***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*? 3](#_Toc469994627)

[2. Preparation 4](#_Toc469994628)

[3. Tutorial 5](#_Toc469994629)

[*3.1.* *Getting started* 5](#_Toc469994630)

[*3.2.* *Input a crime stain profile* 6](#_Toc469994631)

[*3.3.* *Input reference profiles* 7](#_Toc469994632)

[*3.4.* *Input allele frequencies* 8](#_Toc469994633)

[*3.5.* *Weight Calculation* 9](#_Toc469994634)

[*3.6.* *Likelihood Calculation* 10](#_Toc469994635)

[*3.7.* *Results* 11](#_Toc469994636)

[*3.8.* *Setting* 12](#_Toc469994637)

**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).

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). The file named “Kongoh v1.0.1.RData” is the software program of *Kongoh*.

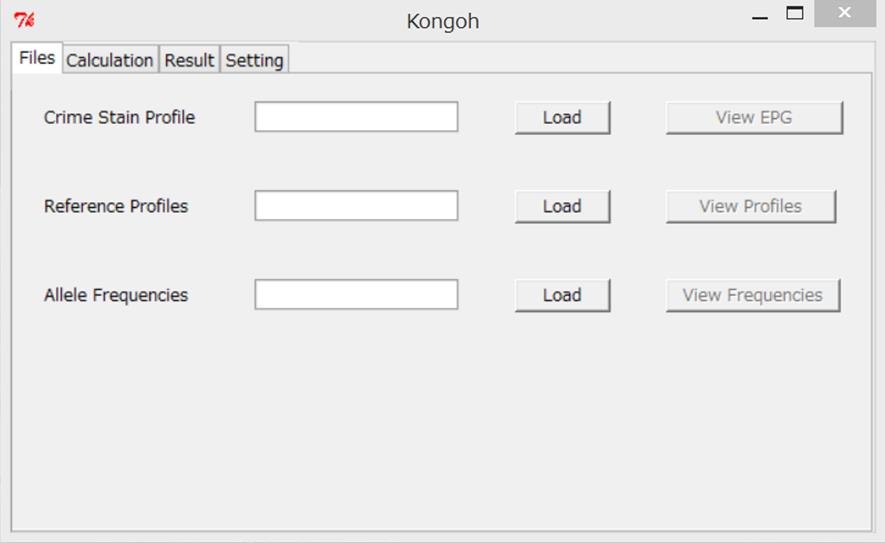
**3. Tutorial**

*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()

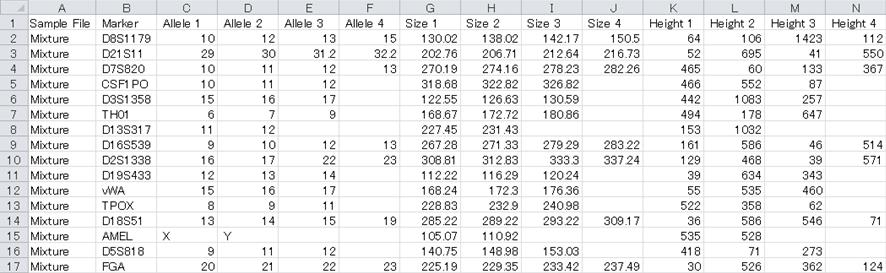
If you have already installed all packages used in *Kongoh*, the “Files” page is opened as shown in Fig. 1.



**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*

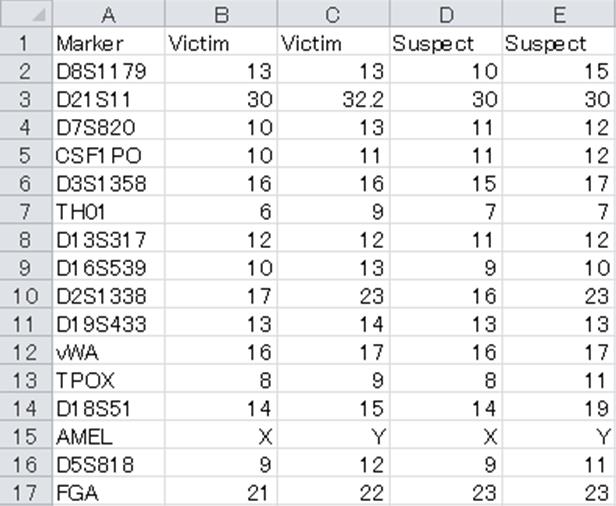
After the “Files” page is opened as shown in Fig. 1, press the “Load” button for the crime stain profile. The profile must be typed by the Identifiler Plus Kit and be analyzed using an ABI 3130xl Genetic Analyzer with no enhancements. You can remove stutter filters of all loci. The input file of the 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.



**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.



**Fig. 3.** The format of the reference profiles.

*3.4. Input allele frequencies*

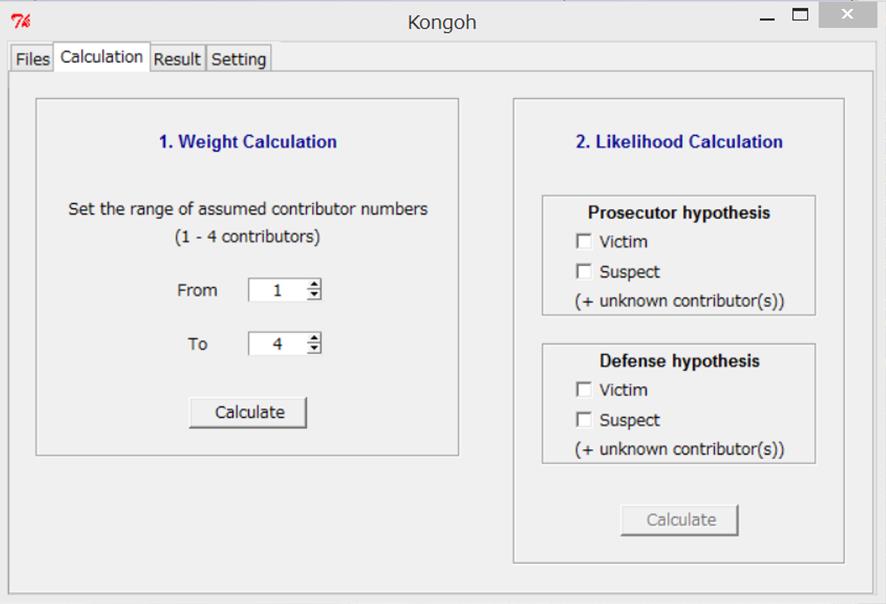
You can input allele frequencies by pressing the “Load” button for the allele frequencies of 15 loci in Identifiler system. The input file of the allele frequencies should be in the .csv format as shown in Fig. 4. This file must include the information of “Allele” and names of each locus. After loading 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.



**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.

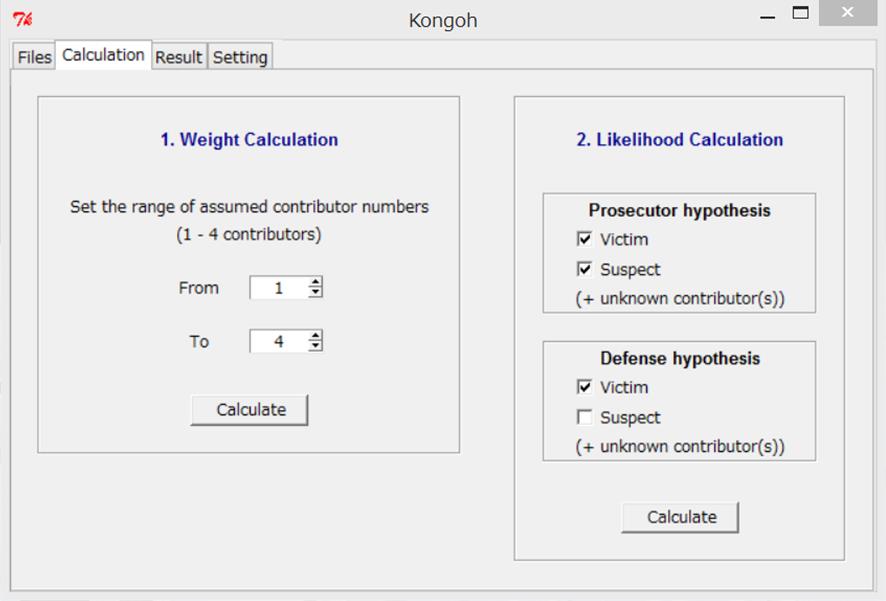


**Fig. 5.** The “Calculation” page where the user can calculate the weights and the likelihoods.

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*

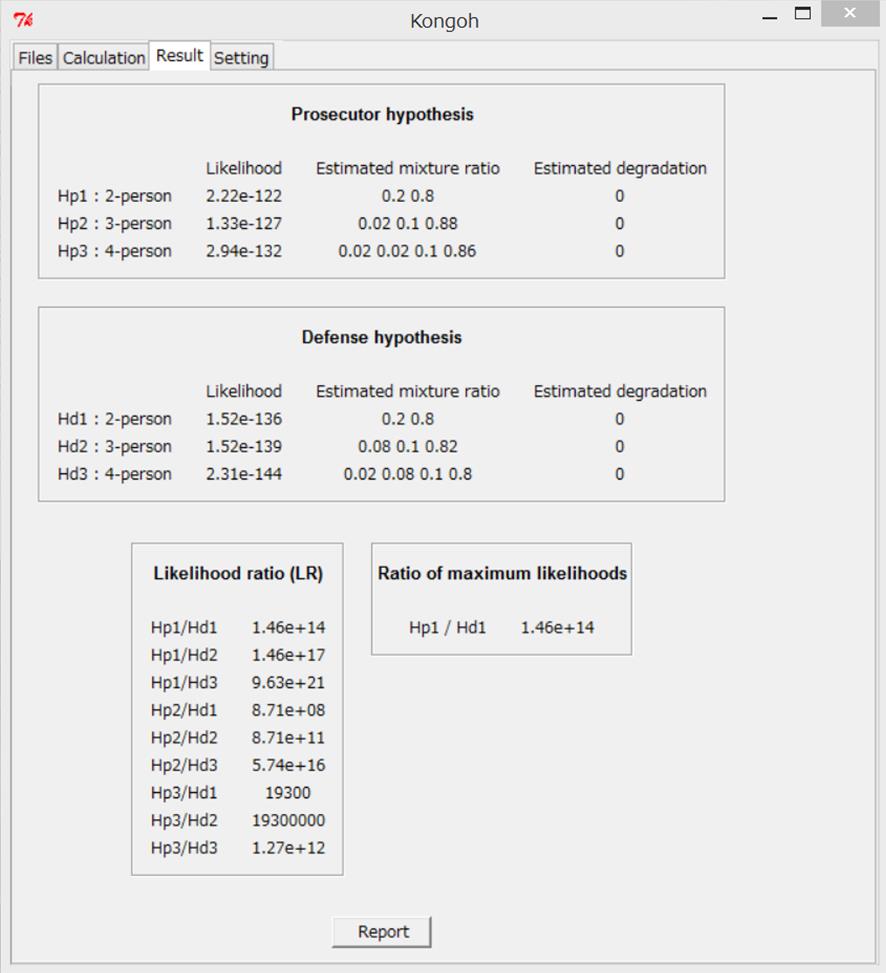
After finishing the weight calculation, you can calculate likelihood values by setting both prosecutor () and defense () hypotheses. Check the individuals to include them as contributors in each hypothesis. Fig. 6 shows an example of setting the hypotheses; : victim + suspect (+ unknown contributors), and : 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*

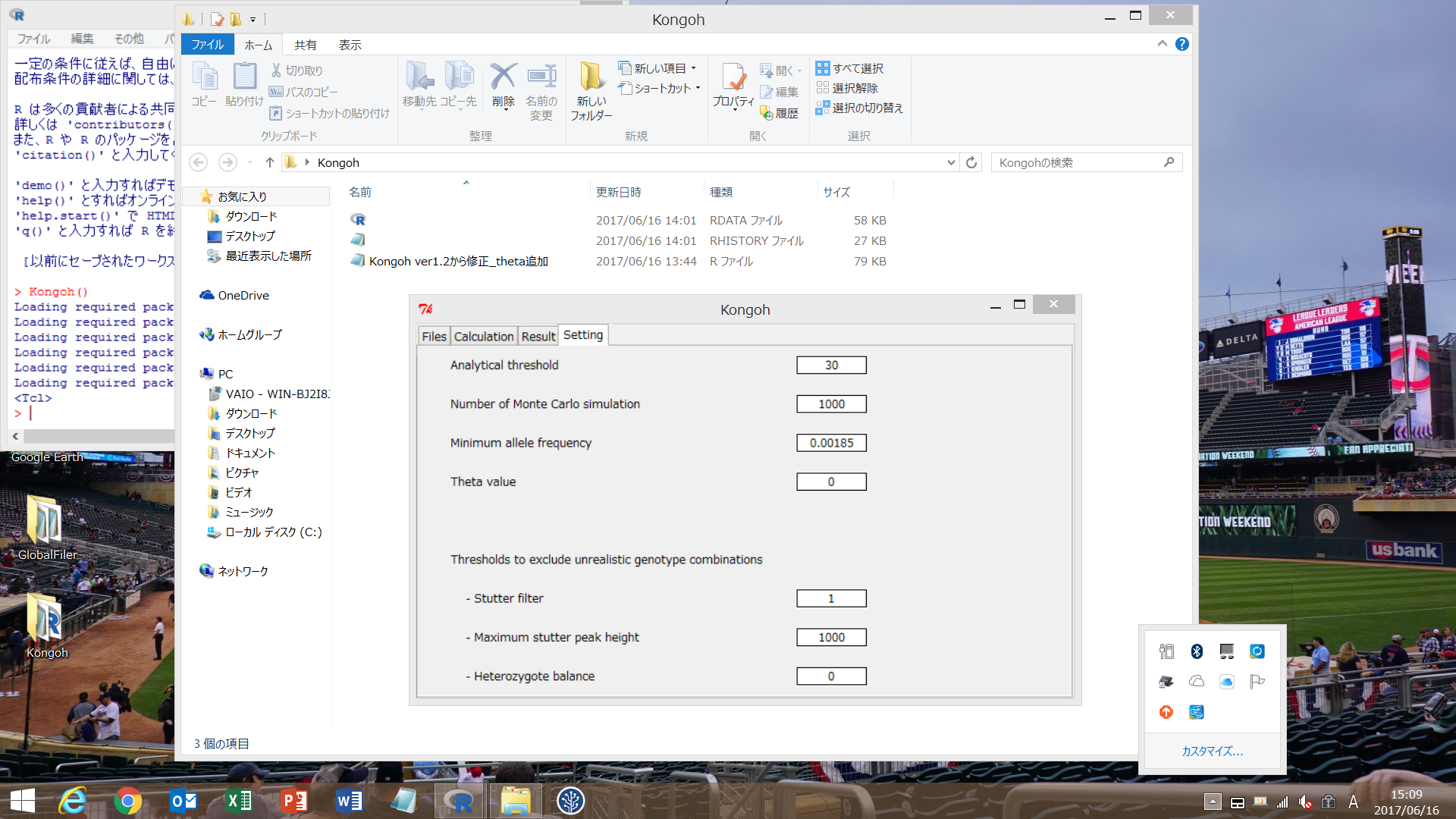
After finishing the likelihood calculation, the “Result” page is automatically opened as shown in Fig. 7. You can view brief overview of the results (i.e., likelihoods and estimated parameters in and , likelihood ratios, and the ratio of maximum likelihood in and that in ). You can export the report 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.