

# Y-profile evidence: Close paternal relatives and mixtures

Sep 13, 2019

Mikkel M Andersen\* and David J Balding

\* [mikl@math.aau.dk](mailto:mikl@math.aau.dk)

The 28th Congress of the International Society for Forensic Genetics  
Prague, the Czech Republic



AALBORG UNIVERSITY  
DENMARK

# Y profiles are rare



**All (modern) Y profiles are rare and shared by close relatives only.**



## Y profiles are rare

**All (modern) Y profiles are rare and shared by close relatives only.**

**Very likely that at most 40 men share a YfilerPlus profile and they are very likely only separated by up to 20 meioses.**



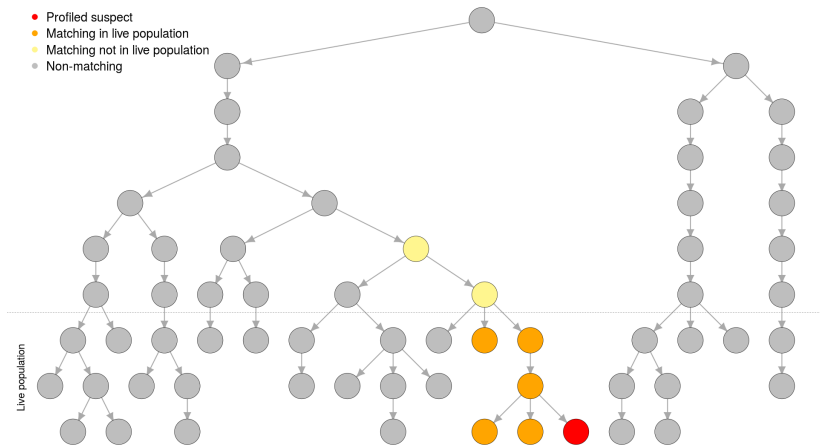
## Y profiles are rare

**All (modern) Y profiles are rare and shared by close relatives only.**

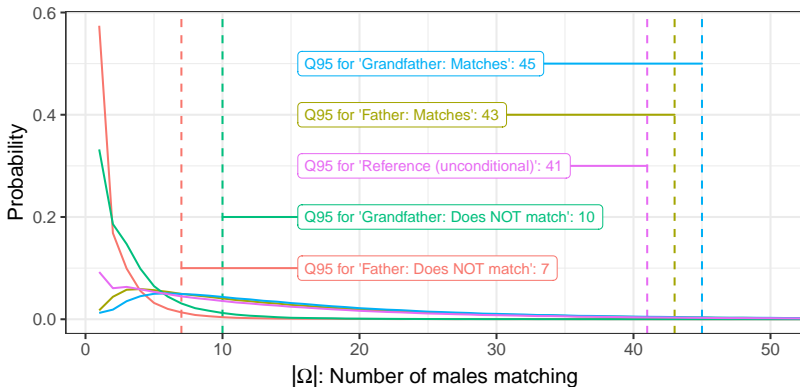
**Very likely that at most 40 men share a YfilerPlus profile and they are very likely only separated by up to 20 meioses.**

All results are for YfilerPlus.

# Simulation model

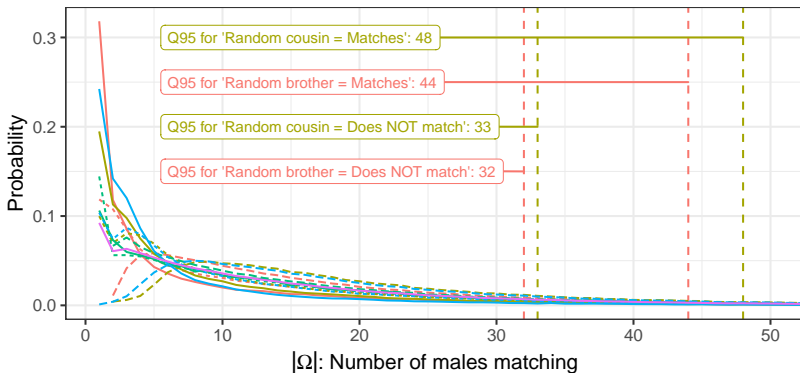


# Close paternal relatives



- Father: Does NOT match
- Grandfather: Matches
- Father: Matches
- Reference (unconditional)
- Grandfather: Does NOT match

# Close paternal relatives



- Does NOT match
- ..... Has none
- - - Matches
- Random brother
- Random cousin
- Random uncle
- Reference (unconditional)
- Random son



# Reporting weight of evidence

- ▶ Match probability
  - ▶ No. of meioses between suspect and members of suspect/reference population
  - ▶ Suspect population specified: e.g. birth years, genealogy, ...
  - ▶ Recently discussed for Y profiles by Caliebe and Krawczak (2018)
- ▶ Very likely that at most 40 men share a YfilerPlus profile
  - ▶ Previously used for autosomal profiles; different from the *LR* reporting
  - ▶ Additional information (on close paternal relatives): matching/non-matching, living elsewhere, ...



## Example with $LR$ 's



Suspect population size  $N$ :  $LR \approx 1/(n/N) = N/n$ .



## Example with $LR$ 's

Suspect population size  $N$ :  $LR \approx 1/(n/N) = N/n$ .

Nothing known about suspect's paternal pedigree:

$N$		$LR$
2,000,000	$2,000,000/40 =$	50,000
200,000	$200,000/40 =$	5,000
20,000	$20,000/40 =$	500



## Example with $LR$ 's

Suspect population size  $N$ :  $LR \approx 1/(n/N) = N/n$ .

Nothing known about suspect's paternal pedigree:

$N$	$LR$
2,000,000	$2,000,000/40 = 50,000$
200,000	$200,000/40 = 5,000$
20,000	$20,000/40 = 500$

Suspect and his father have different profiles:

$N$	$LR$
2,000,000	$2,000,000/10 = 200,000$
200,000	$200,000/10 = 20,000$
20,000	$20,000/10 = 2,000$



## Example with $LR$ 's

Suspect population size  $N$ :  $LR \approx 1/(n/N) = N/n$ .

Nothing known about suspect's paternal pedigree:

$N$	$LR$
2,000,000	$2,000,000/40 = 50,000$
200,000	$200,000/40 = 5,000$
20,000	$20,000/40 = 500$

Suspect and his father have different profiles:

$N$	$LR$
2,000,000	$2,000,000/10 = 200,000$
200,000	$200,000/10 = 20,000$
20,000	$20,000/10 = 2,000$

Suspect has no brothers; has two uncles (one matching); ...



# MIXTURES



# Mixtures

*A person is upper-case; the corresponding person's profile is lower-case.*

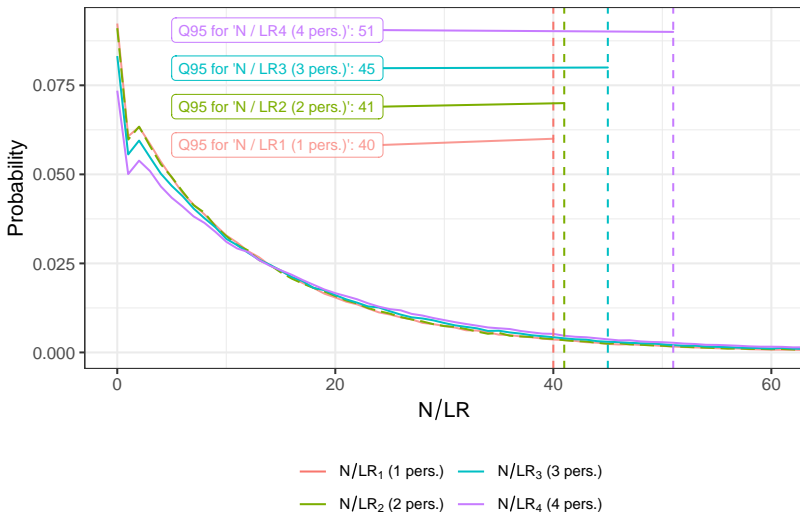
- ▶  $m$ : Observed mixture
- ▶  $Q$ : Queried contributor (with profile  $q$ )
- ▶ Hypotheses
  - ▶  $H_p$ :  $Q + U$  for unknown  $U$  s.t.  $m = q + u$
  - ▶  $H_d$ :  $R + S$  for any pair unknowns  $R$  and  $S$  such that  $m = r + s$

$$\text{LR}_2 = \frac{P(m \mid H_p)}{P(m \mid H_d)} = \frac{n_u/N}{\sum_{(r,s)=m} (n_r/N)(n_s/N)} = \frac{Nn_u}{\sum_{(r,s)=m} n_r n_s}$$

$$N/\text{LR}_2 = \frac{\sum_{(r,s)=m} n_r n_s}{n_u}$$

for population size  $N$  and  $n_a$  the population count of profile  $a$ .

# Mixtures





# SIMULATED VS REAL POPULATIONS





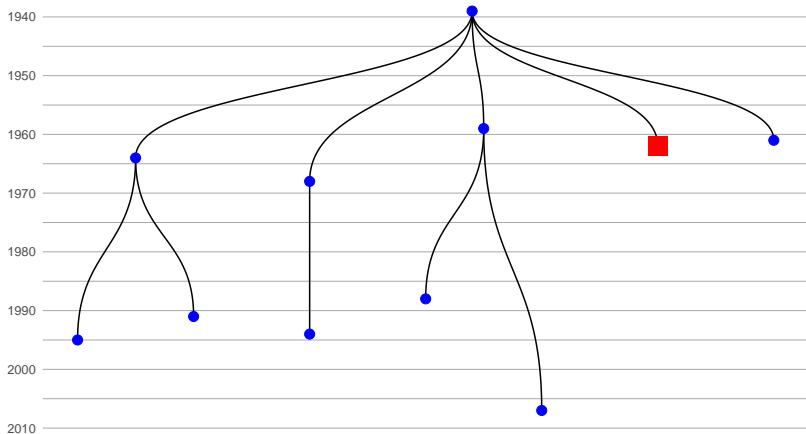
# Population simulations vs real populations

Danish data (retrived in 2017):

- ▶ 5,052,681 males
- ▶ Birthyear range: 1858-2017
- ▶ Unfortunately:
  - ▶ Individuals born 1969 and later: complete data (e.g. father, mother, ...)
  - ▶ Individuals born 1968 and before: most probably incomplete
  - ▶ Many small pedigrees (largest of size 37 males, many much smaller)
- ▶ Merge pedigrees by adding surrogate ancestors
  - ▶ Future: more different ways

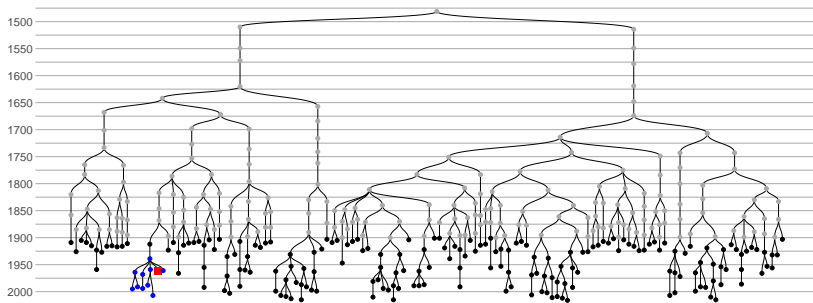
# Observed pedigree

● Pedigree member    ■ Person of interest (PoI)



# *In silico* merge to larger genealogy

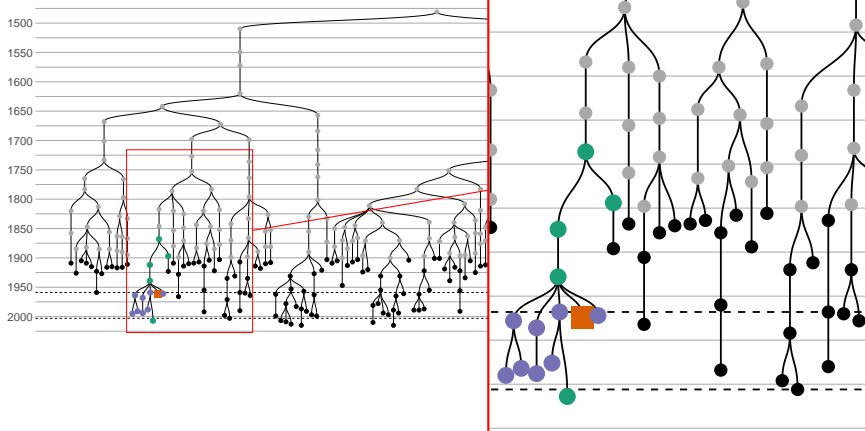
- Pedigree member
- Real male
- Person of interest (PoI)
- Surrogate male



# Randomly assign YfilerPlus haplotypes

Matching in total: 14  
 Matching in live population (1959 – 2003): 9

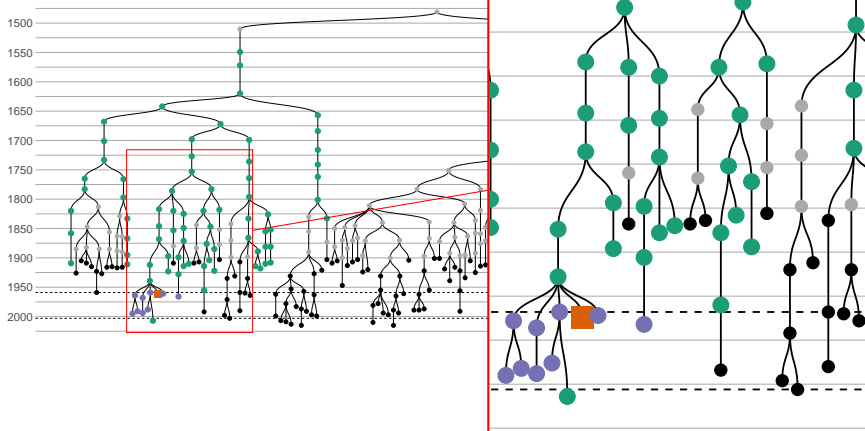
- Matching (in live population)
- Matching (not live)
- Person of interest (PoI)
- Real male
- Surrogate male



# Randomly assign YfilerPlus haplotypes

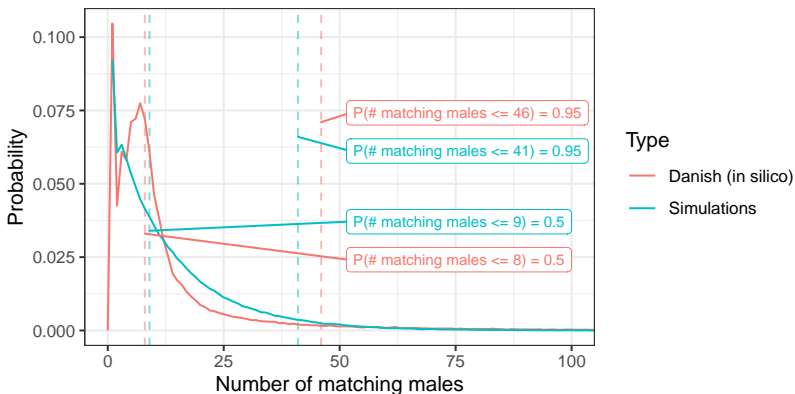
Matching in total: 83  
 Matching in live population (1959 – 2003): 10

- Matching (in live population)
- Matching (not live)
- Person of interest (PoI)
- Real male
- Surrogate male



# How realistic are population simulations?

Doing this for many pedigrees, many merges and many mutation process (a total of 125,000 realisations):





# Thank you for your attention!

## Thank you for your attention!

References:

- ▶ “How convincing is a matching Y-chromosome profile?” (DOI: 10.1371/journal.pgen.1007028)
- ▶ “Y-profile evidence: close paternal relatives and mixtures” (DOI: 10.1016/j.fsigen.2018.10.004)
- ▶ Software: <https://github.com/mikldk/malan>

Advertisements for posters:

- ▶ P316: “*Validation of a population simulation model for the estimation of Y-haplotype frequencies in forensic cases using a large French-Canadian dataset*” (R Landry, MM Andersen, E Milot)
- ▶ P512: “*Isoallelic frequency estimation for STR markers from massive parallel sequencing data*” (MM Andersen and T Tvedebrink)