

RelSearch version 1.0.0

User Manual

Sho Manabe

2024-10-09

Contents

1	Initial setup	3
2	Quick guide	4
3	Files	10
3.1	STR: Victim database	10
3.2	STR: Reference database	11
3.3	STR: Allele frequencies	11
3.4	Y-STR: Victim database	12
3.5	Y-STR: Reference database	12
3.6	mtDNA: Victim database	13
3.7	mtDNA: Reference database	14
4	Setting	15
4.1	Criteria	15
4.2	Relationship	17
4.2.1	Edit	18
4.2.2	Add	19
4.2.3	Delete	20
4.2.4	Reset	20
4.2.5	Family tree	20
4.3	Mutation rate	21
4.3.1	Edit	21
4.3.2	Add	21
4.3.3	Delete	21
4.3.4	Reset	21
4.4	Data storage conditions	21
5	Result	23
5.1	Summary	23
5.2	Selected data in detail	24
5.3	All candidates	24
5.4	Analysis conditions	24

6	Other functions	26
6.1	Project	26
6.1.1	New project	26
6.1.2	Load project	26
6.1.3	Save project	26
6.2	View database	26
6.3	Example files	26
7	Computation time and limits	27
8	Computational principle	29
8.1	STR	29
8.1.1	Likelihood ratio	29
8.1.2	Allele probabilities	30
8.1.3	Allelic drop-out	30
8.1.4	Mutation	31
8.2	Y-STR	32
8.3	mtDNA	33
	Reference	34

Chapter 1

Initial setup

1. Ensure that R ($\geq 4.4.0$) is installed. It is available from the [R Development Core Team website](#).
2. Begin an R session.
3. Execute the following command in R to install required packages.

```
install.packages('https://github.com/manabe0322/RelSearch/releases/download/  
  ↪ v1.0.0/RelSearch_1.0.0.zip', repos = NULL, type = 'win.binary')  
install.packages(c("shiny", "data.table", "dplyr", "DT", "magrittr",  
  ↪ "pedtools", "Rcpp", "ribd", "shinyFeedback", "shinyjs", "shinythemes",  
  ↪ "waiter"))
```

Chapter 2

Quick guide

1. Execute the following commands in R to launch the software interface.

```
library(RelSearch)  
RelSearch()
```

Then, the software interface will open automatically.

2. Load files from each “**Browse...**” button.

The screenshot displays the RelSearch web interface (version 0.11.0) with a blue navigation bar containing links: Load, Result, Project, Database, Settings, Example files, and Manual. The main content area is divided into three columns for STR, Y-STR, and mtDNA analysis. Each column has three sections: Victim database, Reference database, and Allele frequencies (only for STR). Each section contains a 'Browse...' button (highlighted with an orange box) and a 'No file selected' status. An 'Analyze' button is located at the bottom left of the interface.

Note

- RelSearch can analyze data for each marker type separately as well as for all marker types together.
- Victim, reference, and allele frequency databases are required to analyze autosomal STR profiles.
- Victim and reference databases are required to analyze Y-STR profiles.
- Victim and reference databases are required to analyze mtDNA profiles.
- See section [Files](#) for information on each file.

3. Click the **"Analysis"** button.

relsearch ver. 0.11.0 Load Result Project ▾ Database ▾ Settings ▾ Example files Manual

STR

Victim database

Browse... str_victim_example.csv
Upload complete

Reference database

Browse... str_ref_example.csv
Upload complete

Allele frequencies

Browse... str_af_example.csv
Upload complete

Y-STR

Victim database

Browse... y_victim_example.csv
Upload complete

Reference database

Browse... y_ref_example.csv
Upload complete

mtDNA

Victim database

Browse... mt_victim_example.csv
Upload complete

Reference database

Browse... mt_ref_example.csv
Upload complete

Analyze

4. After finishing the analysis, the result window will open automatically.

research ver. 0.11.0 Load **Result** Project ▾ Database ▾ Settings ▾ Example files Manual

Result

Summary Selected data in detail Analysis conditions

Display setting

Default display

Identified pairs

Multiple candidates

Not support lineage

Minimum LR displayed
100

Apply

Download

Show 10 ▾ entries Search:

Victim	Reference	Assumed relationship	LR	Estimated relationship	Paternal lineage	Maternal lineage
All	All	All	All	All	All	All
V7	R7	Father_Son	9.23e+8	Father_Son	Support	Support
V8	R8	Father_Son	6.48e+8	Father_Son	Support	Not support
V14	R14	Brother_Brother	2.85e+7	Brother_Brother	Support	Support
V9	R9	Father_Son	1.34e+7	Father_Son	Support	Not support
V11	R11	Mother_Daughter	1.25e+7			Not support
V16	R16	Brother_Sister	2.01e+6	Brother_Sister		Support
V15	R15	Brother_Sister	1.57e+6	Brother_Sister		Support
V7	R30	Brother_Brother	1.02e+5	Brother_Brother	Support	Support
V10	R10	Mother_Daughter	4.49e+4	Mother_Daughter		Support
V12	R12	Mother_Daughter	3.09e+3	Mother_Daughter		Support

Showing 1 to 10 of 12 entries Previous 1 2 Next

Note

- See section **Result** for navigation of the result window.

7

5. Select **Project** ► **Save project**.

research ver. 0.11.0 Load Result **Project** ▼ Database ▼ Settings ▼ Example files Manual

New project
Load project
Save project ←

Result

Summary Selected data in detail Analysis conditions

Display setting

Default display

Identified pairs

Multiple candidates

Not support lineage

Minimum LR displayed

100

Apply

Download

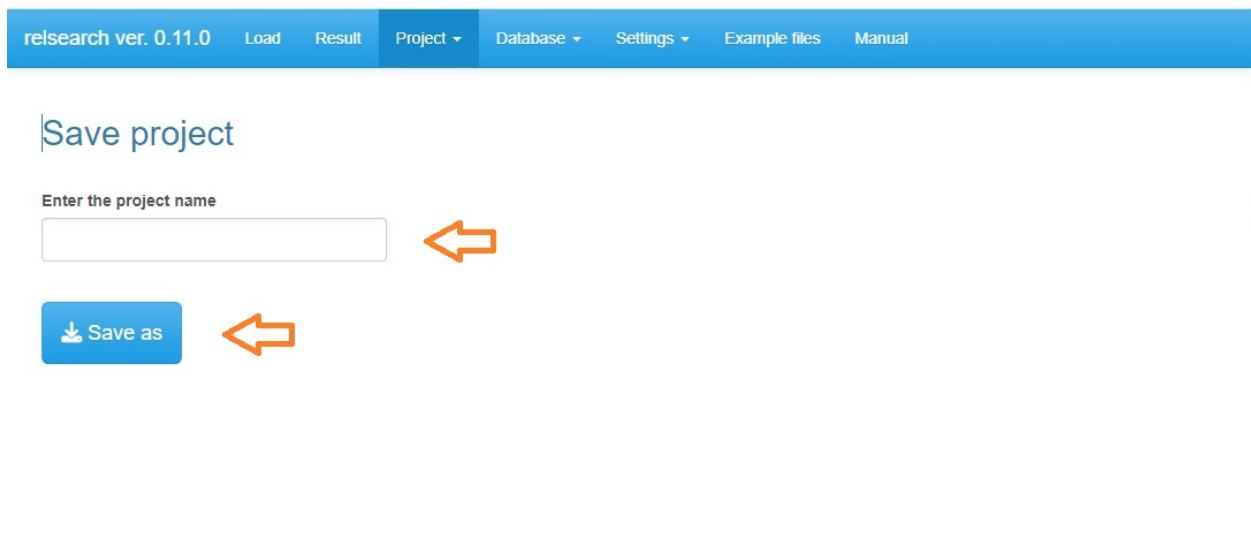
Show 10 entries Search:

Victim	Reference	Assumed relationship	LR	Estimated relationship	Paternal lineage	Maternal lineage
All	All	All	All	All	All	All
V7	R7	Father_Son	9.23e+8	Father_Son	Support	Support
V8	R8	Father_Son	6.48e+8	Father_Son	Support	Not support
V14	R14	Brother_Brother	2.85e+7	Brother_Brother	Support	Support
V9	R9	Father_Son	1.34e+7	Father_Son	Support	Not support
V11	R11	Mother_Daughter	1.25e+7			Not support
V16	R16	Brother_Sister	2.01e+6	Brother_Sister		Support
V15	R15	Brother_Sister	1.57e+6	Brother_Sister		Support
V7	R30	Brother_Brother	1.02e+5	Brother_Brother	Support	Support
V10	R10	Mother_Daughter	4.49e+4	Mother_Daughter		Support
V12	R12	Mother_Daughter	3.09e+3	Mother_Daughter		Support

Showing 1 to 10 of 12 entries Previous 1 2 Next

127.0.0.1:4409/#tab-1131-3

6. Enter the project name and click the **“Save as”** button.



research ver. 0.11.0 Load Result **Project** ▾ Database ▾ Settings ▾ Example files Manual

Save project

Enter the project name

←

Save as ←

Note

- The saved project can be loaded from **Project ► Load project**.
- How to handle projects is described in section **Other functions**.

Chapter 3

Files

3.1 STR: Victim database

SampleName	D3S1358	D3S1358	vWA	vWA	D16S539	D16S539	CSF1PO	CSF1PO
Victim1	17	17	17		11		12	
Victim2	17		18	18			10	11
Victim3	15	17	17	18	10	10	10	12
Victim4	15	18	15	18	9	12	11	12
Victim5	15	15	18		9	9	10	

Note

- File type: .csv
- This file requires the column "SampleName" and columns for each marker (two columns in each).
- The marker with two empty cells (e.g., D16S539 of the sample "Victim2" in the above table) is ignored when calculating the likelihood ratio.
- The marker with one empty cell (e.g., vWA of the sample "Victim1" in the above table) can be regarded as both homozygote (i.e., no drop-out) and heterozygote with drop-out of one allele when calculating the likelihood ratio.

3.2 STR: Reference database

SampleName	Family	Relationship	D3S1358	D3S1358	vWA	vWA	D16S539	D16S539
Reference1	Family 1	Father_Son	17	17	14	17	10	10
Reference2	Family 2	Brother_Brother	16	17	17	17	10	12
Reference3	Family 3	Brother_Brother	16	17	14	18	9	9
Reference3	Family 3	Father_Son	16	17	14	18	9	9
Reference4	Family 4	Brother_Brother	15	16	17		9	12
Reference5	Family 4	Uncle_Nephew	15	15	14	16		

Note

- File type: .csv
- This file requires the columns "SampleName", "Family", "Relationship", and columns for each marker (two columns in each).
- The names of the relationship need to be pre-defined in **Settings ► Relationships** (see section **Setting**).
- If there are multiple missing family members for a reference, add rows for each relationship of the members (e.g., Brother_Brother and Father_Son of the sample "Reference3" in the above table).
- The marker with two empty cells (e.g., D16S539 of the sample "Reference5" in the above table) is ignored when calculating the likelihood ratio.
- The marker with one empty cell (e.g., vWA of the sample "Reference4" in the above table) can be regarded as both homozygote (i.e., no drop-out) and heterozygote with drop-out of one allele when calculating the likelihood ratio.

3.3 STR: Allele frequencies

Allele	D3S1358	vWA	D16S539	CSF1PO	TPOX
10			602	648	100
11			562	620	1072
12	6		517	1267	114
13	3	1	209	208	3
14	79	586	26	52	2
15	1192	78	3	14	

Note

- File type: .csv

- This file requires the columns "Allele" and columns for each marker (one column in each).
- Allele counts in the population database are entered in the columns for each marker.

3.4 Y-STR: Victim database

SampleName	DYS392	DYS518	DYS570	DYS437	DYS385
Victim1	11	37	17	14	13,17
Victim2	13	38	19	14	10,20
Victim3	11	37	17	14	14
Victim4	11	37	16	14	13
Victim5			18		13,17

Note

- File type: .csv
- This file requires the column "SampleName" and columns for each marker (one column in each).
- In the marker with more than one allele, each allele must be separated by a comma without any spaces (e.g., DYS385 in the above table).
- The marker with an empty cell (e.g., DYS392 of the sample 'Victim5' in the above table) is ignored for the analysis.

3.5 Y-STR: Reference database

SampleName	Family	Relationship	DYS392	DYS518	DYS570	DYS437	DYS385
Reference1	Family 1	Father_Son	13	38	19	14	10,20
Reference2	Family 2	Brother_Brother	11	37	17	14	14,17
Reference3	Family 3	Brother_Brother	11	38	16	14	11
Reference3	Family 3	Father_Son	11	38	16	14	11
Reference4	Family 4	Brother_Brother	11	38	19	14	13,15
Reference5	Family 4	Uncle_Nephew	14		18	14	13,18

Note

- File type: .csv

- This file requires the column "SampleName", "Family", "Relationship", and columns for each marker (one column in each).
- The names of the relationship need to be pre-defined in **Settings ► Relationships** (see section **Setting**).
- If there are multiple missing family members for a reference, add rows for each relationship of the members (e.g., Brother_Brother and Father_Son of the sample "Reference3" in the above table).
- In the marker with more than one allele, each allele must be separated by a comma without any spaces (e.g., DYS385 in the above table).
- The marker with an empty cell (e.g., DYS518 of the sample 'Reference5' in the above table) is ignored for the analysis.

3.6 mtDNA: Victim database

SampleName	Range	Haplotype
Victim1	73-340 16024-16365	16183C 16189C 16217C 16311C 73G 263G 309.1C 315.1C
Victim2	73-340 16024-16365	16093C 16114A 16223T 16362C 73G 263G 309.1C 315.1C
Victim3	73-167 240-340 16024-16365	16095T 16189C 16223T 16265C 16274A 16362C 73G 263G
Victim4	73-265 16024-16284	16223T 73G 152C 263G
Victim5		

Note

- File type: .csv
- This file requires the column "SampleName", "Range", and "Haplotype".
- In the 'Range' column, the first and the last positions in each readable sequence must be written with a hyphen between these positions (e.g, 73-340). The readable sequences must be separated by a single blank space.
- In the 'Haplotype' column, the mtDNA types are expressed as the differences to a reference sequence such as the revised Cambridge Reference Sequence [1] (e.g., 73G, 315.1C, and so on). Each mtDNA type must be separated by a single blank space.
- The empty cells in the 'Range' and 'Haplotype' columns (e.g., the sample 'Victim5' in the above table) mean that there is no readable sequence.

3.7 mtDNA: Reference database

SampleName	Family	Relationship	Range	Haplotype
Reference1	Family 1	Mother_Daughter	73-340 16024-16365	16223T 73G 263G 309.1C 315.1C
Reference2	Family 2	Mother_Daughter	73-340 16024-16365	16172C 16223T 73G 150T 263G
Reference3	Family 3	Brother_Brother	73-340 16024-16365	16172C 16189C 16223T 73G 263G
Reference3	Family 3	Father_Son	73-340 16024-16365	16172C 16189C 16223T 73G 263G
Reference4	Family 4	Sister_Sister	73-340 16024-16365	16223T 16362C 73G 263G 315.1C
Reference5	Family 4	Aunt_Niece		

Note

- File type: .csv
- This file requires the column "SampleName", "Family", "Relationship", "Range", and "Haplotype".
- The names of the relationship need to be pre-defined in **Settings ► Relationships** (see section **Setting**).
- If there are multiple missing family members for a reference, add rows for each relationship of the members (e.g., Brother_Brother and Father_Son of the sample "Reference3" in the above table).
- In the 'Range' column, the first and the last positions in each readable sequence must be written with a hyphen between these positions (e.g, 73-340). The readable sequences must be separated by a single blank space.
- In the 'Haplotype' column, the mtDNA types are expressed as the differences to a reference sequence such as the revised Cambridge Reference Sequence [1] (e.g., 73G, 315.1C, and so on). Each mtDNA type must be separated by a single blank space.
- The empty cells in the 'Range' and 'Haplotype' columns (e.g., the sample 'Reference5' in the above table) mean that there is no readable sequence.

Chapter 4

Setting

4.1 Criteria

In RelSearch, there are some criteria to determine whether each victim-reference pair has an assumed relationship or not. For autosomal STR, RelSearch calculates LR values for each victim-reference pair and investigates whether these LR values are greater than an LR threshold or not. For Y-STR, RelSearch analyzes the number of mismatched loci and the total mismatched steps between victim and reference Y-STR profiles to search paternal lineages. For mtDNA, RelSearch investigates the number of mismatched nucleotides between the victim and the reference sequences to search maternal lineages.

To set criteria, select **Setting ► Criteria**. Then the criteria setting window will open automatically.

The criteria for each marker type is as follows:

STR

- Minimum LR: The minimum likelihood ratio for which the victim and reference are estimated to have the assumed relationship.

Y-STR

- Maximum number of mismatched loci: The upper limit of the number of mismatched loci between victim and reference profiles for which paternal lineage is not negated.
- Maximum total mutational steps: The upper limit of total mutational steps (i.e., sum of mutational steps of all loci) between victim and reference profiles for which paternal lineage is not negated.

mtDNA

- Maximum number of inconsistency: The upper limit of inconsistency between victim and reference profiles for which maternal lineage is not negated.

After changing the criteria, press the **Save** button to reflect the changes. Press the **Reset** button if you want to return to default values.

4.2 Relationship

The names of the relationship in each reference profile need to be pre-defined. To define the relationship names, select **Setting ► Relationships**. Then the relationship setting window will open automatically.

There are 44 pre-defined victim-reference relationships as the default. The names of these relationships are determined in the order of victim_reference (e.g., Father_Son means that a victim is the father and a reference is the son). You are free to change these names.

Information on the table is as follows.

- **Sex (Victim):** The biological sex of the victim
- **Sex (Reference):** The biological sex of the reference
- **Pr(IBM = 2):** The probability that two alleles are identity by descent (IBD).
- **Pr(IBM = 1):** The probability that one allele is identity by descent (IBD).
- **Pr(IBM = 0):** The probability that zero allele is identity by descent (IBD).
- **Paternal lineage:** Information on whether each relationship is paternal lineage or not.
- **Maternal lineage:** Information on whether each relationship is maternal lineage or not.

You can edit, add, and delete the relationships from the left sidebar.

4.2.1 Edit

You can edit the relationship names from the **Edit** button.

In the pop-up window, select a relationship from the defined relationships, and enter a new name of the selected relationship. Press the **Save** button to reflect the new name.

4.2.2 Add

You can define a new relationship from the **Add** button. The procedure is as follows.

1. Enter a relationship name.
2. Make a family tree to define the relationship between a victim and a reference. The following is a list of functions to make a family tree.
 - **Add a person:** The function to add an unknown person (UK) to the list of persons.
 - **Delete a person:** The function to delete an unknown person (UK) from the list of person.
 - **Sex:** The function to select a biological sex (male: M or female: F).
 - **Father:** The function to select a person who is the father.
 - **Mother:** The function to select a person who is the mother.
 - **Founder:** The function to select whether the person is a founder of the family tree or not.

Here is an example of defining the brother-brother relationship for the victim-reference pair.

- 2.1. Press the ****Add**** button twice to add the untyped father and mother. UK1 and UK2 are assumed to be the father and mother, respectively.
 - 2.2. Set the sex of the Victim, Ref, UK1, and UK2 to "M", "M", "M", and "F", respectively.
 - 2.3. In the Founder column, select "No" for Victim and Ref, and select "Yes" for UK1 and UK2.
 - 2.4. In the Father column, select "UK1" for Victim and Ref.
 - 2.5. In the Mother column, select "UK2" for Victim and Ref.
3. Press the **View family tree** button to check whether the family tree is correctly set up or not.
4. Press the **Save** button.

Note

Information on the biological sex, IBD probabilities, paternal lineage, and maternal lineage in the new relationship is automatically determined based on your-defined family tree.

4.2.3 Delete

You can delete a defined relationship from the **Delete** button.

In the pop-up window, select a relationship from the defined relationships. Press the **Save** button to delete the selected relationship.

4.2.4 Reset

If you want to return to default settings, press the **Reset** button. In the pop-up window, press the **Restore default** button.

4.2.5 Family tree

You can check the family tree of each defined relationship. Select a row from the relationship table and press the **Family tree** button.

4.3 Mutation rate

Mutation rates of each autosomal STR marker need to be set to calculate the likelihood ratio (LR) for parent-child relationship. The default mutation rates considering the difference between paternal and maternal and mutational steps (-1, +1, -2, and +2) are reported by Morimoto et al [2].

To set mutation rates, select **Setting ► Mutation rates**. Then the setting window for mutation rates will open automatically.

4.3.1 Edit

You can edit the mutation rates from the **Edit** button.

In the pop-up window, select a locus and change the mutation rates. Press the **Save** button to reflect the changes.

4.3.2 Add

You can add a locus from the **Add** button.

In the pop-up window, enter a locus name and enter the mutation rates. Press the **Save** button to add the locus.

4.3.3 Delete

You can delete a locus from the **Delete** button.

In the pop-up window, select a locus. Press the **Save** button to delete the selected locus.

4.3.4 Reset

If you want to return to default settings, press the **Reset** button. In the pop-up window, press the **Restore default** button.

4.4 Data storage conditions

When handling a large number of victim and reference profiles, it may be difficult to store all the results because RelSearch may run out of memory (see section **Computation time and limits**). You can set the data storage conditions.

To set the conditions, select **Setting ► Data storage conditions**. Then the setting window for data storage will open automatically.

- Minimum LR to be stored: The minimum LR to be stored in the project data. If you want to reduce the file size of the project data, the minimum LR should be increased.

- Maximum number of displayed data: The maximum number of data displayed in the Results tab. If this number is set too high, RelSearch may run out of memory.

Press the **Save** button to reflect the changes. Press the **Reset** button if you want to return to default values.

Chapter 5

Result

This section describes the display of results on the Results tab.

5.1 Summary

When the analysis is complete, summary data is automatically displayed. You can change the data displayed using the buttons in the left sidebar. You can also download the displayed data as a .csv file using the **Download** button.

Default display

Default display depends on whether the input data includes autosomal STR data or not.

If the input data includes autosomal STR data, the data that satisfies the criterion of the minimum LR (see section [Setting](#)) is displayed.

If the input data is composed of only Y-STR and/or mtDNA data, the data that satisfies the criteria to support paternal/maternal lineage (see section [Setting](#)) is displayed.

If the number of data that satisfies the above conditions is more than the upper limit (i.e., the maximum number of displayed data explained in section [Data storage conditions](#)), the number of data displayed is the upper limit.

Identified

If the data supports that a victim (or a reference) has the assumed relationship with only one reference (or victim), the victim-reference pair is classified as "Identified".

Inconclusive

If the data supports that a victim (or a reference) has the assumed relationship with multiple references (or victims), the victim-reference pair is categorized as "Inconclusive".

Excluded by Y-STR or mtDNA

The victim-reference pair is excluded if the LR exceeds the minimum LR but the data does not satisfy the criteria to support paternal/maternal lineage when assuming the paternal/maternal lineage relative.

Minimum LR displayed

You can change the minimum LR displayed. After setting the value, press the **Apply** button. Please note that the amount of data stored in the project data is limited according to the data storage conditions (see [Data storage conditions](#)). Even if you set the minimum LR displayed to less than 1, you cannot check the data with $LR < 1$ as long as the minimum LR to be stored in the project data is set to 1.

5.2 Selected data in detail

The detailed data of a selected victim-reference pair is displayed in **Selected data in detail** tab.

1. Select a row in the summary table.
2. Go to **Selected data in detail** tab.

In **Selected data in detail** tab, you can review and download the analyzed STR, Y-STR, and mtDNA data for the selected victim-reference pair. You can check all members of the family to which the selected reference belongs.

STR tab

Y-STR tab

mtDNA tab

5.3 All candidates

All candidate relatives that relate to the selected victim-reference pair is displayed in **All candidates** tab. This function is useful when the victim-reference pair is classified as "Inconclusive".

1. Select a row in the summary table.
2. Go to **All candidates** tab.

In **All candidates** tab, you can download the displayed data.

5.4 Analysis conditions

You can review the following analysis conditions in **Analysis conditions** tab.

- **Database:** Input .csv file names of each database are displayed.

- **Allele probability:** Unobserved alleles in the population database are displayed, if any. You can download the allele probabilities that are estimated based on Dirichlet distribution (see section [Computational principle](#)).
- **Criteria:** The user-defined criteria to determine whether each victim-reference pair has an assumed relationship or not (see section [Setting](#)) is displayed.
- **Assumed relationship:** Information on the user-defined relationships (see section [Setting](#)) is displayed.
- **Mutation rate:** The user-defined mutation rates (see section [Setting](#)) are displayed.
- **Data storage conditions:** The user-defined conditions for data storage (see section [Setting](#)) is displayed.

Chapter 6

Other functions

6.1 Project

6.1.1 New project

To start a new project, go to **Project ► New project** and press the **New project** button.

Important: The current project should be saved before starting a new project because the current project will be deleted.

6.1.2 Load project

To load a previous project, go to **Project ► Load project**. Select a project file from the **Browse...** button and press the **Load project** button.

Important: The current project should be saved before loading another project because the current project will be deleted.

6.1.3 Save project

To save the current project, go to **Project ► Save project**. Enter the project name and press the **Save as** button.

6.2 View database

Each loaded database can be reviewed in **Database** tab.

6.3 Example files

Example files for each database can be downloaded from the **Example files** tab.

Chapter 7

Computation time and limits

When handling a large number of victim and reference profiles, the computational load of RelSearch will be high. This section provides approximate computation times and limits as a guide.

The following 6 databases have been analyzed to estimate computation time and limits.

ID	No. of victims (V)	No. of references (R)	No. of V-R pairs	Actual V-R relationship (n = 1,000 in each)
1	1,000	1,000	1 million	BB
2	2,000	2,000	4 million	BB, FS
3	3,000	3,000	9 million	BB, FS, MD
4	4,000	4,000	16 million	BB, FS, MD, SS
5	5,000	5,000	25 million	BB, FS, MD, SS, UN
6	6,000	6,000	36 million	BB, FS, MD, SS, UN, AN

BB: brothers (marker type: STR, Y-STR, mtDNA)

FS: father-son (marker type: STR, Y-STR)

MD: mother-daughter (marker type: STR, mtDNA)

SS: sisters (marker type: STR, mtDNA)

UN: uncle-nephew (marker type: STR, Y-STR)

AN: aunt-niece (marker type: STR, mtDNA)

The targeted autosomal STR loci are 21 loci typed by the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific, Waltham, MA). The targeted Y-STR loci are 27 loci typed by Yfiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific). The targeted mtDNA regions were the hypervariable regions (HVR) I (16024-16365) and II (73-340).

Computer specifications are Windows 10 64 bit, Intel Core i9-9900K CPU, 3.6 GHz, 32 GB RAM. Data storage conditions are as follows:

- Minimum LR to be stored: 1
- Maximum number of displayed data: 10,000

The following table shows approximate computation times and warnings related to computation limits.

ID	Computational time	Warning message in the R console
1	4 min	
2	12 min	
3	26 min	
4	53 min	
5	94 min	It seems your data is too big.
6	157 min	It seems your data is too big.

All 6 databases could be analyzed, but a warning message related to computation limits was displayed in the R console when analyzing database IDs 5 and 6 (i.e., performing more than 25 million comparisons between victim and reference profiles). In addition, the number of victim-reference pairs that meet the criteria for kinship estimation has reached the upper limit of the data displayed (i.e., 10,000) when analyzing database IDs 5 and 6. These results suggest that the upper limit of computation is less than 25 million comparisons. Please note that the limit is based only on the computer specifications and databases used in this study.

Chapter 8

Computational principle

8.1 STR

8.1.1 Likelihood ratio

The likelihood ratio (LR) is calculated for each victim-reference pair by assuming the following two hypotheses:

H_1 : The victim and the reference are a certain relationship,

H_2 : The victim and the reference are unrelated.

The equation of the LR is as follows:

$$LR = \frac{Pr(G_v, G_r | H_1)}{Pr(G_v, G_r | H_2)} \quad (1)$$

where G_v and G_r denote the genotypes of the victim and the reference, respectively.

In RelSearch, users can set the minimum threshold of the LR to estimate that the victim and reference have the relationship assumed in H_1 . The default threshold is 100, which is determined based on the verbal scale "Moderate Support" written in the Scientific Working Group on DNA Analysis Methods (SWGDM) guideline [3].

If full profiles are obtained for a victim-reference pair under the assumption of the relationship other than the parent-child (i.e., without considering drop-out and mutations in calculation), the numerator and the denominator of the LR are calculated using the probabilities that two, one and zero alleles are identity by descent (IBD).

$$\begin{aligned} Pr(G_v, G_r | H) = & Pr(G_v, G_r | IBD = 2)Pr(IBD = 2 | H) \\ & + Pr(G_v, G_r | IBD = 1)Pr(IBD = 1 | H) \\ & + Pr(G_v, G_r | IBD = 0)Pr(IBD = 0 | H) \end{aligned} \quad (2)$$

The specific method for these calculations is based on the method of Wenk et al [4].

8.1.2 Allele probabilities

The allele probabilities are assumed to be based on Dirichlet distributions in each locus [2]. The probability of allele x (p_x) can be estimated using the following formula:

$$p_x = \frac{\alpha_x + 1}{\sum_{x=1}^X (\alpha_x + 1)} \quad (3)$$

where α_x is the number of observations of the x th allele ($x = 1, 2, \dots, X$) and X is the number of observed allele types.

If some alleles of victims or references are unobserved in the population database, these alleles are added to the observed allele types as zero observation (i.e., $\alpha_x = 0$), and p_x of all alleles are re-estimated [2].

8.1.3 Allelic drop-out

When only one allele is observed and another allele may have dropped out in a locus, RelSearch assumes that the genotype is both homozygous (i.e., no drop-out allele) and heterozygous (i.e., one drop-out allele). The probability of observing the profile is the same in all candidate genotypes. The method corresponds to the "Method 1" proposed by Dørum et al [5].

Here, G_v^* and G_r^* denote the observed profiles of the victim and the reference, respectively. G_{v_i} and G_{r_j} denote the i th candidate genotype of the victim and the j th candidate genotype of the reference, respectively.

Suppose that G_v^* is 15/- (i.e., partial profile) and G_r^* is 16/17. In this case, RelSearch considers the following candidate genotypes of the victim:

G_{v_1} : 15/15,

G_{v_2} : 15/16,

G_{v_3} : 15/17, and

G_{v_4} : 15/Q.

Q denotes any undetected alleles other than 15, 16, and 17.

The candidate genotype of the reference is only 16, 17 because two heterozygote alleles are detected.

In this case, the likelihood function considering allelic drop-out is as follows:

$$\begin{aligned}
& Pr(G_v^* : 15/-, G_r^* : 16/17|H) \\
& = Pr(G_v^* : 15/-, G_r^* : 16/17|G_{v_i} : 15/15, G_r : 16/17)Pr(G_{v_i} : 15/15, G_r : 16/17|H) \\
& + Pr(G_v^* : 15/-, G_r^* : 16/17|G_{v_i} : 15/16, G_r : 16/17)Pr(G_{v_i} : 15/16, G_r : 16/17|H) \\
& + Pr(G_v^* : 15/-, G_r^* : 16/17|G_{v_i} : 15/17, G_r : 16/17)Pr(G_{v_i} : 15/17, G_r : 16/17|H) \\
& + Pr(G_v^* : 15/-, G_r^* : 16/17|G_{v_i} : 15/Q, G_r : 16/17)Pr(G_{v_i} : 15/Q, G_r : 16/17|H)
\end{aligned} \tag{4}$$

This equation is generalized as follows:

$$Pr(G_v^*, G_r^*|H) = \sum_i \sum_j Pr(G_v^*, G_r^*|G_{v_i}, G_{r_j})Pr(G_{v_i}, G_{r_j}|H) \tag{5}$$

$Pr(G_v^*, G_r^*|G_{v_i}, G_{r_j})$ is assumed to be the same irrespective of the candidate genotypes. Therefore, the likelihood function can be rewritten as follows:

$$Pr(G_v^*, G_r^*|H) = \sum_i \sum_j Pr(G_{v_i}, G_{r_j}|H) \tag{6}$$

$Pr(G_{v_i}, G_{r_j}|H)$ values are calculated according to equation (1) under the assumption of the relationship other than the parent-child. When assuming the parent-child relationship, mutational events are considered for the calculation as explained in the next session.

8.1.4 Mutation

When assuming the parent-child relationship, mutational events are considered for the calculation of $Pr(G_{v_i}, G_{r_j}|H)$ in equation (2). Suppose that a victim-reference pair is assumed to be parent-child, and G_{v_i} is a/b and G_{r_j} is c/d. RelSearch assumes all inheritance patterns from the parent to the child including mutational events. The LR is calculated according to [6,7].

$$LR = \frac{(\mu_{a \rightarrow c} + \mu_{b \rightarrow c})p_d + (\mu_{a \rightarrow d} + \mu_{b \rightarrow d})p_c}{4p_cp_d} \tag{7}$$

where p_a, p_b, p_c , and p_d denote the probabilities of the allele a, b, c , and d , respectively. $\mu_{a \rightarrow c}$ denotes the probability that the parent allele a is inherited as the allele c to the child. $\mu_{a \rightarrow c}, \mu_{b \rightarrow c}, \mu_{a \rightarrow d}$, and $\mu_{b \rightarrow d}$ depend on the sex of the parent (i.e., paternal or maternal), and the mutational steps are assumed to be -2, -1, +1, and +2 steps. The mutation rates reported by Morimoto et al [2] are used as the default values.

8.2 Y-STR

RelSearch analyzes the number of mismatched loci and the total mutational steps between victim and reference Y-STR profiles. The following table shows some examples of the analysis results in one locus.

Victim profile	Reference profile	Mismatch	Mutational steps	Ignore due to allelic drop-out
15	15		0	
15	16	Yes	1	
15	17	Yes	2	
-	15			Yes
15	15, 17			Yes
15	16, 17	Yes	1	
15, 16	15, 17	Yes	1	
15, 16	15, 18	Yes	2	
15, 17	16, 18	Yes	Unable to calculate	

RelSearch ignores loci where the victim and reference profile match when considering the allelic drop-out. Suppose that the victim profile is 15, and the reference profile is 15, 17 in a duplicated marker (e.g., DYS385). If the allele 17 is assumed to be dropped out in the victim profile, the assumed victim profile is the same as the reference profile.

If multiple mutational events are assumed in a duplicated marker, RelSearch does not calculate the mutational steps for the marker due to the difficulty of the estimation of the actual mutational process (i.e., "Unable to calculate" in the above table). Especially in DYF387S1, three alleles may be observed and the mutational process related to these alleles is too complex.

After analyses of all victim-reference profiles, RelSearch investigates whether each victim-reference profile supports paternal lineage or not based on the following criteria.

- Maximum number of mismatched loci: The upper limit of the number of mismatched loci between victim and reference profiles for which paternal lineage is not negated.
- Maximum total mutational steps: The upper limit of total mutational steps between victim and reference profiles for which paternal lineage is not negated.

The default criteria for the maximum number of mismatched loci is 4 and for the maximum total mutational steps is 5, which is determined based on a previous study [8].

8.3 mtDNA

RelSearch investigates the number of mismatched nucleotides between the victim and the reference sequences. There is one criterion to support maternal lineage as follow:

- Maximum number of inconsistency: The upper limit of inconsistency between victim and reference profiles for which maternal lineage is not negated.

Information on the ranges of the mtDNA sequences in each profile is needed because RelSearch targets the ranges read in both victim and reference profiles. The mtDNA profiles are expressed as the differences to a reference sequence such as the revised Cambridge Reference Sequence [1] (e.g., 73G, 315.1C, and so on). RelSearch only recognizes differences in strings of each mtDNA profile; therefore, the rule of the notation for mtDNA profiles should be aligned in each profile by software users.

The default maximum number of inconsistency is 1, which is determined according to the SWGDAM guideline for mtDNA analysis interpretation [9].

Reference

- [1] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nature Genetics* 23 (1999) 147.
- [2] C. Morimoto, H. Tsujii, S. Manabe, S. Fujimoto, E. Hirai, Y. Hamano, K. Tamaki, Development of a software for kinship analysis considering linkage and mutation based on a Bayesian network, *Forensic Science International: Genetics* 47 (2020) 102279.
- [3] Scientific Working Group on DNA Analysis Methods, Recommendations of the SWGDAM ad hoc working group on genotyping results reported as likelihood ratios., (2018).
- [4] R.E. Wenk, M. Traver, F.A. Chiafari, Determination of sibship in any two persons, *Transfusion* 36 (1996) 259–262.
- [5] G. Dørum, D. Kling, C. Baeza-Richer, M. García-Magariños, S. Sæbø, S. Desmyter, T. Egeland, Models and implementation for relationship problems with dropout, *International Journal of Legal Medicine* 129 (2015) 411–423.
- [6] F. Ricciardi, K. Slooten, Mutation models for DVI analysis, *Forensic Science International: Genetics* 11 (2014) 85–95.
- [7] T. Egeland, N. Pinto, A. Amorim, Exact likelihood ratio calculations for pairwise cases, *Forensic Science International: Genetics* 29 (2017) 218–224.
- [8] H. Liu, X. Li, J. Mulero, A. Carbonaro, M. Short, J. Ge, A convenient guideline to determine if two Y-STR profiles are from the same lineage, *ELECTROPHORESIS* 37 (2016) 1659–1668.
- [9] Scientific Working Group on DNA Analysis Methods, Interpretation guidelines for mitochondrial DNA analysis by forensic DNA testing laboratories., (2019).