Kongoh version 1.0.1 User Manual

23rd June 2017

Sho Manabe

Department of Forensic Medicine, Kyoto University Graduate School of Medicine

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

Contents

1. What is *Kongoh*?

Kongoh (named after the Japanese word “mixture”) is an open-source software for DNA evidence interpretation based on a quantitative continuous model. *Kongoh* performs a Monte Carlo simulation on the basis of probability distributions of the biological parameters determined by means of empirical data, and then the peak heights generated by the simulation are approximated by gamma distributions. The software is a graphical user interface written in R language and the source code is freely available at GitHub (https://github.com/manabe0322/Kongoh/releases).

The profile typed by AmpFℓSTR® Identifiler® Plus PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA) can be interpreted using *Kongoh* in current version. The Identifiler Plus system is run for 28 amplification cycles. PCR products are analyzed using an ABI 3130xl Genetic Analyzer (Thermo Fisher Scientific) with no enhancements.

In *Kongoh*, there is no requirement to designate the peak located at the position of the -1 backward stutter peak as an allele or stutter because the derivation of the peak in the stutter position can be determined probabilistically. Thus, we can remove stutter filters of all loci. *Kongoh* also considers allelic drop-out, which is the event of peaks under the analytical threshold. Drop-in is not considered in *Kongoh*, but spontaneous drop-in peaks could be explained by additional unknown contributors.

Likelihood ratios are calculated by the ratio of maximum likelihoods in prosecution and defense hypotheses. Likelihoods of 1–4 persons’ contributions are automatically calculated; therefore, the number of contributors does not need to be determined manually prior to analysis.

2. Preparation

Before you start, make sure you have installed the R software and some packages properly. The R software is available from the R Development Core Team website (http://www.R-project.org). After the R software is installed, download the following packages required for running the *Kongoh* program.

Required packages: tcltk, tcltk2, gtools, MASS, truncnorm, and snow.

The *Kongoh* program is freely available at GitHub (https://github.com/manabe0322/Kongoh/releases). The file named “Kongoh v1.0.1.RData” is the software program of *Kongoh*.

3. Tutorial

3.1. Getting started

After installation, click on an icon named “Kongoh v1.0.1.RData” to start an R session. *Kongoh* is launched by the following command:

```
Kongoh()
```

If you have already installed all packages used in *Kongoh*, the “Files” page is opened as shown in Fig. 1.

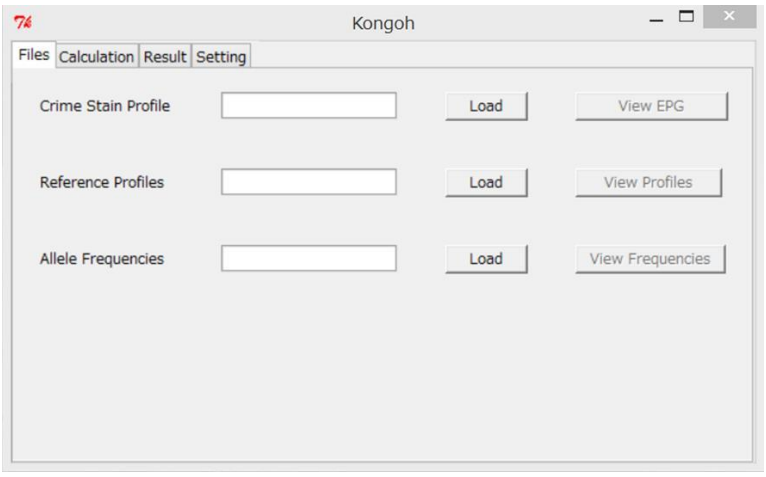

Fig. 1. The “Files” page where the user can import a crime stain profile, reference profiles, and allele frequencies.

3.2. Input a crime stain profile

After the "Files" page is opened as shown in Fig. 1, press the "Load" button for the crime stain profile. The profile must be typed by the Identifiler Plus Kit and be analyzed using an ABI 3130xl Genetic Analyzer with no enhancements. You can remove stutter filters of all loci. The input file of the crime stain profile should be in the .csv format as shown in Fig. 2. This file must include the information of "Sample File", "Marker", "Allele", "Size", and "Height". This file is also exported from the GeneMapper® software. After loading the file for a crime stain profile, the electropherogram can be saved into a PDF file by pressing the "View EPG" button.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Sample File	Marker	Allele 1	Allele 2	Allele 3	Allele 4	Size 1	Size 2	Size 3	Size 4	Height 1	Height 2	Height 3	Height 4
2	Mixture	D8S1179	10	12	13	15	130.02	138.02	142.17	150.5	64	106	1423	112
3	Mixture	D21S11	29	30	31.2	32.2	202.76	206.71	212.64	216.73	52	695	41	550
4	Mixture	D7S820	10	11	12	13	270.19	274.16	278.23	282.26	465	60	133	367
5	Mixture	CSF1PO	10	11	12		318.68	322.82	326.82		466	552	87	
6	Mixture	D3S1358	15	16	17		122.55	126.63	130.59		442	1083	257	
7	Mixture	TH01	6	7	9		168.67	172.72	180.86		494	178	647	
8	Mixture	D13S317	11	12			227.45	231.43			153	1032		
9	Mixture	D16S539	9	10	12	13	267.28	271.33	279.29	283.22	161	586	46	514
10	Mixture	D2S1338	16	17	22	23	308.81	312.83	333.3	337.24	129	468	39	571
11	Mixture	D19S433	12	13	14		112.22	116.29	120.24		39	634	343	
12	Mixture	vWA	15	16	17		168.24	172.3	176.36		55	535	460	
13	Mixture	TPOX	8	9	11		228.83	232.9	240.98		522	358	62	
14	Mixture	D18S51	13	14	15	19	285.22	289.22	293.22	309.17	36	586	546	71
15	Mixture	AMEL	X	Y			105.07	110.92			535	528		
16	Mixture	D5S818	9	11	12		140.75	148.98	153.03		418	71	273	
17	Mixture	FGA	20	21	22	23	225.19	229.35	233.42	237.49	30	526	362	124

Fig. 2. The format of the crime stain profile.

3.3. Input reference profiles

You can input reference profiles by pressing the “Load” button for the reference profiles. The profiles must include 15 STR loci in Identifiler system. The input file of the reference profiles should be in the .csv format as shown in Fig. 3. This file must include the information of “Marker” and names of each profile. After loading the file for reference profiles, you can view the profiles by pressing the “View Profiles” button.

	A	B	C	D	E
1	Marker	Victim	Victim	Suspect	Suspect
2	D8S1179	13	13	10	15
3	D21S11	30	32.2	30	30
4	D7S820	10	13	11	12
5	CSF1PO	10	11	11	12
6	D3S1358	16	16	15	17
7	TH01	6	9	7	7
8	D13S317	12	12	11	12
9	D16S539	10	13	9	10
10	D2S1338	17	23	16	23
11	D19S433	13	14	13	13
12	vWA	16	17	16	17
13	TPOX	8	9	8	11
14	D18S51	14	15	14	19
15	AMEL	X	Y	X	Y
16	D5S818	9	12	9	11
17	FGA	21	22	23	23

Fig. 3. The format of the reference profiles.

3.4. Input allele frequencies

You can input allele frequencies by pressing the “Load” button for the allele frequencies of 15 loci in Identifiler system. The input file of the allele frequencies should be in the .csv format as shown in Fig. 4. This file must include the information of “Allele” and names of each locus. After loading the file for allele frequencies, you can view the frequencies by pressing the “View Frequencies” button. To calculate the frequencies of rare alleles which are not observed in the population database, you should set the minimum allele frequency in the “Setting” tab as described in section 3.8.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Allele	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
2	5						0.001852									
3	6						0.223621									
4	7	0.001852		0.003332	0.010367		0.266938	0.001852	0.001852				0.001852		0.002962	
5	8			0.12699	0.001852		0.066642	0.267308	0.002221				0.454646		0.007034	
6	9	0.001852		0.045909	0.04702		0.398741	0.129211	0.353943				0.114402		0.086264	
7	9.1			0.001852												
8	9.2										0.001852					
9	9.3						0.035172									
10	10	0.132914		0.219178	0.223621		0.009256	0.115143	0.196964				0.363199	0.002592	0.201037	
11	10.1			0.001852												
12	10.2										0.001852					
13	10.3			0.001852												
14	11	0.109219		0.328767	0.208071			0.221399	0.187338		0.004073		0.363199	0.004813	0.292484	
15	11.2										0.001852					
16	12	0.122917		0.235098	0.418734	0.002221		0.202518	0.178823		0.040726		0.035913	0.04813	0.235468	
17	12.2										0.005553					
18	13	0.225102		0.035172	0.069234	0.001852		0.051462	0.072936		0.287671	0.001852	0.001852	0.199556	0.166975	
19	13.2										0.030359					
20	14	0.205109		0.006294	0.018141	0.029248		0.013328	0.008515		0.34987	0.194372	0.001852	0.22251	0.009256	
21	14.2										0.088486					
22	15	0.134765		0.001852	0.005553	0.39615		0.001852	0.001852		0.051092	0.027027		0.168456	0.001852	
23	15.2										0.115143					
24	16	0.064421			0.001852	0.306553				0.008886	0.005553	0.184376		0.125879		
25	16.2										0.019622					
26	17	0.006664				0.199926				0.097742		0.282858		0.081822		0.003702
27	17.1													0.001852		
28	17.2										0.003332					
29	18	0.001852				0.06368				0.1592	0.001852	0.225842		0.048501		0.021844
30	19					0.003332				0.209182		0.074047		0.036653		0.067382
31	20									0.105887		0.010367		0.022214		0.089226
32	21					0.001852				0.01518		0.002962		0.01592		0.131063
33	22									0.050722		0.001852		0.013328		0.201777
34	22.2															0.002221
35	23									0.146983				0.007775		0.205479
36	23.2															0.005183
37	24									0.106108				0.004073		0.157349
38	24.2															0.001852
39	25									0.061829				0.001852		0.073676
40	25.2															0.002221
41	26									0.029248				0.001852		0.03221
42	27		0.001852							0.008886				0.001852		0.008145
43	28		0.042577							0.002592						0.002592
44	28.2		0.005183													
45	29		0.246946													

Fig. 4. The format of the allele frequencies.

3.5. Weight Calculation

After loading three required files, press the “Calculation” tab. The “Calculation” page is opened as shown in Fig. 5.

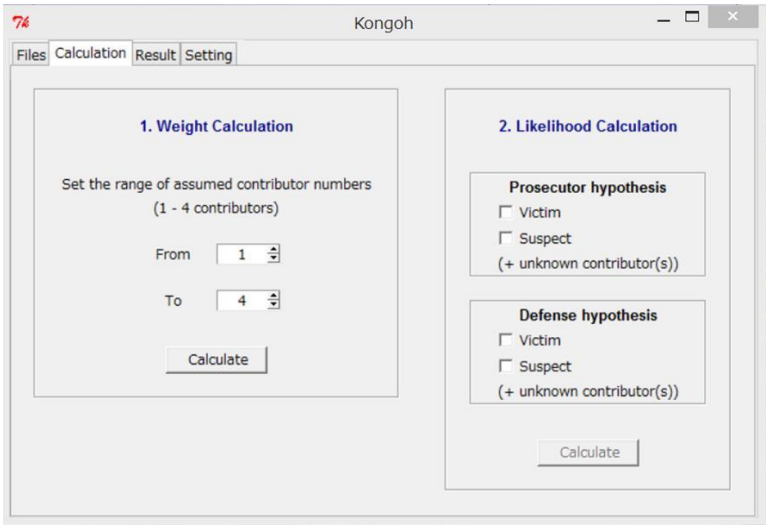

Fig. 5. The “Calculation” page where the user can calculate the weights and the likelihoods.

The weight values of each genotype combination in 1–4 contributors are automatically calculated by pressing the “Calculate” button. The user can change the range of the assumed numbers of contributors by choosing the number of contributors from spin boxes.

3.6. Likelihood Calculation

After finishing the weight calculation, you can calculate likelihood values by setting both prosecutor (H_p) and defense (H_d) hypotheses. Check the individuals to include them as contributors in each hypothesis. Fig. 6 shows an example of setting the hypotheses; H_p : victim + suspect (+ unknown contributors), and H_d : victim (+ unknown contributors). You do not need to select the number of unknown contributors because *Kongoh* automatically calculate likelihoods of all assumed numbers set in the weight calculation. After setting each hypothesis, press the "Calculate" button to calculate likelihood values.

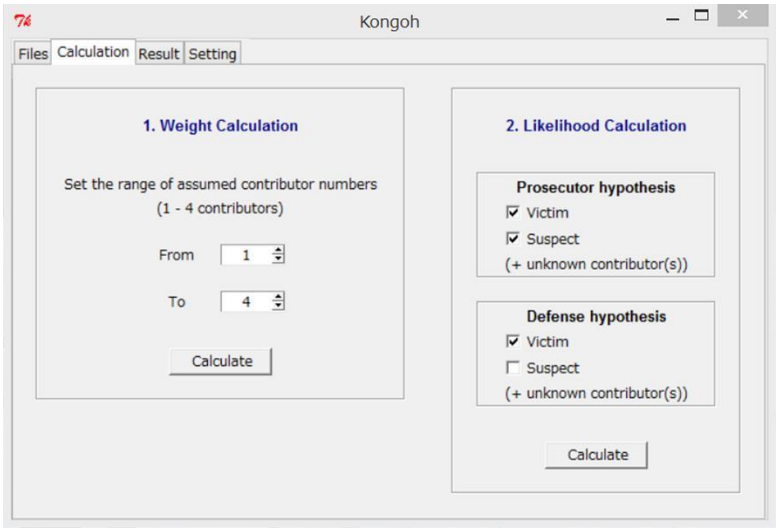

Fig. 6. An example of setting the hypotheses.

3.7. Results

After finishing the likelihood calculation, the "Result" page is automatically opened as shown in Fig. 7. You can view brief overview of the results (i.e., likelihoods and estimated parameters in H_p and H_d, likelihood ratios, and the ratio of maximum likelihood in H_p and that in H_d). You can export the report into a .csv file by pressing the "Report" button. The report includes detailed information such as set parameter values, hypotheses, likelihoods in each locus, and estimated mixture ratios including information of each contributor.

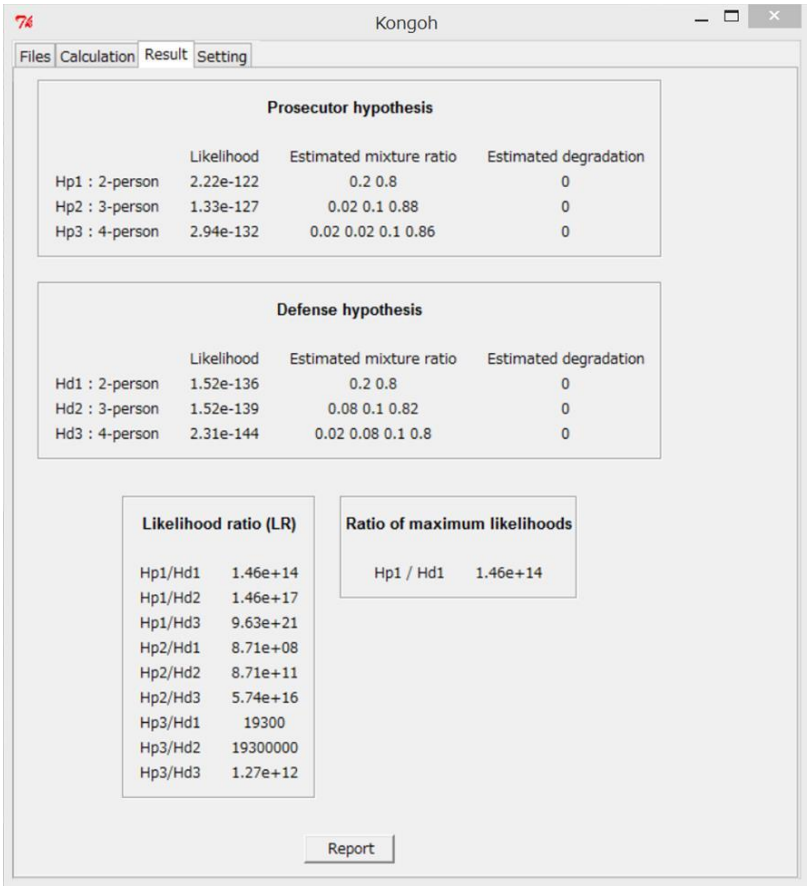

Fig. 7. An example of the "Result" page.

3.8. Setting

If you want to change some parameters, press the "Setting" tab and change each parameter value before calculating weight values. The "Setting" page is shown in Fig. 8.

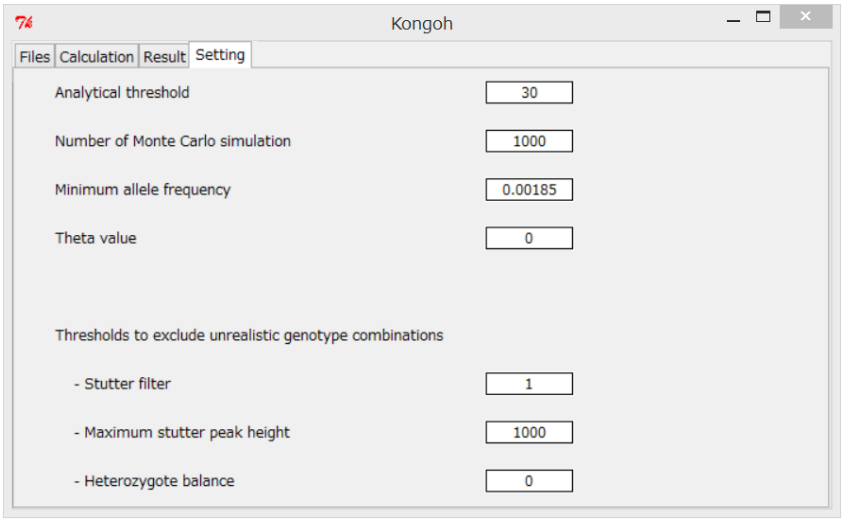

Fig. 8. The "Setting" page. Default parameter values are already entered in each entry box.

The parameters are the analytical threshold, the number of Monte Carlo simulation, and the minimum allele frequency in current version. You also set the following three thresholds to exclude unrealistic genotype combinations.

(i) If a stutter ratio is greater than the set value, it is not possible that the stutter position's peak is derived from only the stutter product,

(ii) If a stutter position's peak is greater than the set value, it is not possible that the peak is derived from only the stutter product, and

(iii) If a heterozygote balance is less than the set value, the two peaks are not derived from only a single contributor.