

Kinship and pedigree analysis: Methods and applications

Magnus Dehli Vigeland and Thore Egeland

Solutions for exercise set VII. DNA based disaster victim identification

Some answers are given in the exercise. Here we add some comments and remaining answers.

The **pedsuite** R packages are used throughout:

```
library(pedsuite)
library(dvir) # not a core package, so must be loaded separately
```

Exercise VII-1

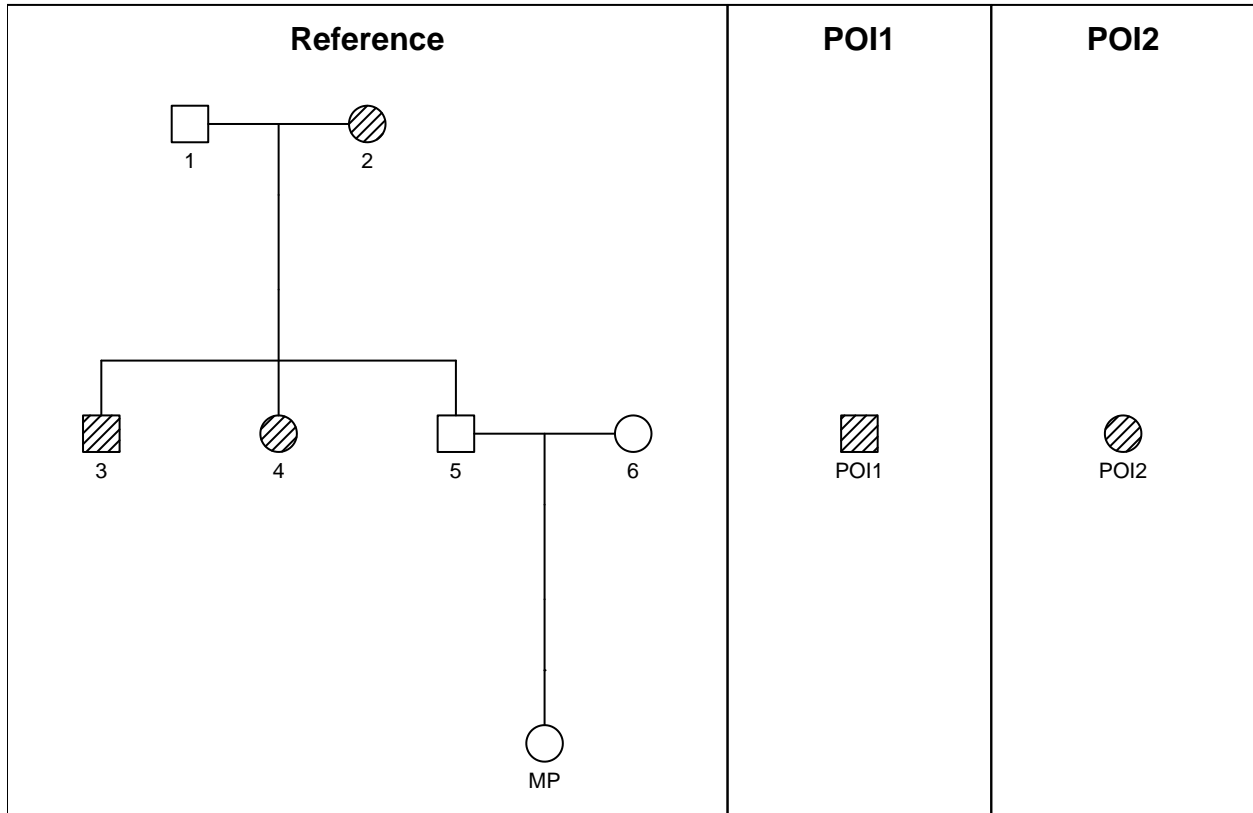
a)

Summary and plot:

```
summary(mpi)

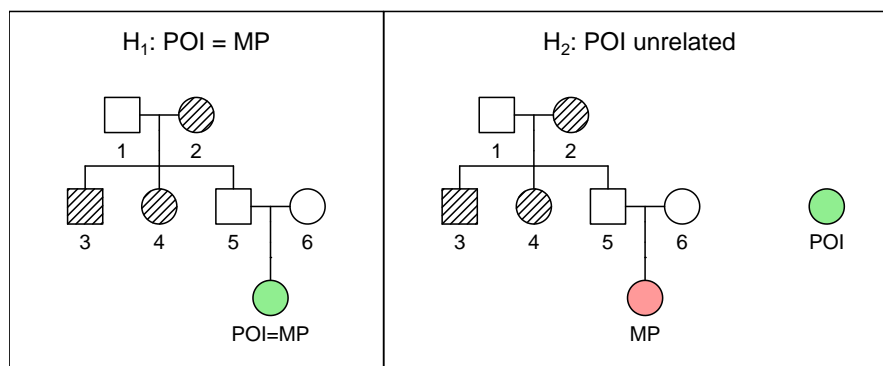
## Ped list with 3 connected components (# members = 7, 1, 1).
## In total 9 individuals (4 males, 5 females, 0 unknown).
##
## --- component 1 ---
## Pedigree with 7 members (3 males, 4 females, 0 unknown).
## 3 generations, 3 founders, 3 leaves.
## 20 attached markers (9 - 26 alleles).
## 3 typed members.
##
## --- component 2 ---
## Singleton (male) labelled 'POI1'.
## 20 attached markers (9 - 26 alleles).
##
## --- component 3 ---
## Singleton (female) labelled 'POI2'.
## 20 attached markers (9 - 26 alleles).
```

```
plotPedList(mpi, hatched = typedMembers)
```



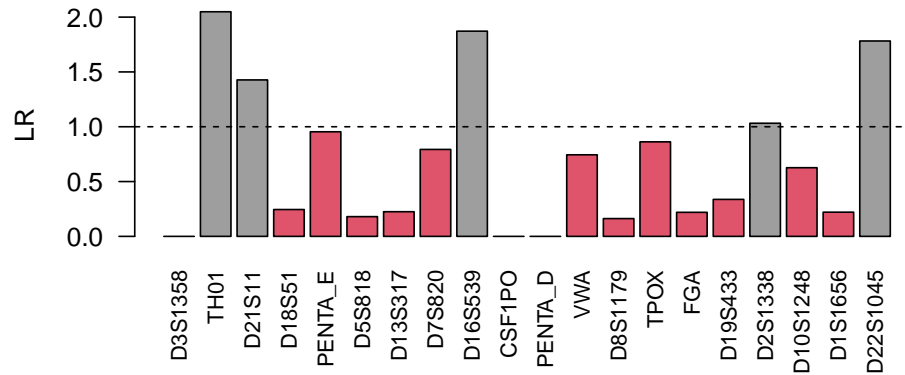
b) Three references: the grandmother, an uncle and an aunt, all on the paternal side.

```
ref = mpi$Reference
missingPersonPlot(ref, missing = "MP")
```



c) The commands give $LR = 0$, which excludes POI1 as the missing person (adding the usual caveats).

d) The markers with $LR = 0$ are D3S1358, CSF1PO and PENTA_D, as can be seen in the plot:



e) Using similar code as for POI1, we find:

```
poi2 = mpi$POI2
mpiTest2 = missingPersonLR(ref, missing = "MP", poi = poi2)
```

```
mpiTest2
```

```
##      H1:H2
## 2369.091
```

The LRs for the individual markers are found in the LRperMarker slot of the output. We sort them and round them to 2 decimal places:

```
round(sort(mpiTest2$LRperMarker), 2)
```

```
## D8S1179 D13S317 D7S820 D22S1045 TPOX D16S539 D5S818 D3S1358
## 0.16 0.65 0.79 0.90 1.00 1.06 1.24 1.32
## PENTA_D VWA D21S11 D10S1248 FGA D1S1656 TH01 D19S433
## 1.34 1.35 1.43 1.51 1.64 1.86 2.05 2.22
## PENTA_E D2S1338 CSF1PO D18S51
## 2.33 2.91 3.25 15.22
```

The marker D18S51 gives the largest LR, with LR = 15.22.

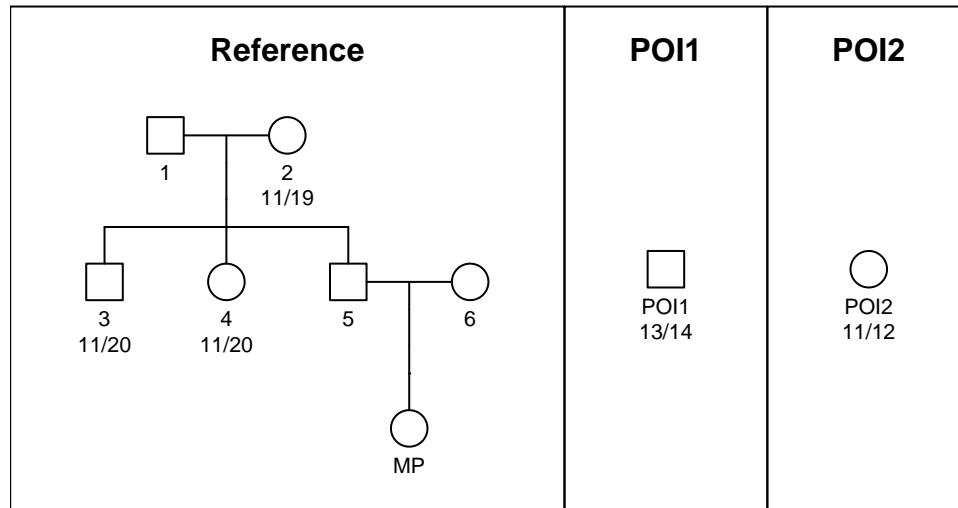
f) As claimed in the exercise, marker D18S51 shows substantially higher LR for POI2 than for POI1:

```
cbind(LR1 = mpiTest1$LRperMarker["D18S51"],
      LR2 = mpiTest2$LRperMarker["D18S51"])
```

```
##      LR1      LR2
## D18S51 0.2449387 15.2168
```

To understand this we inspect the genotypes for the marker:

```
plotPedList(mpi, marker = "D18S51")
```



Furthermore, the allele frequencies of this marker are (rounded to 4 decimal places):

```
afreq(mpi, "D18S51") |> round(4)
```

```
##      9      10      11      12  12.2      13      14      15      16  16.2      17
## 0.0004 0.0124 0.0083 0.1349 0.0001 0.1268 0.1873 0.1425 0.1163 0.0001 0.1009
##      18  18.1  18.3      19      20      21      22      23      24      25      26
## 0.0770 0.0001 0.0001 0.0534 0.0207 0.0075 0.0059 0.0023 0.0014 0.0010 0.0003
##      27
## 0.0003
```

We see that the rare allele “11” (frequency 0.008) carried by POI2 is observed several times in the reference pedigree, while neither of POI1’s alleles occur in the reference. This explains the difference in the LR’s for this particular marker.

g) The exclusion power calculation gives the following output:

```
ep = missingPersonEP(ref, missing = "MP")
ep
```

```
## Potential mismatches: 8 (D3S1358, D7S820, CSF1PO, PENTA_D, VWA, TPOX, D19S433, D2S1338)
## Expected mismatches: 1.679
## P(at least 1 mismatch): 0.863
```

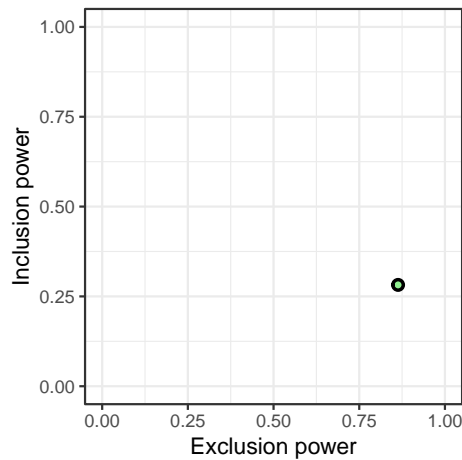
We conclude that there are 8 markers for which an unrelated POI may give exclusion. The overall probability of at least one incompatible marker is $EP = 0.86$, which is normally considered acceptable.

h) We estimate the inclusion power as suggested in the exercises:

```
ip = missingPersonIP(ref, missing = "MP", nsim = 1000, threshold = 10000, seed = 17)
ip
```

```
## Mean LR: 46998.29
## Mean log10(LR): 3.223
## Estimated power:
## P(LR >= 10000) = 0.282
```

The output shows that LR exceeds 10000 with probability 0.282. *Note:* The answer depends to some extent on the seed, but with as many as 1000 simulations the numbers will not change too much.



i)

Comment: In general, plots are more relevant when there are many points, not just one, as here. We recognize the estimates of EP (0.86) and IP (0.282), calculated previously.

Exercise VII-2

a) The plot was produced with the following command:

```
plotDVI(grave, widths = c(2,7), titles = c("PM data", "AM data"))
```

b) The total *a priori* number of solutions is $\text{ncomb}(5,5,3,3) = 52564$. The arguments are, respectively:

- the number of female victims
- the number of female missing persons
- the number of male victims
- the number of male missing persons

c) The parents of MP6 are first cousins, hence she has inbreeding coefficient $f = 1/16$. This can be seen by manually tracing the paths, or in R with the pedsuite functions `verbalise()` and `inbreeding()`:

```
verbalise(grave$am, ids = parents(grave$am, "MP6"))
```

```
## First cousins
## R5-UN2-[GF2,R3]-M0-MP5
```

```
inbreeding(grave$am, "MP6")
```

```
## [1] 0.0625
```

d) The entries $LR_{i,j}$, $i, j = 1, \dots, 8$, of the pairwise LR matrix are the likelihood ratios comparing the assignment $V_i = MP_j$ to the hypothesis of no identification.

```
prw = pairwiseLR(grave)
prw$LRmatrix
```

```

##          MP1          MP2          MP3          MP4          MP5          MP6
## V1 479971259          NA 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
## V2          NA 67760107189          NA          NA          NA          NA
## V3          0          NA 6.409841e+14 0.000000e+00 0.000000e+00 0.000000e+00
## V4          0          NA 0.000000e+00 1.803600e+12 1.803600e+12 0.000000e+00
## V5          0          NA 0.000000e+00 1.030067e+11 1.030067e+11 0.000000e+00
## V6          0          NA 0.000000e+00 0.000000e+00 0.000000e+00 8.817392e+12
## V7          NA          0          NA          NA          NA          NA
## V8          NA          0          NA          NA          NA          NA
##          MP7          MP8
## V1          NA          NA
## V2          0 0.5512209
## V3          NA          NA
## V4          NA          NA
## V5          NA          NA
## V6          NA          NA
## V7 16946051 295.8389523
## V8          0 0.2684890

```

The LR for the pairing $V1 = MP1$ is $4.8e+08$.

e) `excl = findExcluded(grave)`

```

## Finding exclusions
## Max incompatible markers = 2
##
## PM samples excluded against all missing: None
##
## Missing persons excluded against all PM samples: None
##
## Pairings excluded in total: 15

```

`excl$exclusionMatrix`

```

##    MP1 MP2 MP3 MP4 MP5 MP6 MP7 MP8
## V1  0  NA  4  2  2  9  NA  NA
## V2  NA  0  NA  NA  NA  NA  4  0
## V3  4  NA  0  2  2  9  NA  NA
## V4  6  NA  4  0  0  5  NA  NA
## V5  6  NA  4  0  0  7  NA  NA
## V6 10  NA  5  1  1  0  NA  NA
## V7  NA 12  NA  NA  NA  NA  0  0
## V8  NA  9  NA  NA  NA  NA  2  0

```

The output gives the number of exclusions, i.e., the number of incompatible markers, for each assignment $V_i = MP_j$. NA indicates that no comparison is done because of sex incompatibility (set `ignoreSex = TRUE` to count exclusions in these cases too).

f) `und = findUndisputed(grave)`

```

## Finding undisputed matches
## Pairwise LR threshold = 10000
##
## Step 1:
## Computing matrix of pairwise LR
## 5 undisputed matches
## V1 = MP1 (LR = 4.8e+08)
## V2 = MP2 (LR = 6.78e+10)
## V3 = MP3 (LR = 6.41e+14)
## V6 = MP6 (LR = 8.82e+12)
## V7 = MP7 (LR = 1.69e+07)
## Reducing DVI dataset
##
## Step 2:
## Computing matrix of pairwise LR
## No further undisputed matches

```

```
und$undisputed
```

```
## NULL
```

This are the *undisputed* identifications in the dataset. By default, a pairing $V_i = MP_j$ is considered to be undisputed if the pairwise $LR_{i,j} \geq 10000$ and all other entries in the same row/column of the LR matrix are ≤ 1 . See `?findUndisputed` for details and further options.

```
g) joint = jointDVI(grave)
```

```

## DVI dataset:
## 8 victims (3M/5F): V1, V2, V3, V4, V5, V6, V7, V8
## 8 missing (3M/5F): MP1, MP2, MP3, MP4, MP5, MP6, MP7, MP8
## 5 typed refs: R3, R5, R4, R1, R2
## 1 ref family: F1
## Number of markers, PM and AM: 23
## Checking DVI dataset consistency
## No problems found
## Finding undisputed matches
## Pairwise LR threshold = 10000
##
## Step 1:
## Computing matrix of pairwise LR
## 5 undisputed matches
## V1 = MP1 (LR = 4.8e+08)
## V2 = MP2 (LR = 6.78e+10)
## V3 = MP3 (LR = 6.41e+14)
## V6 = MP6 (LR = 8.82e+12)
## V7 = MP7 (LR = 1.69e+07)
## Reducing DVI dataset
##
## Step 2:
## Computing matrix of pairwise LR
## No further undisputed matches
##
## Reduced DVI dataset:

```

```
## 3 victims (1M/2F): V4, V5, V8
## 3 missing (1M/2F): MP4, MP5, MP8
## 10 typed refs: R3, R5, R4, R1, R2, MP3, MP1, MP2, MP7, MP6
## 1 ref family: F1
## Number of markers, PM and AM: 23
##
## Calculating pairing combinations
## Assignments to consider in the joint analysis: 10
##
## Time used: 3.09 secs
```

```
joint[1:5, ]
```

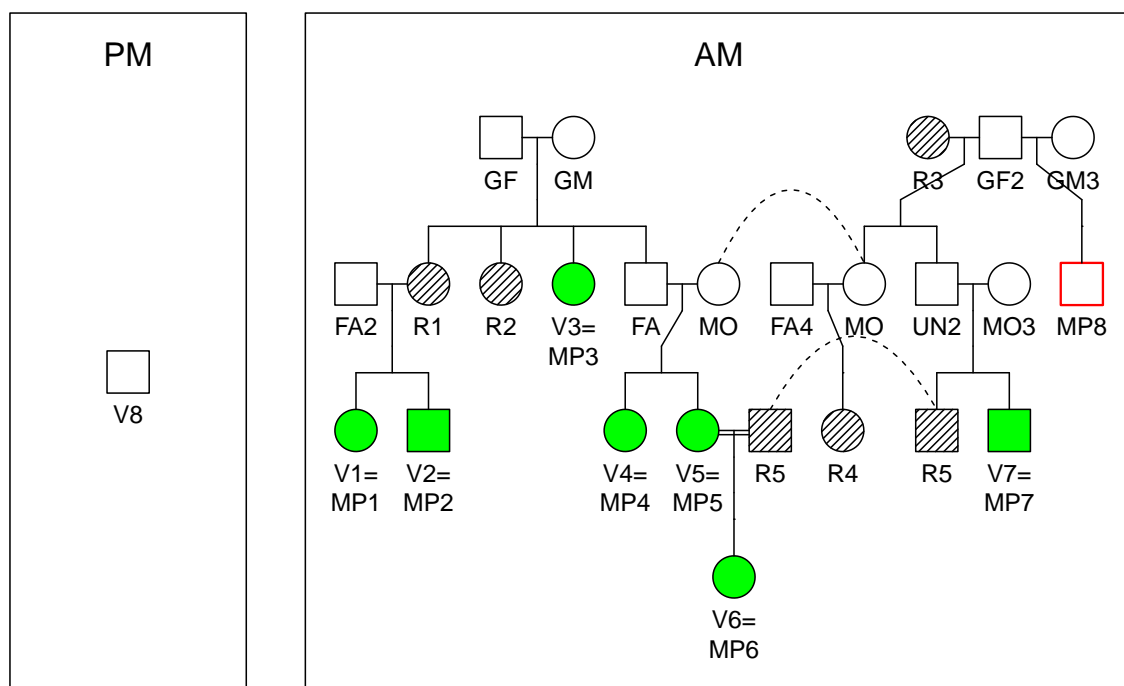
```
##   V1 V2 V3 V4 V5 V6 V7 V8   loglik           LR   posterior
## 1 MP1 MP2 MP3 MP4 MP5 MP6 MP7 * -737.0038 1.626550e+90 6.904734e-01
## 2 MP1 MP2 MP3 MP4 MP5 MP6 MP7 MP8 -737.8061 7.291528e+89 3.095266e-01
## 3 MP1 MP2 MP3 * MP5 MP6 MP7 MP8 -773.8441 1.628294e+74 6.912139e-17
## 4 MP1 MP2 MP3 * MP5 MP6 MP7 * -774.8047 6.230422e+73 2.644825e-17
## 5 MP1 MP2 MP3 MP4 * MP6 MP7 MP8 -782.6972 2.327428e+70 9.879972e-21
```

Comment: The optimal solution is $V_i = M_i$, for $i = 1, \dots, 7$ and no match for V_8 . The LR comparing this solution to the one where $V_8 = MP_8$ is $13.74/6.160 = 2.2$. Hence, it is not possible to reliably distinguish these solutions. This is not unexpected as there is scarce evidence in the pedigree for identifying MP_8 . The remaining solutions, from number 3 and onwards, are much less likely. In fact, the posterior for the two top solutions, calculated using a flat prior for each of the 52564 possible assignments, i.e., $1/52674$, is close to 1.

h) Here is a suitable command.

Note: The `format` argument is not necessary; I used it here just to avoid overlapping labels.

```
plotSolution(grave, joint, format = "[S]=\n[M]")
```



Exercise VII-3

- a) There are no problems with the data, no nonidentifiable missing persons, and no exclusions.
- b) The identifications $MP4 = V4$ and $MP5 = V5$ do not reach the threshold required for undisputed identification. However, $MP4$ and $MP5$ are full sisters, and this enables identification based on a joint analysis.
- c) We can conclude in favor of $H_1: \{MP1, MP2\} = \{V1, V2\}$, but we cannot tell them apart, i.e., $MP1 = V1$ or $V2$ and $MP2 = V1$ or $V2$. In this case with only one reference family, the AM and PM output is identical. We find $GLR = 5.982177e+13$, the **Conclusion: Symmetric match** with the comment **Full siblings: $\{MP1, MP2\} = \{V1, V2\}$**

Exercise VII-4

- a) No
- b) The LR in favor of $V7$ and $V8$ being siblings against unrelated is 562709.
- c) Output from Familias:

DVI module - Results

Summary				
Project name is: Untitled			Number of PM samples: 8	
Number of matches: 3			Number of AM families: 5	
Family id	Unide...	Prior	Posterior	LR
F2	V6	0.009901	0.999914	1079808.9
F3	V2	0.009901	0.998633	67917.208
F5	V4	0.009901	0.999997	31898442

Figure 1: Result of search in planecrash exercise