

Lecture 8: Complex DVI cases with DIVIANA

ISFG-GHEP Online School 2024

Kinship and pedigree analysis: Methods and applications

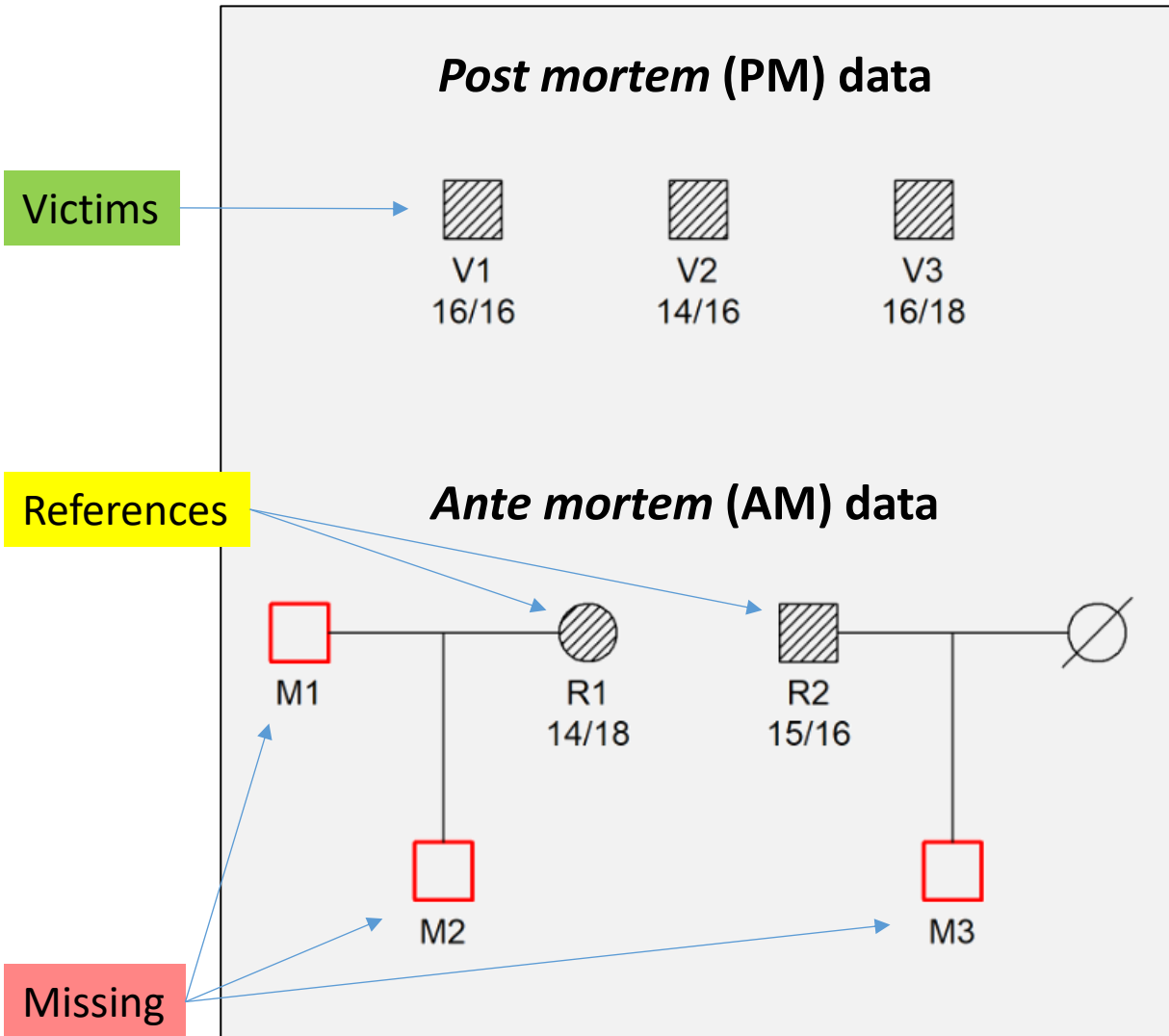


Plan

- Disaster victim identification
 - Terminology and standard methods
- Challenging cases
 - Families with multiple missing
- New statistical tools
 - Generalised likelihood ratio (GLR)
 - Relatedness triangle
- New software: **DIVIANA**



Disaster victim identification (DVI)



Computational approaches

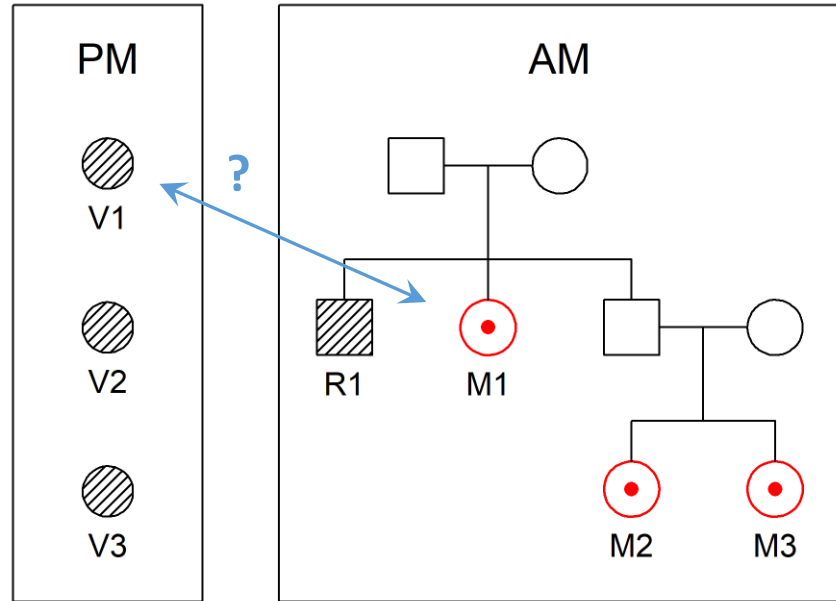
➤ Pairwise LR

	M1	M2	M3
V1	1	0	0.39
V2	1	16142	0
V3	1	0	6266

➤ Joint / global

Vigeland & Egeland (Scientific Reports, 2021):
Joint DNA-based disaster victim identification

Joint analysis of DVI data



Victims and R1 typed with 23 STRs.

Pairwise LR

	M1	M2	M3
V1	121.3	398.6	398.6
V2	0.6	36.6	36.6
V3	0.00	1.3	1.3

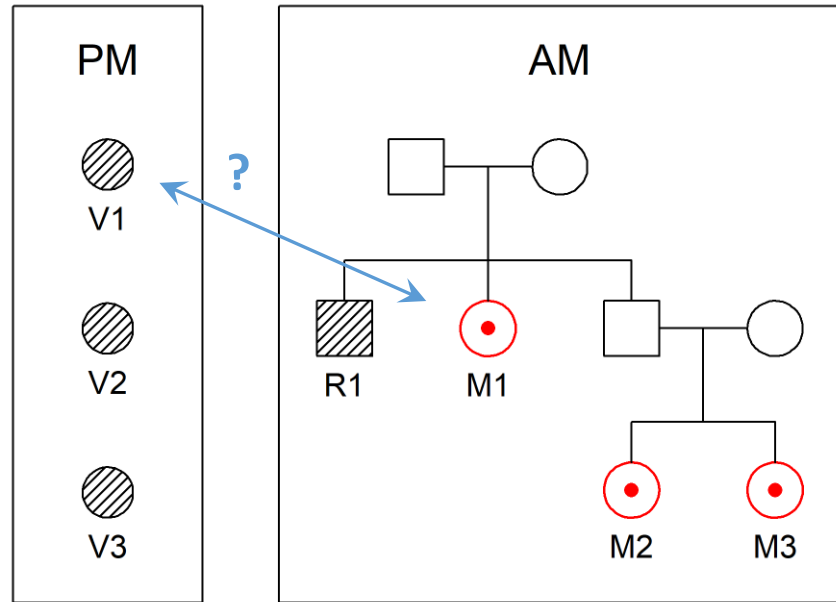
Joint likelihoods

	V1	V2	V3	loglik
1	M1	M2	M3	-257.7
2	M1	M3	M2	-257.7
3	*	M2	M3	-268.9
4	*	M3	M2	-268.9
5	M2	M1	M3	-272.6
6	M3	M1	M2	-272.6
7	M1	*	M2	-276.5
8	M1	*	M3	-276.5
9	M2	M3	M1	-276.6
⋮				
32	*	*	*	-286.0
33	*	M1	*	-286.5
34	*	*	M1	-292.2

A tool for interpreting joint results:

Generalised likelihood ratio (GLR)

$$\text{GLR} = \frac{\max_{i \in H_1} L_i}{\max_{j \in H_2} L_j}$$



Victims and R1 typed with 23 STRs.

$H_1: V1 = M1$
 $H_2: V1 \neq M1$

$$\text{GLR} = \frac{L_1}{L_3} = \frac{e^{-257.7\dots}}{e^{-268.9\dots}} = 70\,582$$

Strong evidence

Joint likelihoods

	V1	V2	V3	loglik	
1	M1	M2	M3	-257.7	✓
2	M1	M3	M2	-257.7	✓
3	*	M2	M3	-268.9	✗
4	*	M3	M2	-268.9	✗
5	M2	M1	M3	-272.6	✗
6	M3	M1	M2	-272.6	✗
7	M1	*	M2	-276.5	✓
8	M1	*	M3	-276.5	✓
9	M2	M3	M1	-276.6	✗
⋮					
32	*	*	*	-286.0	✗
33	*	M1	*	-286.5	✗
34	*	*	M1	-292.2	✗

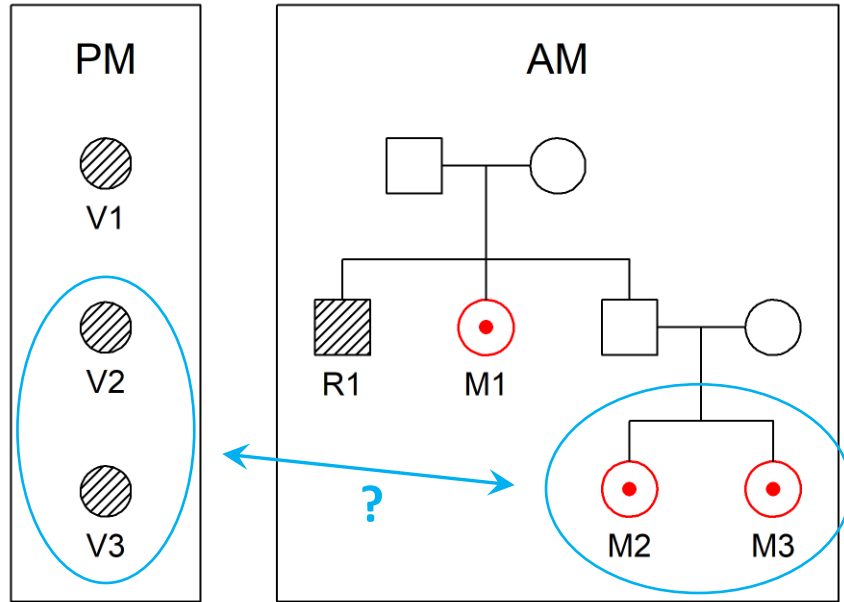
Interpretation of $GLR = x$

$$GLR = \frac{\max_{i \in H_1} L_i}{\max_{j \in H_2} L_j}$$

*The best explanation of the data given H1
is x times more likely
than the best explanation given H2*

Symmetric matches

$$\text{GLR} = \frac{\max_{i \in H_1} L_i}{\max_{j \in H_2} L_j}$$



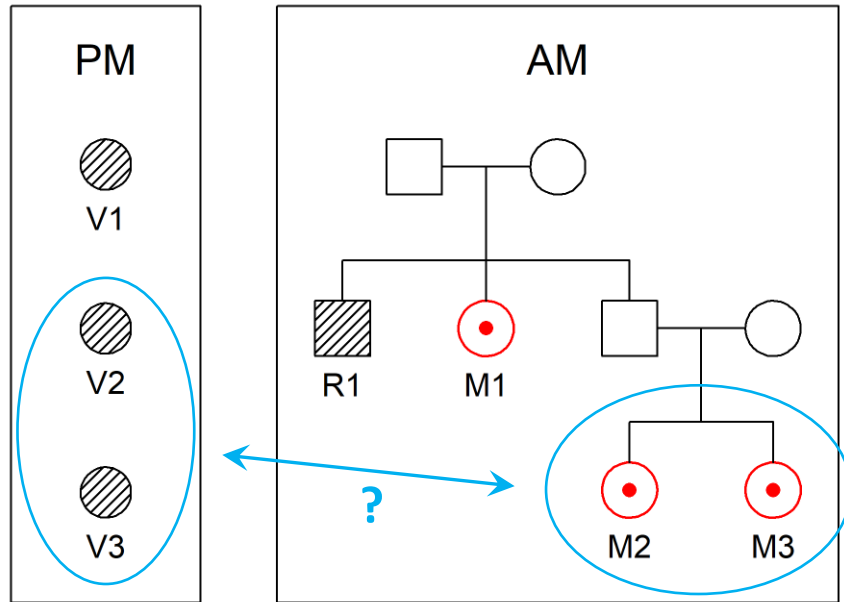
Cannot separate these! →

Joint likelihoods

	V1	V2	V3	loglik
1	M1	M2	M3	-257.7
2	M1	M3	M2	-257.7
3	*	M2	M3	-268.9
4	*	M3	M2	-268.9
5	M2	M1	M3	-272.6
6	M3	M1	M2	-272.6
7	M1	*	M2	-276.5
8	M1	*	M3	-276.5
9	M2	M3	M1	-276.6
⋮				
32	*	*	*	-286.0
33	*	M1	*	-286.5
34	*	*	M1	-292.2

Symmetric matches

$$\text{GLR} = \frac{\max_{i \in H_1} L_i}{\max_{j \in H_2} L_j}$$



$$H_1: \{V2, V3\} = \{M2, M3\}$$

$$H_2: \{V2, V3\} \neq \{M2, M3\}$$

$$\text{GLR} = 2.9 \times 10^6$$

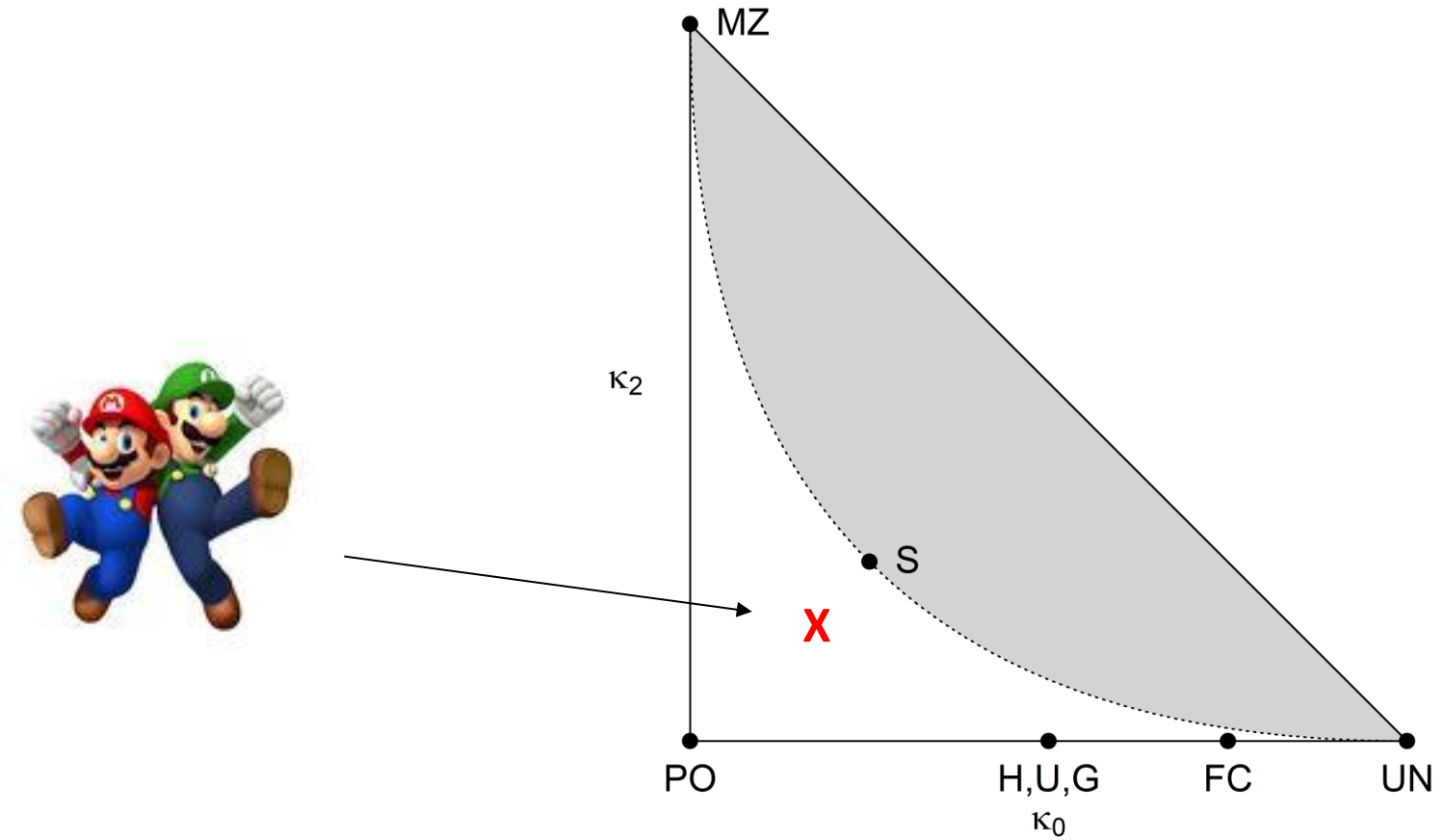
Strong evidence!

Joint likelihoods

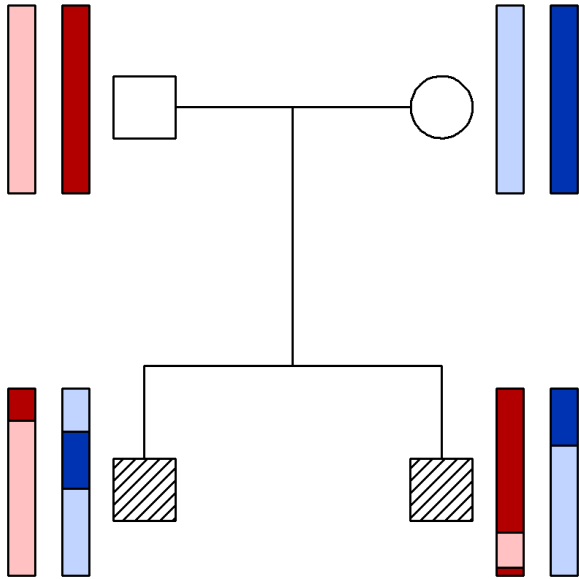
	V1	V2	V3	loglik	
1	M1	M2	M3	-257.7	✓
2	M1	M3	M2	-257.7	✓
3	*	M2	M3	-268.9	✓
4	*	M3	M2	-268.9	✓
5	M2	M1	M3	-272.6	✗
6	M3	M1	M2	-272.6	✗
7	M1	*	M2	-276.5	✗
8	M1	*	M3	-276.5	✗
9	M2	M3	M1	-276.6	✗
⋮					
32	*	*	*	-286.0	✗
33	*	M1	*	-286.5	✗
34	*	*	M1	-292.2	✗

Conclusion: V2 and V3 are M2 and M3 (but we don't know who-is-who)

The relatedness triangle

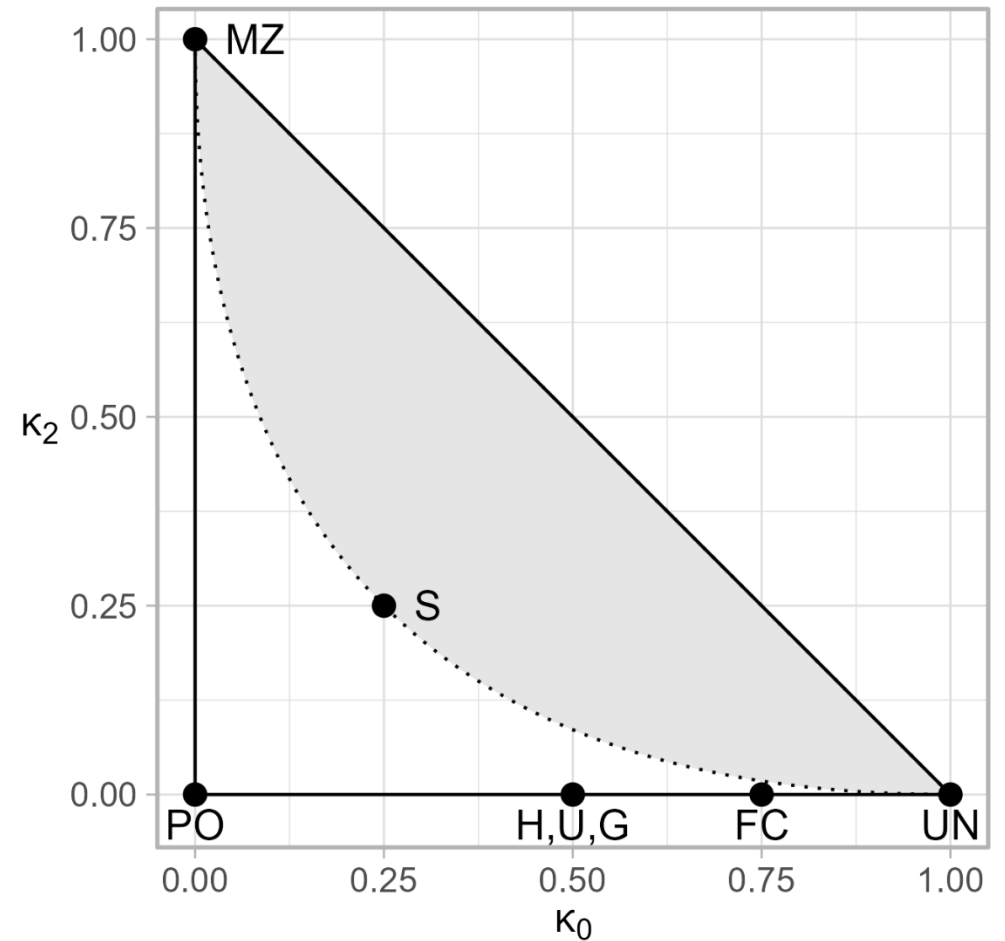


The "kappa" coefficients



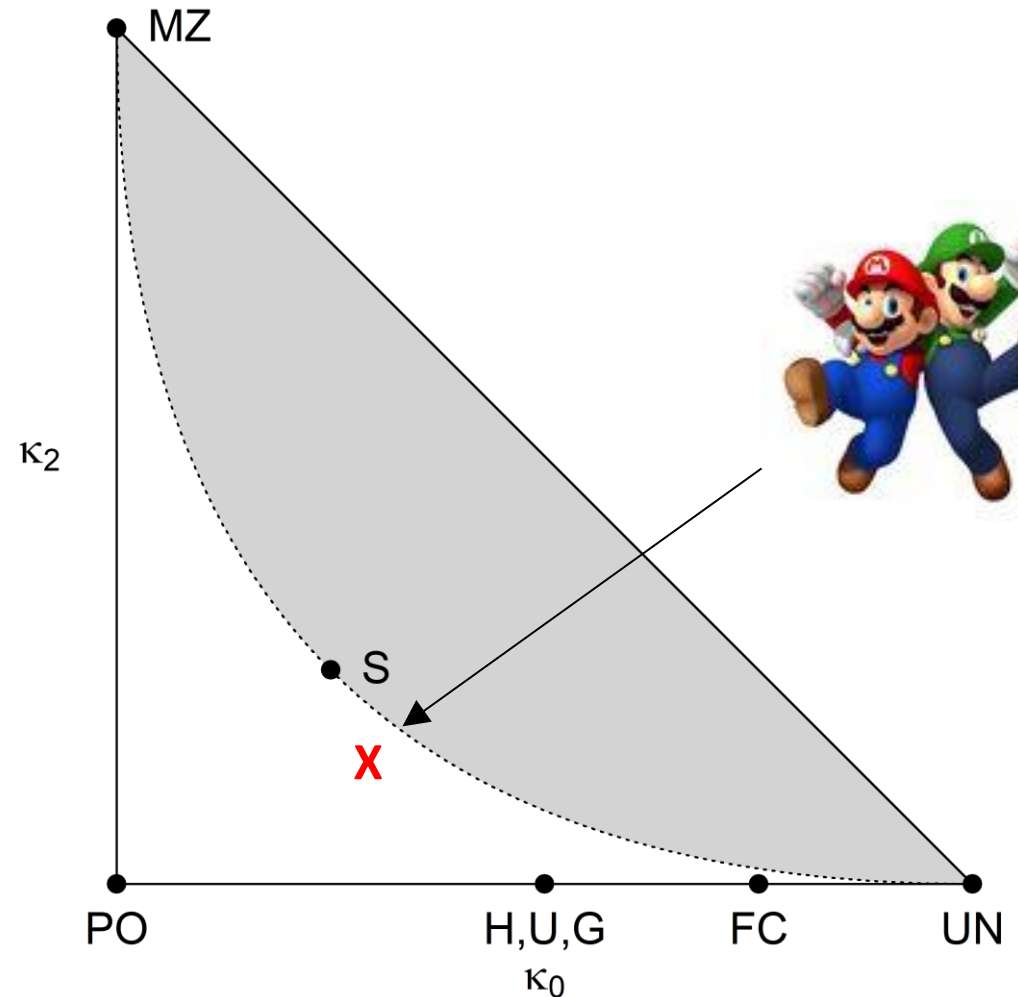
K_0 = proportion of genome with IBD=0
 K_1 = proportion of genome with IBD=1
 K_2 = proportion of genome with IBD=2

$$K_0 + K_1 + K_2 = 1$$



Maximum likelihood estimation

- Thompson (1975)
- Input: Genotypes from two individuals
- Output: Most likely point in the triangle



DIVIANA





AM data



PM data



Mutation model



Frequency database



Pedigrees

Overview



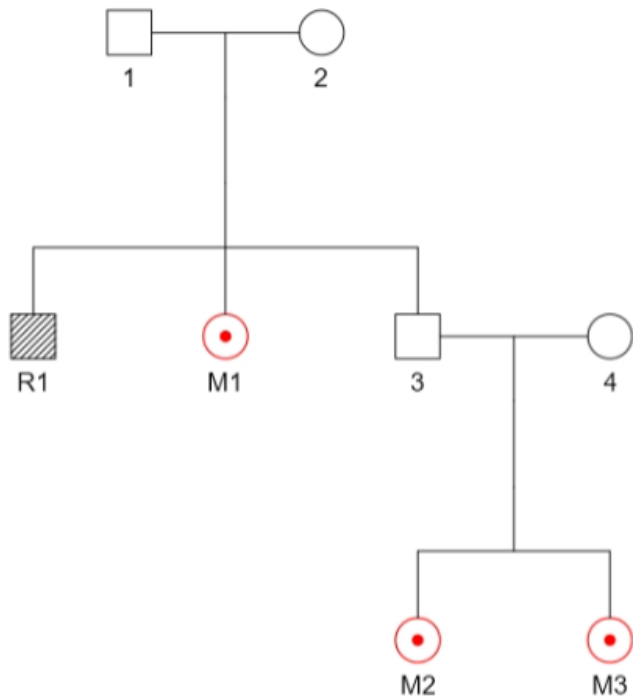
New



Family 1/1



Edit



Labels



3 victims (0M/3F)
3 missing (0M/3F)
1 typed ref
1 ref family
Number of markers, PM and AM: 23



AM data

- Click on individuals to make changes
- Click **Undo** if you make a mistake
- Toggle *missing/nonmissing*
- Mark as *untyped*
- Select reference individual

Pedigree builder

- Add child
- Sibling
- Parents
- Swap sex
- Remove

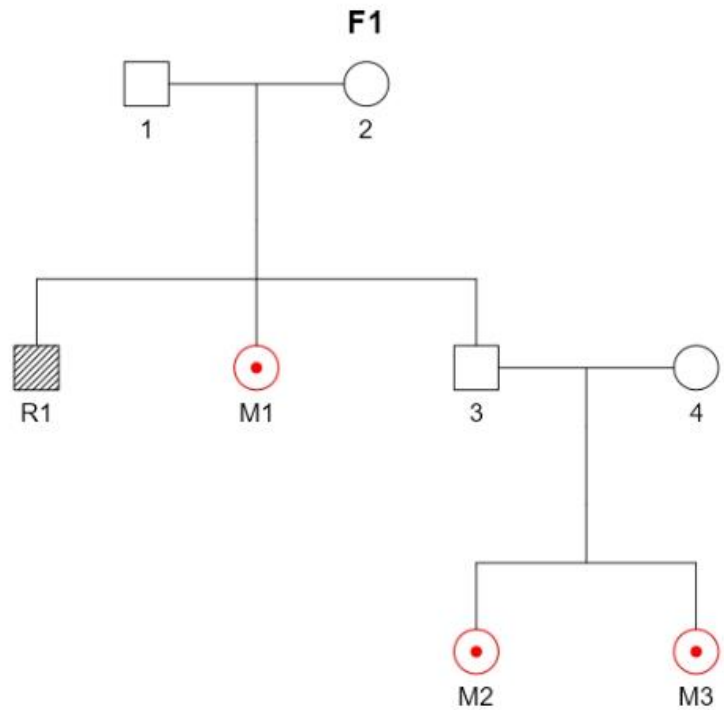
Edit role:

- Missing
- Untyped

Reference

Select

Reset Undo



Save Cancel

i 3 victims (0M/3F)
3 missing (0M/3F)
1 typed ref
1 ref family
Number of markers, PM and AM: 23



AM data +

PM data +

Mutation model +

Frequency database +

Pedigrees

Overview

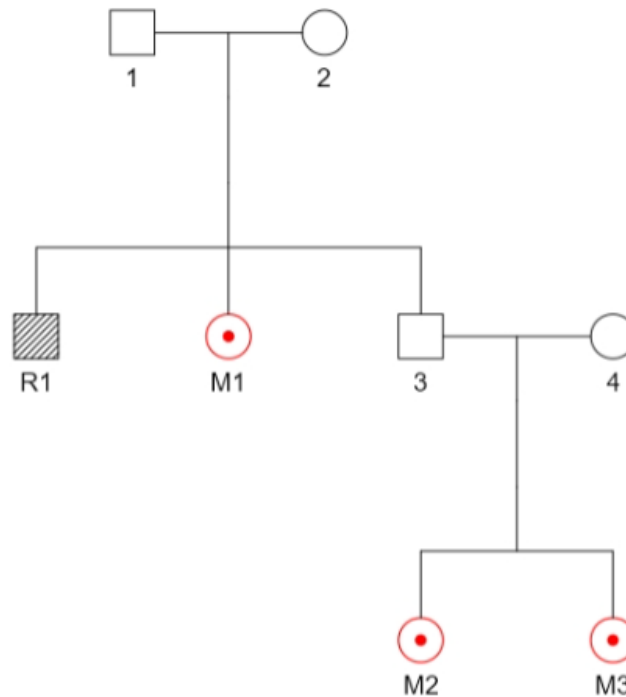
New



Family 1/1



Edit



Labels



3 victims (0M/3F)
3 missing (0M/3F)
1 typed ref
1 ref family
Number of markers, PM and AM: 23



AM data

Familias

Browse...

No file selected

	D3S1358	TH01	D21S11	D18S51	PENTA_E	D5S818	D13S317
R1	16/17	6/9.3	29/31	14/16	7/12	11/11	10/13

PM data

Mutation model

Frequency database

Pedigrees

Overview

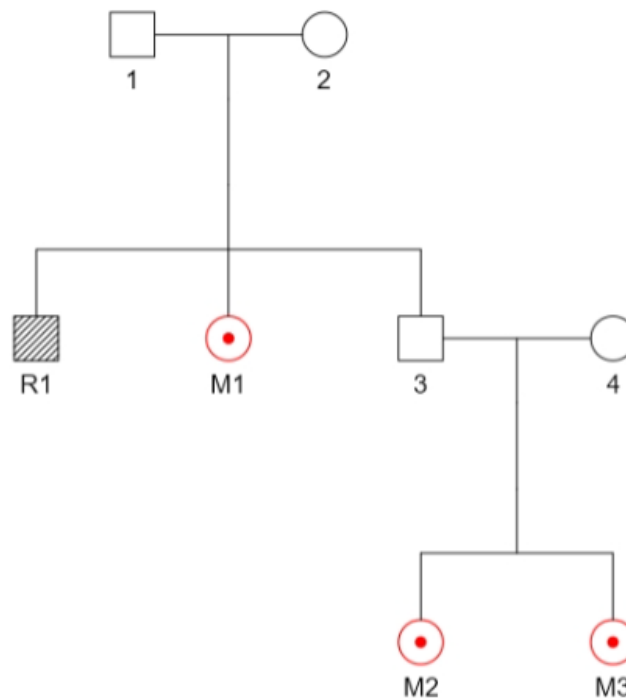
New



Family 1/1



Edit



Labels



3 victims (0M/3F)
 3 missing (0M/3F)
 1 typed ref
 1 ref family
 Number of markers, PM and AM: 23



AM data



PM data



Mutation model



Frequency database



Pedigrees

Overview



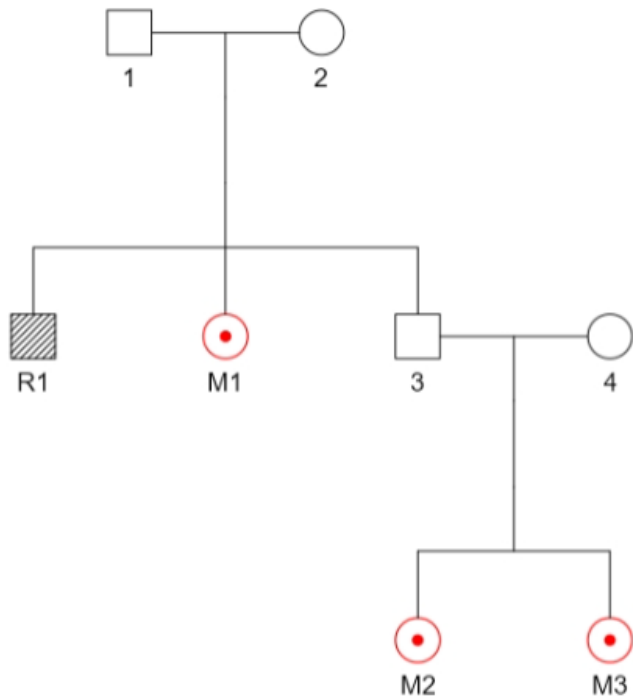
New



Family 1/1



Edit



Labels



3 victims (0M/3F)
3 missing (0M/3F)
1 typed ref
1 ref family
Number of markers, PM and AM: 23



AM data +

PM data +

Mutation model -

Marker 1/23: **D3S1358**

No model Equal Prop Stepwise

	Rate	Rate2	Range
Female:	0,001		
Male:	0,001		

Apply to this marker

Apply to all markers

Frequency database +

Pedigrees

Overview

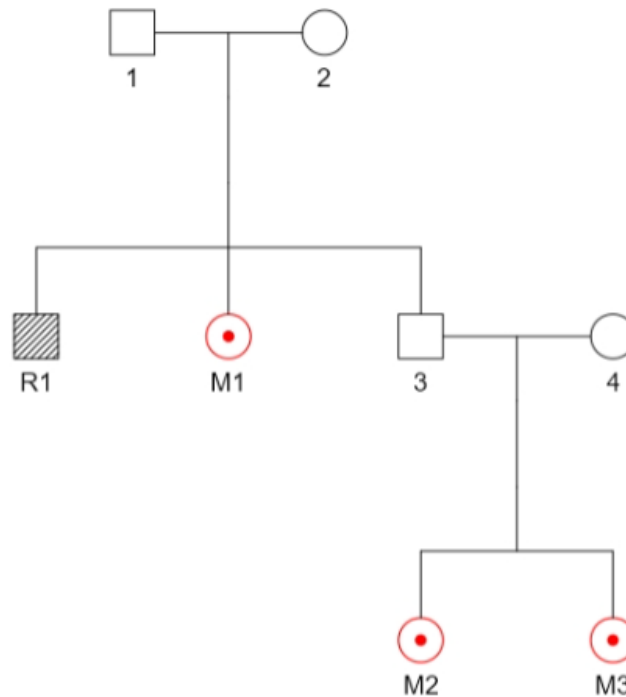
New



Family 1/1



Edit



Labels +



3 victims (0M/3F)
 3 missing (0M/3F)
 1 typed ref
 1 ref family
 Number of markers, PM and AM: 23



AM data +

PM data +

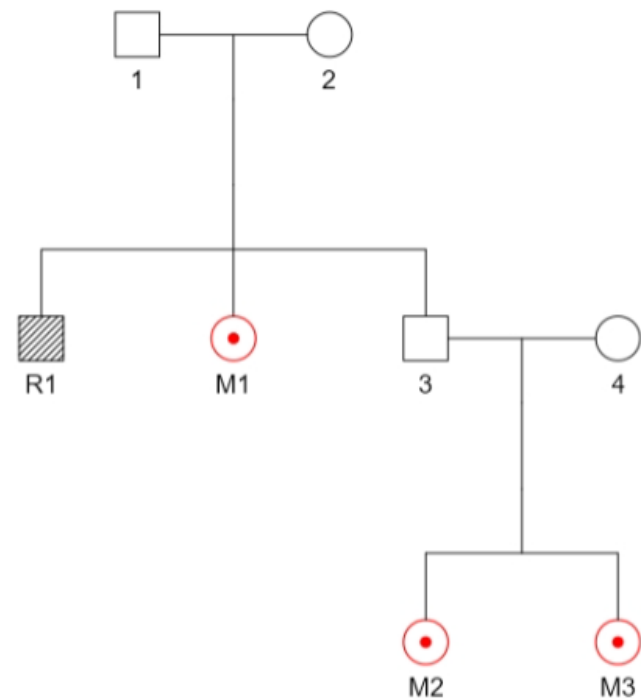
Mutation model +

Frequency database +

Pedigrees

Overview -

New ◀ Family 1/1 ▶ Edit



Labels +



3 victims (0M/3F)
3 missing (0M/3F)
1 typed ref
1 ref family
Number of markers, PM and AM: 23



SOLVE

Identifications AM PM log

Family	Missing	Sample	LR	GLR	Conclusion	Comment
F1	M1	V1	1.21×10^2	7.06×10^4	Match (GLR)	Joint analysis {M1,M2,I
F1	M2	V2/V3		2.95×10^6	Symmetric match	Full siblings: {M2, M3}
F1	M3	V2/V3		2.95×10^6	Symmetric match	Full siblings: {M2, M3}



Settings

LR threshold

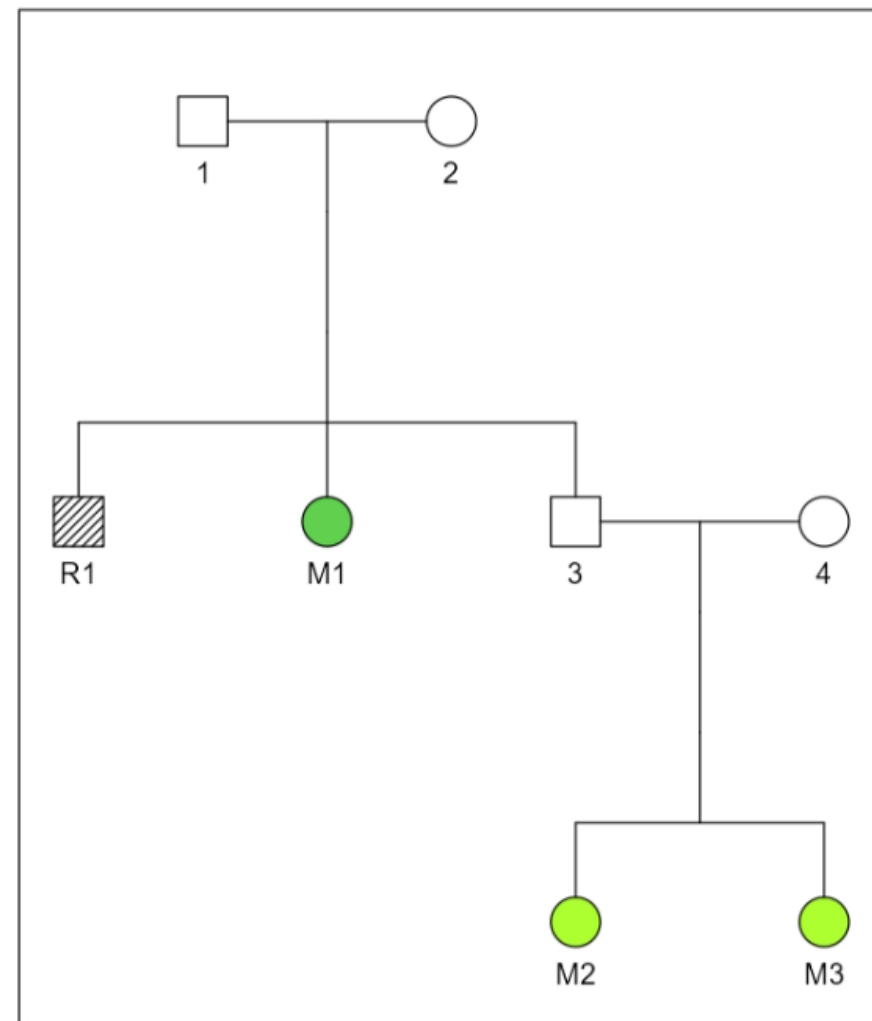
10000

Ignore Sex



Download

Result plot





SOLVE

Identifications AM PM log

Settings

LR threshold

1e4

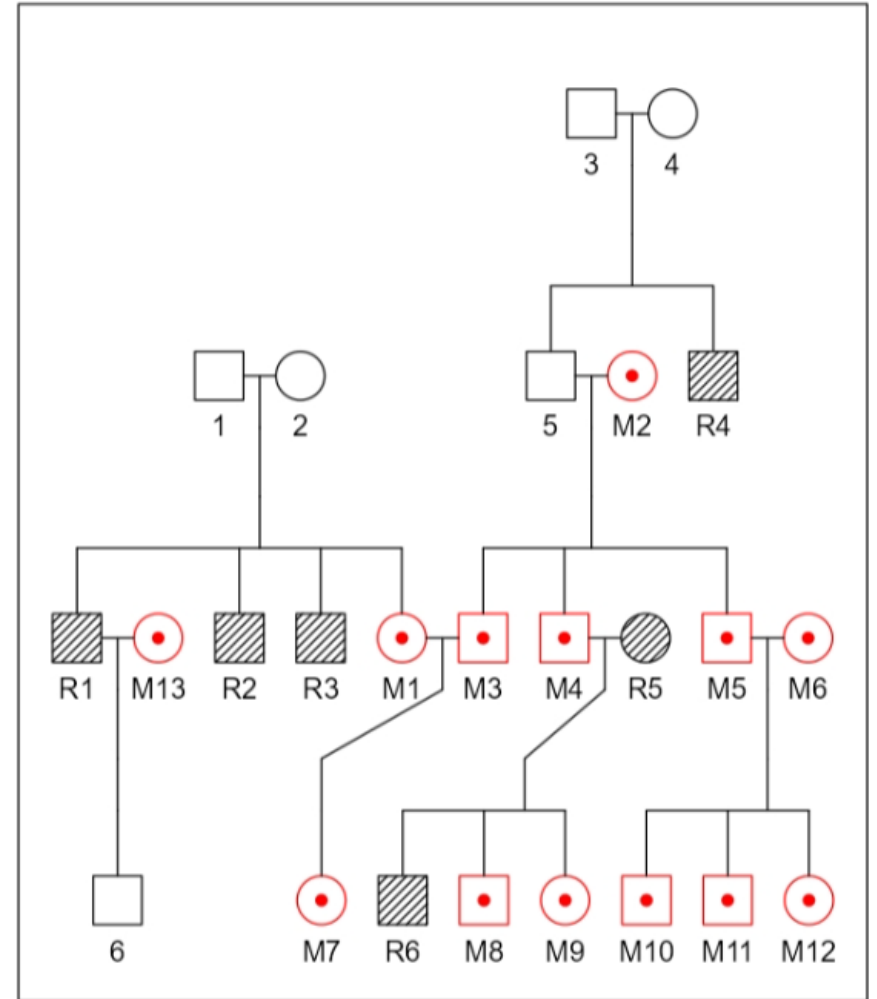
Ignore Sex



Download

Family	Missing	Sample	LR	GLR	Conclusion	Comment
F1	M1					
F1	M2					
F1	M3					
F1	M4					
F1	M5					
F1	M6					
F1	M7					
F1	M8					
F1	M9					
F1	M10					
F1	M11					
F1	M12					
F1	M13					

Result plot





SOLVE

Identifications AM PM log

Settings

LR threshold

1e4

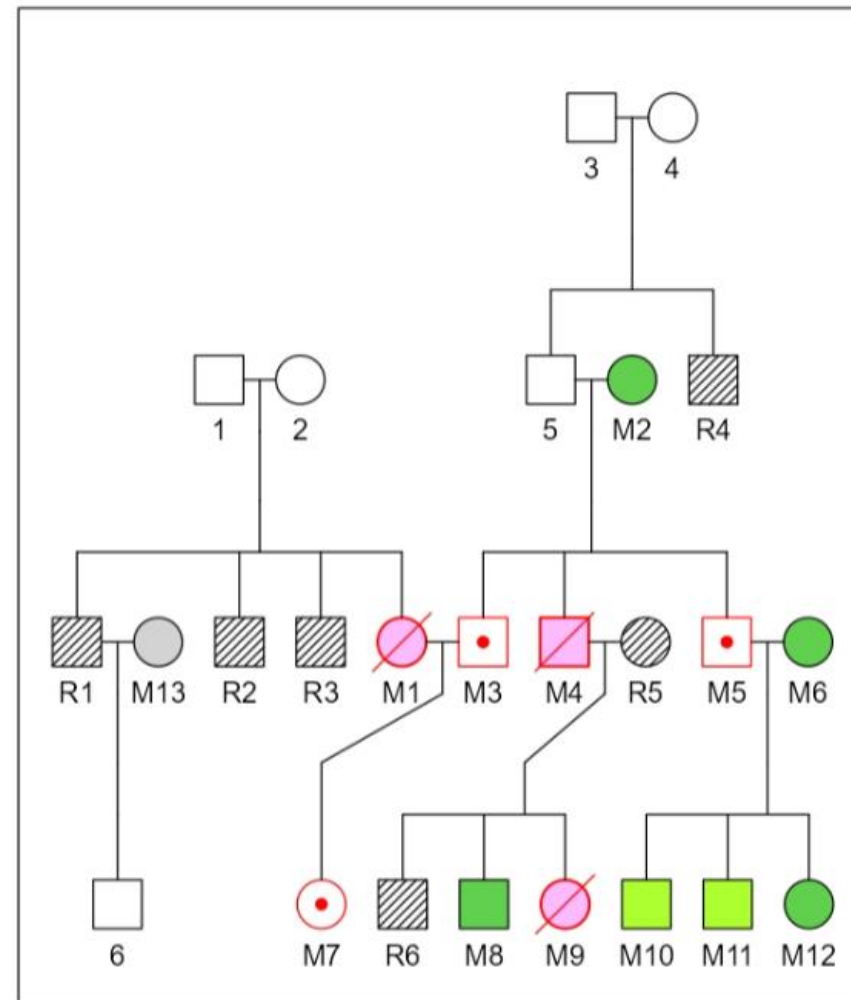
Ignore Sex



Download

Family	Missing	Sample	LR	GLR	Conclusion	Comment
F1	M1				Excluded	5+ inconsistencies
F1	M2	V5	1.70×10^3	1.44×10^6	Match (GLR)	Joint analysis (M2,M3,I
F1	M3				Inconclusive GLR	
F1	M4				Excluded	3+ inconsistencies
F1	M5				Inconclusive GLR	
F1	M6	V1	1.00	7.72×10^8	Match (GLR)	Joint analysis (M2,M3,I
F1	M7				Inconclusive GLR	
F1	M8	V4	2.85×10^6		Undisputed	Step 1
F1	M9				Excluded	3+ inconsistencies
F1	M10	V2		2.58×10^6	Symmetric match	V2 also matches M11
F1	M11	V2		2.58×10^6	Symmetric match	V2 also matches M10
F1	M12	V3	4.78	1.02×10^7	Match (GLR)	Joint analysis (M2,M3,I
F1	M13				Nonidentifiable	Unrelated to R1, R2, R

Result plot





Live demo!



Where can I
find DIVIANA??

Beta version online

<https://magnusdv.shinyapps.io/diviana>

GitHub

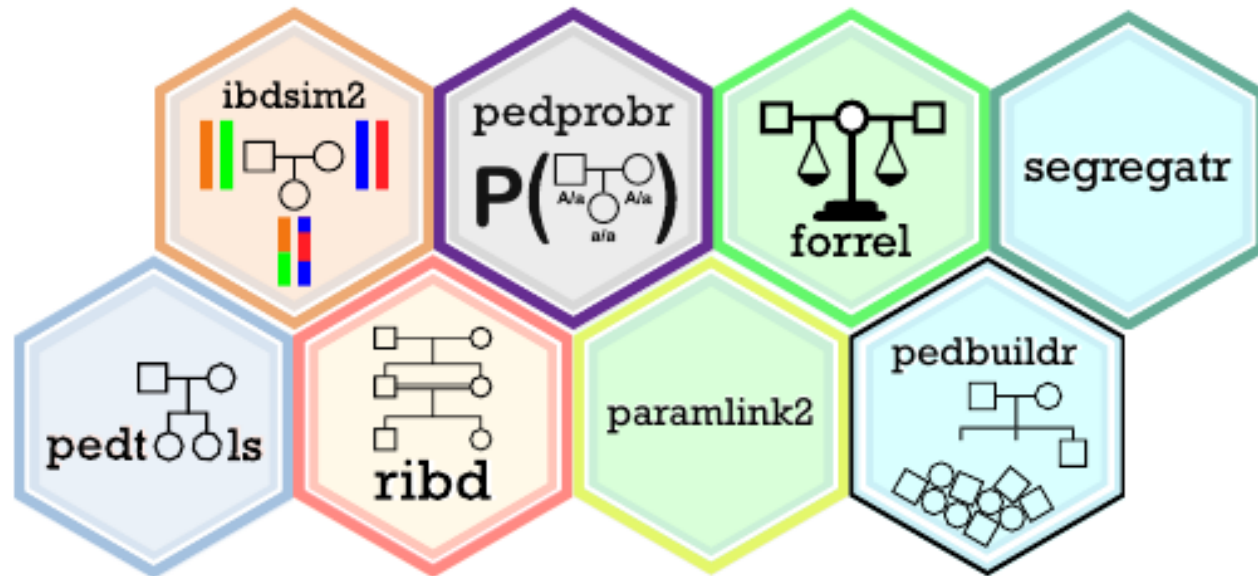
<https://github.com/magnusdv/diviana>

```
# Install dev versions of pedsuite packages
pkgs = c("pedtools", "ribd", "forrel", "dvir", "diviana")
for(pkg in pkgs)
  remotes::install_github(paste0("magnusdv/", pkg))
```

```
diviana::launchApp()
```

The pedsuite:

A collection of packages for pedigree analysis in R



Home page: <https://magnusdv.github.io/pedsuite>

Source code: <https://github.com/magnusdv>

Developers: Magnus D. Vigeland, Thore Egeland

PEDIGREE ANALYSIS IN R



Magnus Dehli Vigeland





Diviana eudoreella