FORCE Interlab study – Part 2

The purpose of this exercise is to show how FORCE SNP data can be used for various applications such as kinship testing, DVI scenario, ancestry/phenotype predictions. A few case scenarios are given below. All participants will be given the same tasks, same DNA data, same reference data, and based on these perform interpretations.

Any suitable software/tool may be used to solve the issues. It is up to each participant to decide. Those who have no preferences or have previous experience can, however, use *FamLink*. Daniel Kling, the developer of *FamLink*, has put a lot of work into updating *FamLink* to handle large SNP panels like the FORCE. A number of new features/tools have also been implemented. *FamLink* can be downloaded at http://famlink.se/f_download.html (please use version 2.3 or higher). A getting started guide for *FamLink* (version 2.3 and above) is available through https://famlink.se/getting_started_famlink2.pdf.

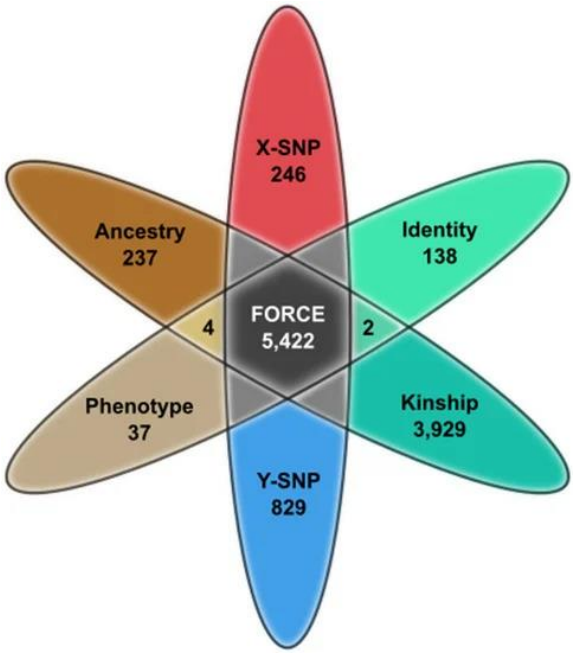
Genotype and reference data ("FORCE_Interlab_Part_2.zip") can be found at https://github.com/atillmar/FORCE. We would like to send a big thank you to Chris Phillips and María de la Puente who compiled the allele frequency data!

Solutions, matching statistics and conclusions should be reported using the "Report_FORCE_Interlab_Part2_template.xlsx" form, and sent to andreas.tillmar@rmv.se before ***December 15.***

If you have any questions or comments, just send an email to andreas.tillmar@rmv.se

Good luck!

Andreas, Kim & Charla



Part 2:1 – Kinship testing

Background: There is a legal interest in establishing the claimed biological relationship between individual A and individual B.

Question: Are A and B related as first cousin once removed (see Figure 1)?

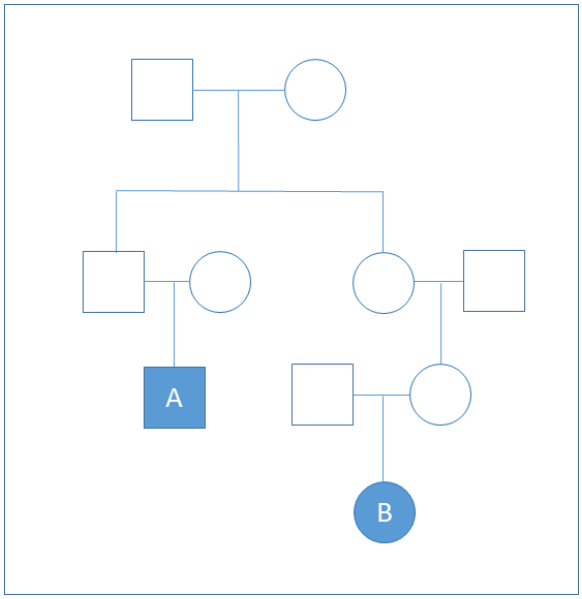


Figure 1. Pedigree for H1.

DNA data: 3930 autosomal kinship SNPs.

Task: Calculate the likelihood ratio (LR) for the hypotheses: H1: "A and B are related as first cousin once removed" and H2: "A and B are unrelated".

Reference marker data: European allele frequencies and the genetic positions of the SNPs given in the files "FORCE_EUR_221006.freq" and "FORCE_221006.map", respectively. Assume linkage equilibrium and ignore complicating factors such as population substructure and mutations.

Report: LR and conclusion

(Hint: Use FamLink -> Select pedigree -> ….->Results)

Part 2:2 – DVI

Background: Four bodies have been found in an old mine. It is believed that these bodies come from a group of five young individuals that disappeared in the 1960s.

Task: Establish the identity of the four found bodies (PM1, PM2, PM3 and PM4). Reference data (AM1, AM2, AM3, AM4 and AM5) is available for the five missing individuals (Family 1-Family 5, see information below in Figure 2 and Table 1). Full or partial SNP profiles (PM1, PM2, PM3 and PM4) are available from the analyses of the human remains.

Reference marker data: European allele frequencies and the genetic positions of the SNPs given in the files "FORCE_EUR_221006.freq" and "FORCE_221006.map", respectively. Assume linkage equilibrium and ignore complicating factors such as population substructure and mutations.

Report: Overall conclusion and individual matching statistics (LRs and/or posterior probabilities).

(Hint: FamLink->Blind search… or FamLink->Select pedigree….->DVI search)

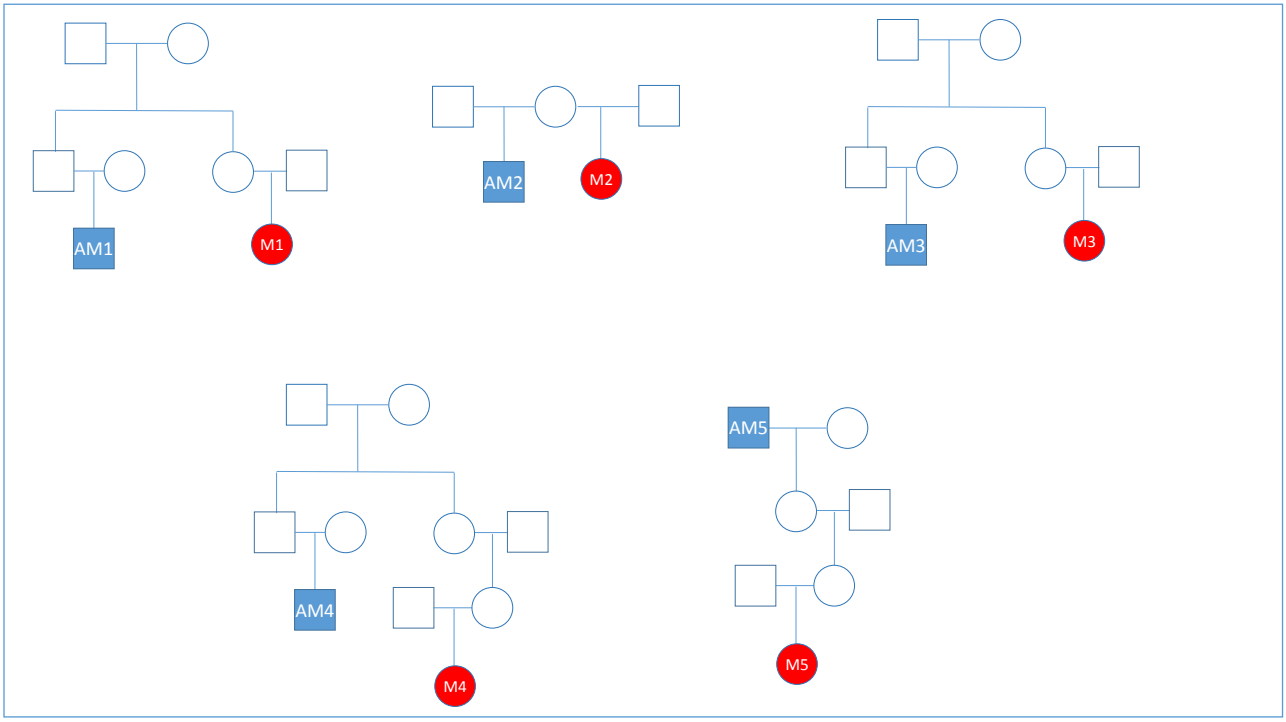


Figure 2. Pedigrees of the reference families and relationships between AM individuals and the missing.

Table 1. Information about the reference families and relationship to the missing

Reference Family	Reference individual	Relationship to the missing
1	AM1	First cousin
2	AM2	Half sibling
3	AM3	First cousin
4	AM4	First cousin once removed
5	AM5	Maternal great-grandfather

Part 2:3 – Ancestry/phenotypic prediction

Background: Two crime scene samples (Sample C and Sample D) have been collected in a triple murder case. It is important to establish the identity of the unknown donors of these samples.

Task: Perform biogeographical ancestry predictions, Y-haplogroup prediction and phenotypic (eye, hair and skin color) predictions based on the SNP data for sample C. Perform biogeographical ancestry prediction and phenotypic (eye, hair and skin color) prediction based on the SNP data for sample D sample D.

Reference marker data: AIM autosomal SNP frequencies for African, American, East Asian, European, South Asian, Middle East and Oceanic metapopulations and the genetic positions of the SNPs are given in the files connected to this exercise. Reference data for the Y SNPs and their haplogroup definitions can be found here: https://raw.githubusercontent.com/genid/Yleaf/master/yleaf/data/hg38/new_positions.txt. This information has been implemented in *FamLink*

Report: Predictions and associated statistics/probabilities.

(Hint: FamLink -> Phenotype/Ancestry/Haplogroup)