ForenStatCI version 1.0.0 User Manual

31th August 2019

Sho Manabe

Department of Forensic Medicine, Kyoto University Graduate School of Medicine

manabe@fp.med.kyoto-u.ac.jp

Contents

1. What is *ForenStatCI*?

ForenStatCI is an open-source software for estimating credible intervals of the strength of DNA evidence interpretation. The software is a graphical user interface written in R language, and the source code is freely available at GitHub (https://github.com/manabe0322/ForenStatCI/releases).

Based on allele counts of the population database, the software automatically calculates the parameters of the Dirichlet distribution and expected allele frequencies. The credible intervals of four statistical indexes (i.e., random match probability, combined probability of inclusion, LR for mixture interpretation using a binary model, LR for pairwise kinship analysis) are estimated based on the Dirichlet distribution. Users can set the number of random sampling of allele frequencies, a theta correction to consider subpopulation effect, and parameters of the prior Dirichlet distribution (i.e., 1 (a uniform prior) or 1 / K, where K is the number of alleles in a locus). The software also calculates the expected values of the four indexes based on the expected allele frequencies.

2. Tutorial

2.1. Getting started

First, ensure that the R software has been installed. It is available from the R Development Core Team website (http://www.R-project.org).

The *ForenStatCI* program is freely available at GitHub (https://github.com/manabe0322/ForenStatCI/releases) and can be accessed by clicking the file named “ForenStatCI v1.0.0.RData”.

After the installation of the R software and the *ForenStatCI* program, start an R session. Subsequently, load the “ForenStatCI v1.0.0.RData” file from “Load Workspace” in the “File” tab (Fig. 1).

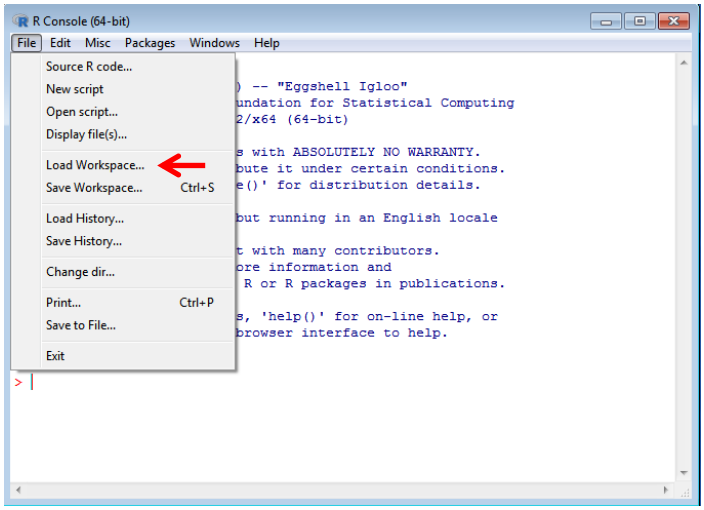

Fig. 1. “Load Workspace” in the “File” tab.

After loading the file, the *ForenStatCI* software is launched by the following command:

```
ForenStatCI()
```

All required packages used in *ForenStatCI* (tcltk, tcltk2, MCMCpack, and gtools) are automatically installed.

After all packages are loaded, the main window opens as shown in Fig. 2. There are five functions in the *ForenStatCI* as follows:

- Estimate expected allele frequencies
- Estimate the credible interval of the random match probability
- Estimate the credible interval of the combined probability of inclusion
- Estimate the credible interval of the likelihood ratio for mixture interpretation
- Estimate the credible interval of the likelihood ratio for kinship analysis

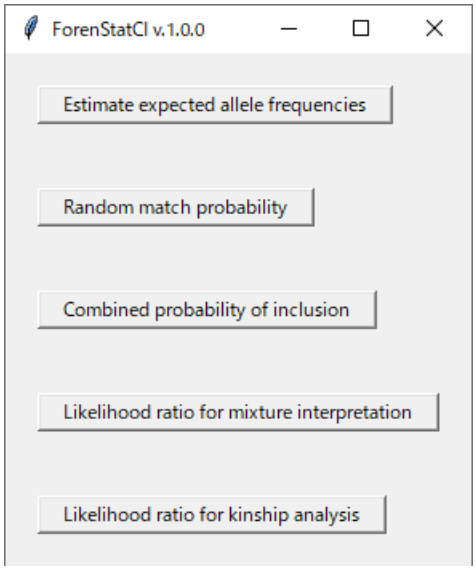

Fig. 2. The figure shows the main window of the *ForenStatCI*.

2.2. Estimate expected allele frequencies

To estimate expected allele frequencies, press the "Estimate expected allele frequencies" button after the main window is opened as shown in Fig. 2. Then a new window is opened as shown in Fig. 3.

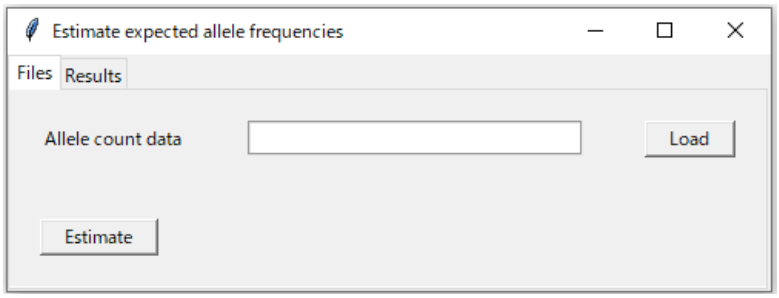

Fig. 3. The "Files" tab of "Estimate expected allele frequencies".

The user can input a .csv file of the observed number of alleles in a population database from the "Load" button. Fig. 4 shows an example of the observed number of alleles in each locus (Fujii K. *et al*. *Leg Med (Tokyo)*. 2015; 17(5):306-8.).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA
2	5											4	
3	6											663	
4	7			1	37		1				1	811	
5	8			5	3	1357				7		194	
6	8.1									1			
7	9			1077	151	354	8					1190	
8	9.1									110			
9	9.2										1		
10	9.3											113	
11	10			602	648	100	387		6	805		27	
12	10.1									7			
13	10.2										3		
14	10.3												
15	11			562	620	1072	319		14	1034	10		
16	11.1									1			

Fig. 4. An example of the observed number of alleles in each locus.

After loading the observed number of alleles, press the “Estimate” button. Then the “Results” tab is opened as shown in Fig. 5.

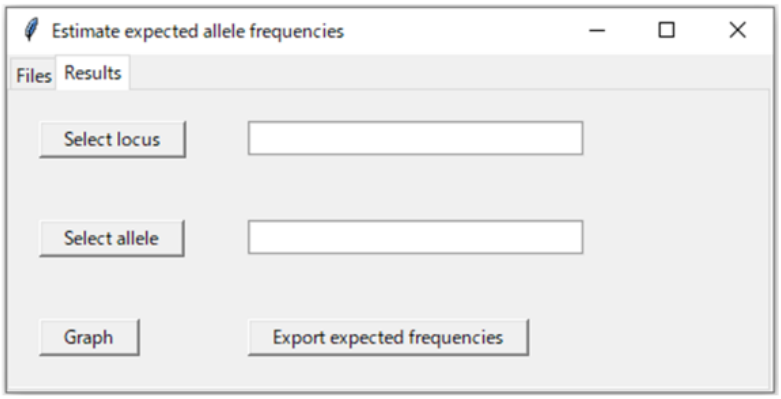

Fig. 5. The “Results” tab of “Estimate expected allele frequencies”.

To draw the Dirichlet distribution and the expected frequency of an allele in a locus, the user can select a locus and an allele from the “Select locus” button and the “Select allele” button, respectively. The Dirichlet distribution and the expected frequency are displayed by pressing the “Graph” button. Fig. 6 shows an example of the Dirichlet distribution of allele 12 in D3S1358.

The user can export the expected allele frequencies of all loci into a .csv file by pressing the “Export expected frequencies” button. Fig. 7 shows an example of the expected allele frequencies of all loci.

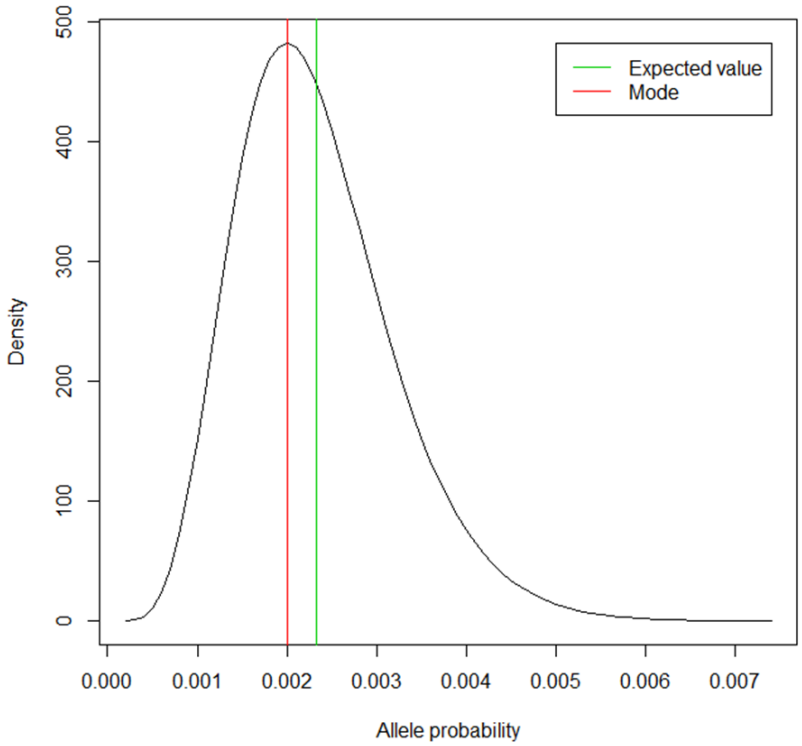

Fig. 6. An example of the Dirichlet distribution of allele 12 in D3S1358. The green line is the expected frequency of allele 12. The red line is the mode of the frequency.

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1		D3S1358	vWA	D16S539	CSF1PO	TPOX	D8S1179	D21S11	D18S51	D2S441	D19S433	TH01	FGA	D
2	5											0.001662		
3	6											0.220671		
4	7			0.000664	0.012616		0.000664				0.000662	0.269857		
5	8			0.001993	0.001328	0.451313				0.002653		0.064806		
6	8.1									0.000663				
7	9			0.358021	0.050465	0.117979	0.002987					0.395813		
8	9.1									0.036816				
9	9.2										0.000662			
10	9.3											0.037886		
11	10			0.200266	0.215471	0.033566	0.128775		0.002317	0.26733		0.009305		
12	10.1									0.002653				
13	10.2										0.001325			
14	10.3													
15	11			0.186981	0.206175	0.356597	0.106206		0.004965	0.343284	0.003644			(
16	11.1									0.000663				

Fig. 7. An example of the expected allele frequencies of all loci.

2.3. Estimate the credible interval of the random match probability

To estimate the credible interval of the random match probability (RMP), press the "Random match probability" button after the main window is opened as shown in Fig. 2. Then a new window is opened as shown in Fig. 8.

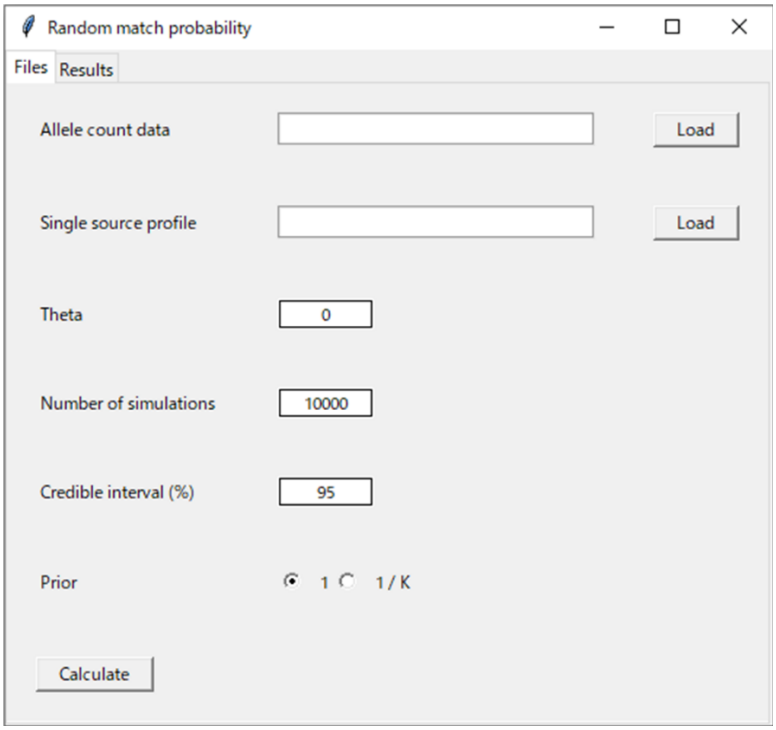

Fig. 8. The "Files" tab of "Random match probability".

The user can input a .csv file of the observed number of alleles in a population database (Fig. 4) and that of a single source profile (Fig. 9) from the "Load" button.

	A	B	C	D
1	Sample Name	Marker	Allele 1	Allele 2
2	Example	D3S1358	15	16
3	Example	vWA	16	16
4	Example	D16S539	10	11
5	Example	CSF1PO	11	11
6	Example	TPOX	11	11
7	Example	D8S1179	11	13
8	Example	D21S11	29	31.2
9	Example	D18S51	13	14
10	Example	D2S441	10	14
11	Example	D19S433	12	16.2
12	Example	TH01	6	9
13	Example	FGA	22	24
14	Example	D22S1045	11	16
15	Example	D5S818	9	13
16	Example	D13S317	11	13
17	Example	D7S820	10	12
18	Example	SE33	17	30.2
19	Example	D10S1248	13	13
20	Example	D1S1656	15	17.3
21	Example	D12S391	19	20
22	Example	D2S1338	18	24

Fig. 9. An example of the single source profile.

The user then set a theta correction to consider subpopulation effect, the number of random sampling of allele frequencies, the range of the credible interval (%), and parameters of the prior Dirichlet distribution (i.e., 1 (a uniform prior) or 1 / K, where K is the number of alleles in a locus). After imputing two required files and setting each parameter, press the “Calculate” button. Then the “Results” tab is opened as shown in Fig. 10.

The credible interval of the user-selected range, quantiles of simulated RMP values, and the expected RMP value are displayed in the “Results” tab. The RMP distribution, the credible interval, and the expected value are displayed by pressing the “Graph” button (Fig. 11). The report can be exported into a .csv file by pressing the “Report” button. All simulated RMP values can be exported as well into a.csv file by pressing the “Save all simulated RMP” button.

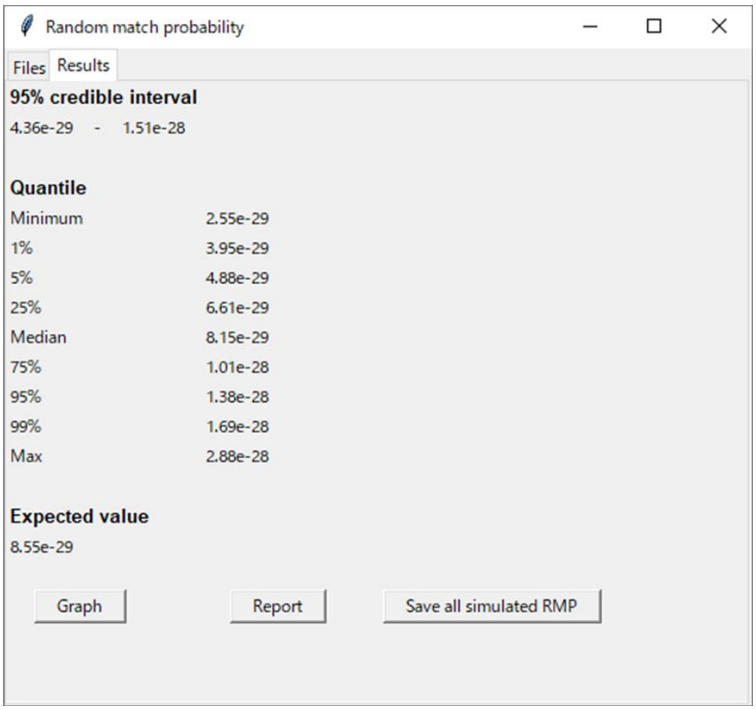

Fig. 10. The "Results" tab of "Random match probability".

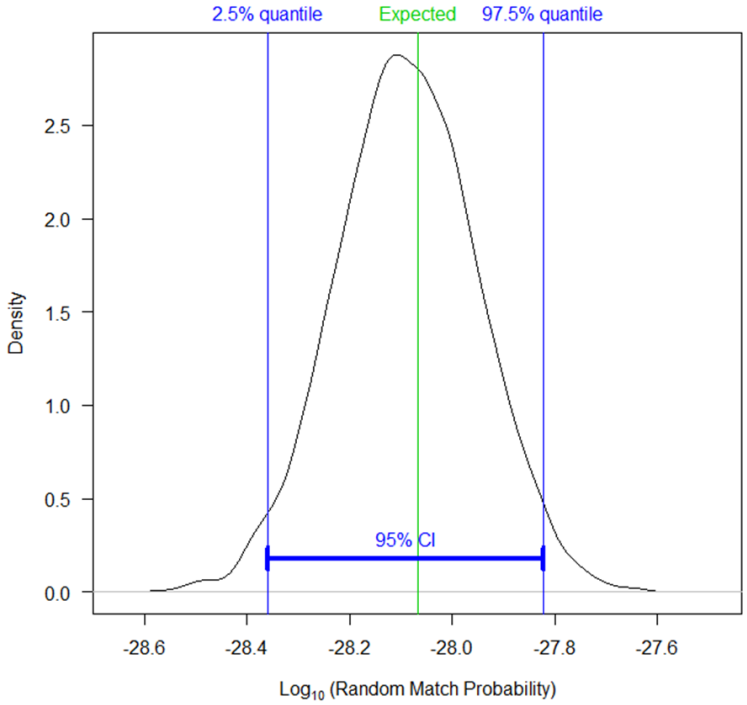

Fig. 11. An example of the RMP distribution. The blue lines represent the credible interval. The green line is the expected RMP value.

2.4. Estimate the credible interval of the combined probability of inclusion

To estimate the credible interval of the combined probability of inclusion (CPI), press the "Combined probability of inclusion" button after the main window is opened as shown in Fig. 2. Then a new window is opened as shown in Fig. 12.

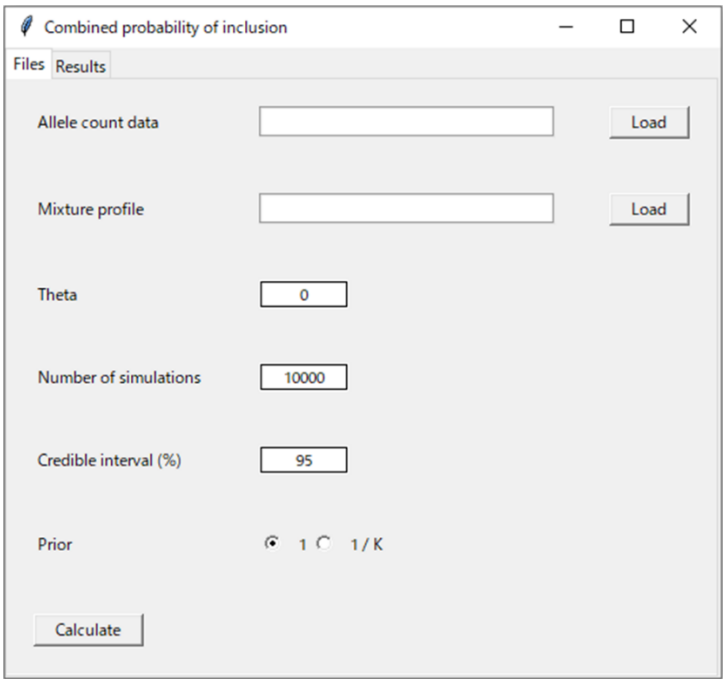

Fig. 12. The "Files" tab of "Combined probability of inclusion".

The user can input a .csv file of the observed number of alleles in a population database (Fig. 4) and that of a mixture profile (Fig. 13) from the "Load" button.

	A	B	C	D	E	F
1	Sample Name	Marker	Allele 1	Allele 2	Allele 3	Allele 4
2	Example	D3S1358	15	16	17	18
3	Example	vWA	16	17	18	
4	Example	D16S539	9	10	11	12
5	Example	CSF1PO	10	11	12	13
6	Example	TPOX	8	11		
7	Example	D8S1179	10	14	15	
8	Example	D21S11	29	30	32.2	
9	Example	D18S51	12	13	15	22
10	Example	D2S441	9.1	11		
11	Example	D19S433	13	15	15.2	16.2
12	Example	TH01	7	9	9.3	
13	Example	FGA	19	21	25.2	
14	Example	D22S1045	11	16	17	
15	Example	D5S818	10	11		
16	Example	D13S317	9	10	11	12
17	Example	D7S820	9	10	12	
18	Example	SE33	17	21	25.2	28.2
19	Example	D10S1248	13	14	15	16
20	Example	D1S1656	14	15	16	
21	Example	D12S391	18	19	22	
22	Example	D2S1338	18	24		

Fig. 13. An example of the mixture profile.

The user then set a theta correction to consider subpopulation effect, the number of random sampling of allele frequencies, the range of the credible interval (%), and parameters of the prior Dirichlet distribution. After imputing two required files and setting each parameter, press the “Calculate” button. Then the “Results” tab is opened as shown in Fig. 14.

The credible interval of the user-selected range, quantiles of simulated CPI values, and the expected CPI value are displayed in the “Results” tab. The CPI distribution, the credible interval, and the expected value are displayed by pressing the “Graph” button. The report can be exported into a .csv file by pressing the “Report” button. All simulated CPI values can be exported as well into a.csv file by pressing the “Save all simulated CPI” button.

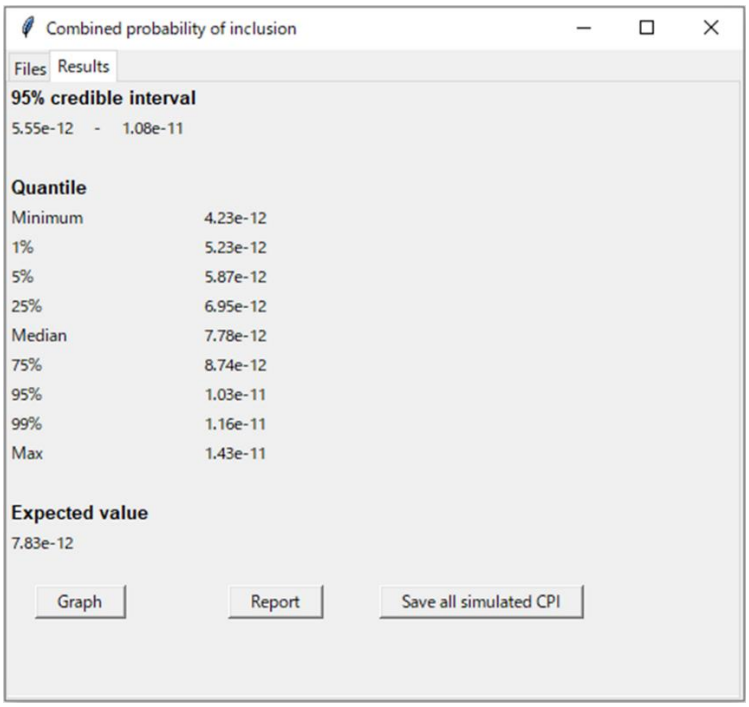

Fig. 14. The "Results" tab of "Combined probability of inclusion".

2.5. Estimate the credible interval of the likelihood ratio for mixture interpretation

To estimate the credible interval of the likelihood ratio (LR) for mixture interpretation by a binary model, press the "Likelihood ratio for mixture interpretation" button after the main window is opened as shown in Fig. 2. Then a new window is opened as shown in Fig. 15.

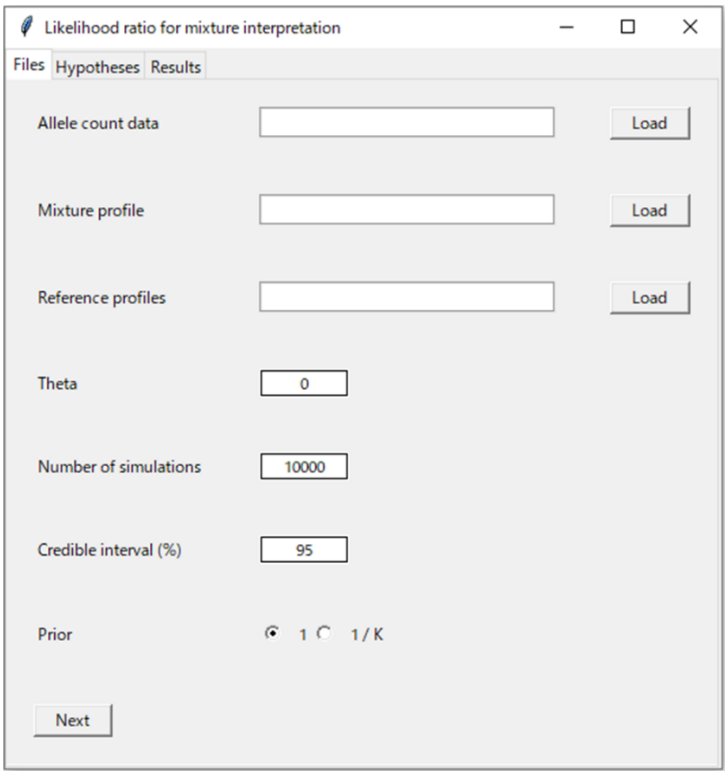

Fig. 15. The "Files" tab of "Likelihood ratio for mixture interpretation".

The user can input a .csv file of the observed number of alleles in a population database (Fig. 4), that of a mixture profile (Fig. 13), and that of reference profiles (Fig. 16) from the "Load" button.

	A	B	C	D
1	Sample Name	Marker	Allele 1	Allele 2
2	Victim	D3S1358	17	18
3	Victim	vWA	17	18
4	Victim	D16S539	9	12
5	Victim	CSF1PO	11	13
6	Victim	TPOX	8	11
7	Victim	D8S1179	14	15
8	Victim	D21S11	29	30
9	Victim	D18S51	15	22
10	Victim	D2S441	9.1	11
11	Victim	D19S433	15.2	16.2
12	Victim	TH01	7	7
13	Victim	FGA	19	21
14	Victim	D22S1045	11	16
15	Victim	D5S818	10	10
16	Victim	D13S317	9	10
17	Victim	D7S820	9	12
18	Victim	SE33	21	25.2
19	Victim	D10S1248	14	15
20	Victim	D1S1656	14	16
21	Victim	D12S391	18	22
22	Victim	D2S1338	18	24
23	Suspect	D3S1358	15	16
24	Suspect	vWA	16	17
25	Suspect	D16S539	10	11
26	Suspect	CSF1PO	10	12
27	Suspect	TPOX	8	8
28	Suspect	D8S1179	10	14
29	Suspect	D21S11	29	32.2
30	Suspect	D18S51	12	13
31	Suspect	D2S441	11	11
32	Suspect	D19S433	13	15
33	Suspect	TH01	9	9.3
34	Suspect	FGA	21	25.2
35	Suspect	D22S1045	16	17
36	Suspect	D5S818	10	11
37	Suspect	D13S317	11	12
38	Suspect	D7S820	10	10
39	Suspect	SE33	17	28.2
40	Suspect	D10S1248	13	16
41	Suspect	D1S1656	14	15
42	Suspect	D12S391	18	19
43	Suspect	D2S1338	18	18

Fig. 16. An example of the reference profiles.

The user then set a theta correction to consider subpopulation effect, the number of random sampling of allele frequencies, the range of the credible interval (%), and parameters of the prior Dirichlet distribution. After imputing three required files and setting each parameter, press the “Next” button. Then the “Hypotheses” tab is opened as shown in Fig. 17.

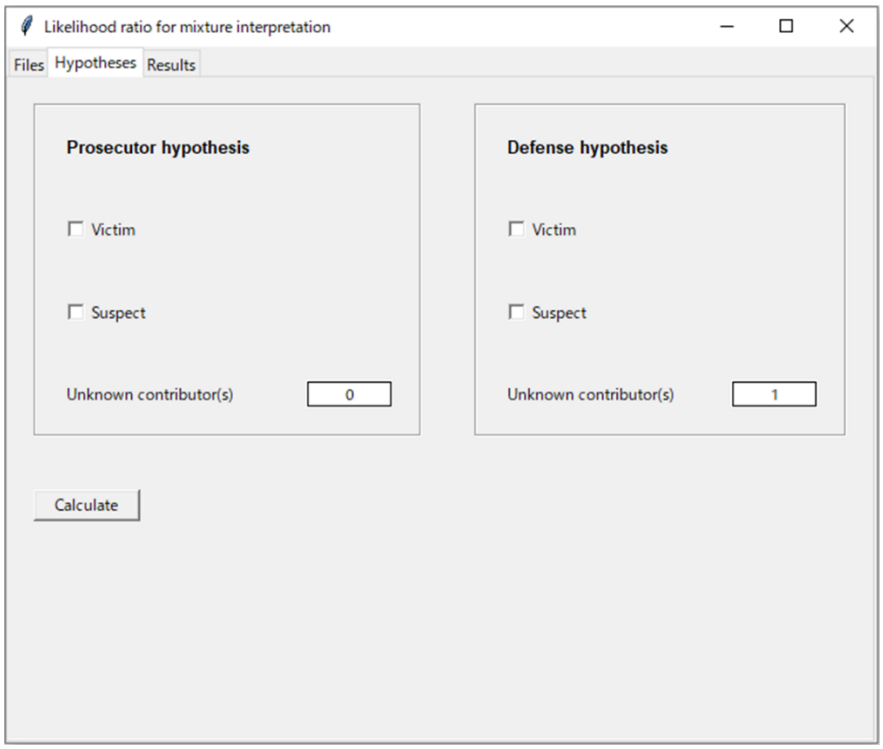

Fig. 17. The “Hypotheses” tab of “Likelihood ratio for mixture interpretation”.

The user can calculate the LR value by setting both prosecutor and defense hypotheses. Check the individuals, to include them as contributors in each hypothesis, and set the number of unknown contributor(s). After setting each hypothesis, press the “Calculate” button, and the “Results” tab is opened as shown in Fig. 18.

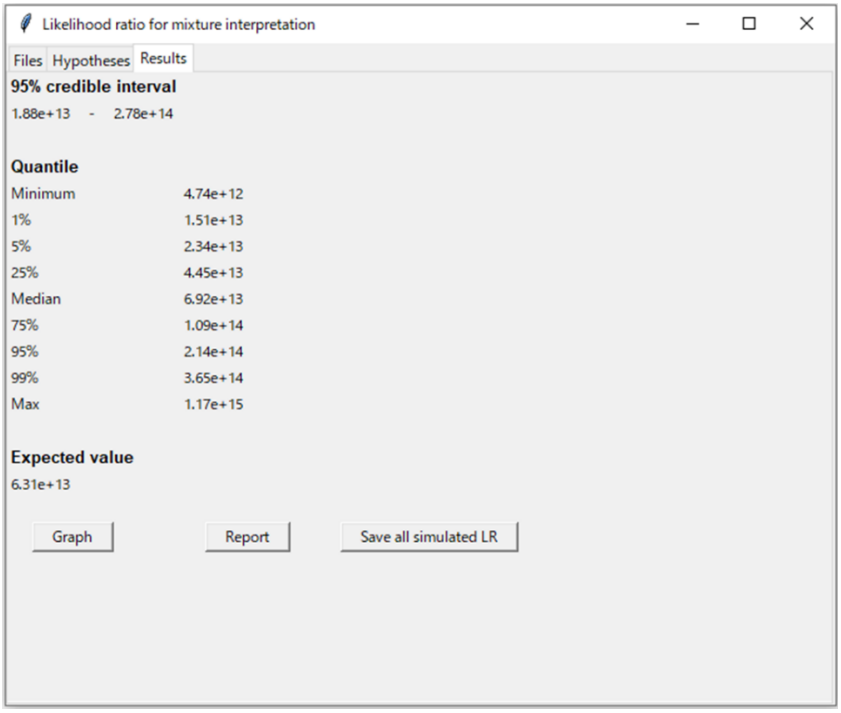

Fig. 18. The "Results" tab of "Likelihood ratio for mixture interpretation".

The credible interval of the user-selected range, quantiles of simulated LR values, and the expected LR value are displayed in the "Results" tab. The LR distribution, the credible interval, and the expected value are displayed by pressing the "Graph" button. The report can be exported into a .csv file by pressing the "Report" button. All simulated LR values can be exported as well into a.csv file by pressing the "Save all simulated LR" button.

2.6. Estimate the credible interval of the likelihood ratio for kinship analysis

To estimate the credible interval of the likelihood ratio (LR) for pairwise kinship analysis, press the "Likelihood ratio for kinship analysis" button after the main window is opened as shown in Fig. 2. Then a new window is opened as shown in Fig. 19.

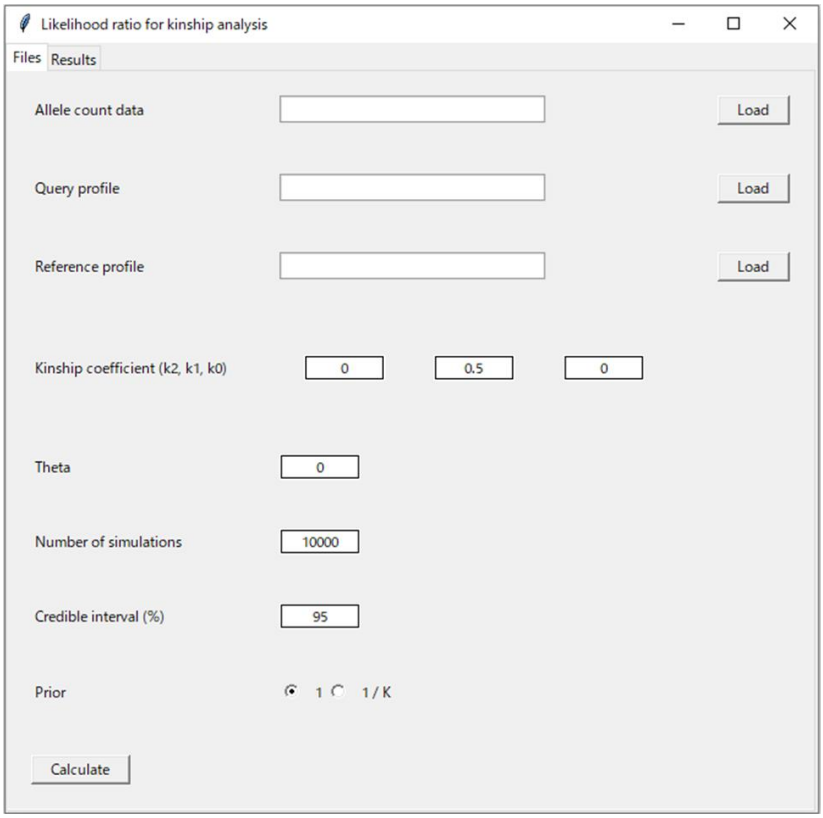

Fig. 19. The "Files" tab of "Likelihood ratio for kinship analysis".

The user can input a .csv file of the observed number of alleles in a population database (Fig. 4), that of a query profile, and that of a reference profile from the "Load" button. The query and the reference profile are the same format as the single source profile used in the analysis of RMP (Fig. 9).

The user then set values of kinship coefficient (k2, k1, and k0) to hypothesize the relationship between the query and the reference individuals. Table 1 shows values of kinship coefficient for commonly encountered relationships.

Table 1 Values of kinship coefficient for commonly encountered relationships

Relationship	k2	k1	k0
Parent-child	0	0.5	0
Siblings	0.25	0.25	0.25
Half siblings	0	0.25	0.5
Grandparent-child	0	0.25	0.5
Uncle-nephew	0	0.25	0.5
Cousins	0	0.125	0.75
Second cousins	0	0.03125	0.9375

The user can also set a theta correction to consider subpopulation effect, the number of random sampling of allele frequencies, the range of the credible interval (%), and parameters of the prior Dirichlet distribution. After imputing three required files and setting each parameter, press the “Calculate” button. Then the “Results” tab is opened as shown in Fig. 20.

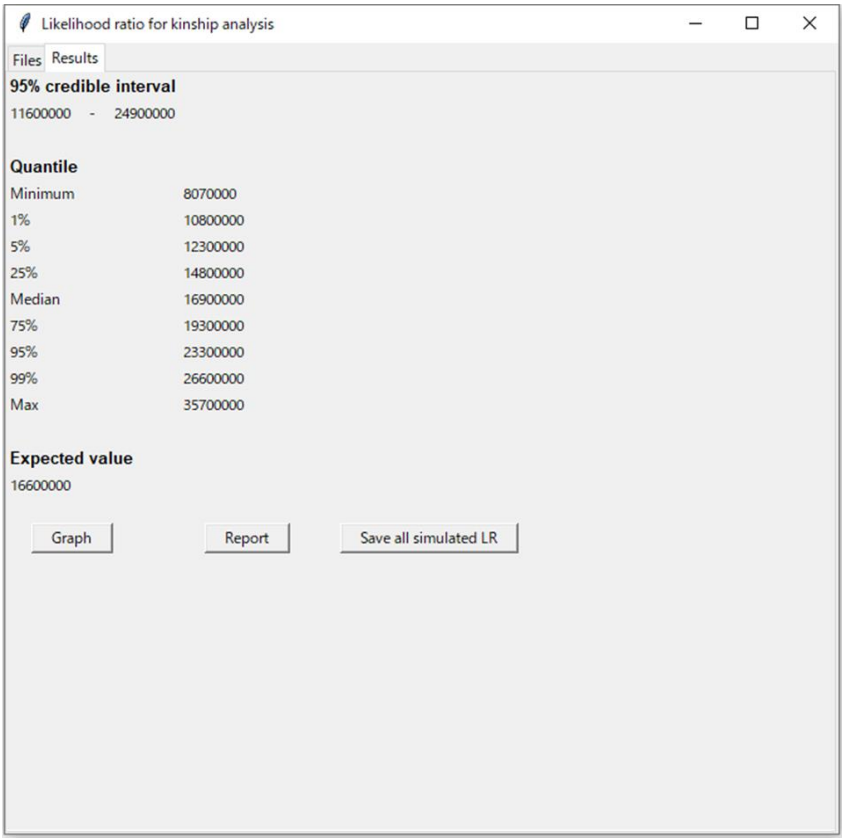

Fig. 20. The "Results" tab of "Likelihood ratio for kinship analysis".

The credible interval of the user-selected range, quantiles of simulated LR values, and the expected LR value are displayed in the "Results" tab. The LR distribution, the credible interval, and the expected value are displayed by pressing the "Graph" button. The report can be exported into a .csv file by pressing the "Report" button. All simulated LR values can be exported as well into a.csv file by pressing the "Save all simulated LR" button.