'Plot LR vs PrD' and 'Export results'.

{LR per Locus}	LR
D3S1358	15.19
VWA	0.7854
D16S539	78.33
D2S1338	2.409
D8S1179	4.618
D21S11	37.95
D18S51	0.06652
D19S433	8660
TH01	11.9
FGA	8.671

{Overall LR}  
2.345e+10

Plot LR vs PrD      Export results





Hp: Victim 1 + victim 2+ Suspect  
Hd: Victim 1+ Victim 2+ Unknown

Drop-in=0.05

Drop-out=0.10

Theta=0

Sensitivity plot

{LR per Locus}	LR
D3S1358	15.19
VWA	0.7854
D16S539	78.33
D2S1338	2.409
D8S1179	4.618
D21S11	37.95
D18S51	0.06652
D19S433	8660
TH01	11.9
FGA	8.671

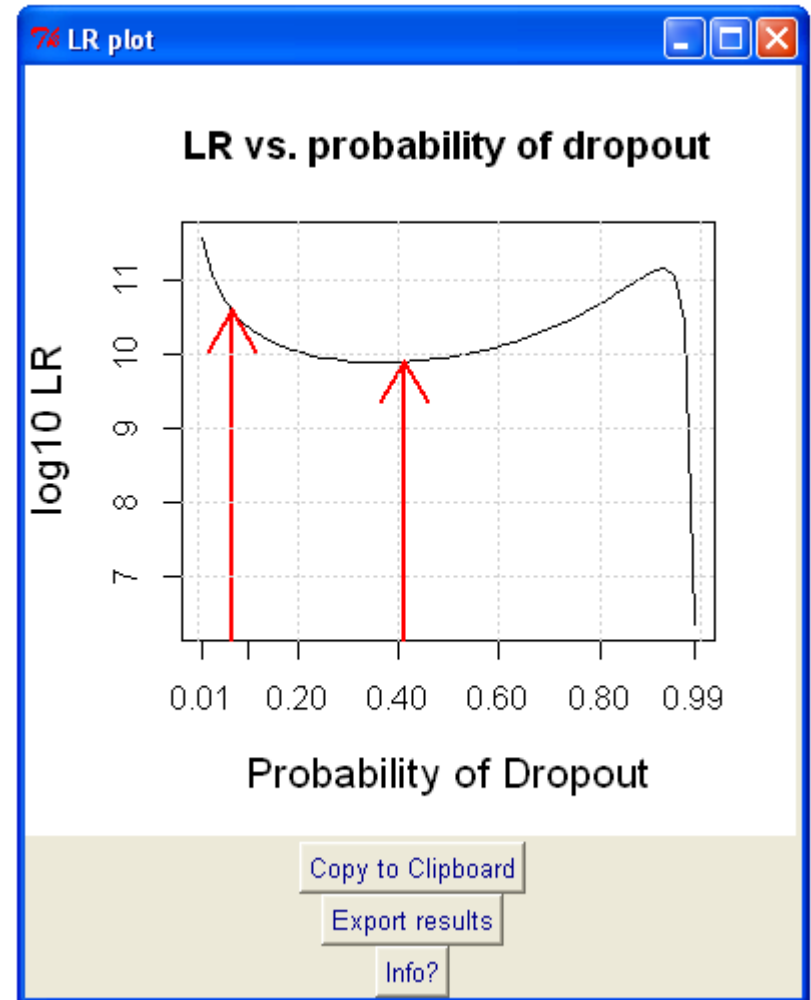
{Overall LR}  
2.345e+10

Plot LR vs PrD      Export results



# Sensitivity analysis

## Drop-out ranges [Export results]





# Hypotheses

(I)

Hp: Victim 1 + victim 2+ Suspect

Hd: Victim 1+ Victim 2+ Unknown

(II)

Hp: Victim 1 + Unknown+ Suspect

Hd:Victim 1+ 2 Unknowns

(III)

Hp: Victim 2 + Unknown+ Suspect

Hd: Victim 1+ 2 Unknowns