X-chromosomal markers in Forensic Genetics

GHEP 2025 Virtual workshop series.

March ~~10,17~~ and **24th**

Daniel Kling and Andreas Tillmar

Teachers

Daniel Kling. PhD

- Forensic Expert
- National Board of Forensic Medicine. Sweden
- Worked in the field for 15 years
- Developer of Familias. FamLink and FamLinkX
- Applied biostatistics. relationship inference. genetic genealogy

Andreas Tillmar. PhD

- Forensic geneticist & Associate professor
- National Board of Forensic Medicine. Sweden and Linköping University. Sweden
- Worked in the field for over 15 years
- Technical leadership mixed with R&D
- Applied biostatistics. relationship inference. population genetics. genetic genealogy.
- Lead author of the ISFG Commission on X-chromosomal testing

Session 3 – Applications and examples (March 24)

16:00	Introduction
16:15-17:00	Summary of theory and some more advanced topics
17:00-17:10	Short break
17:10-18:00	Examples
18:05-18:40	Exercises
18:40-19:00	Summary

Presentations. exercises etc are available at https://familias.name/GHEP2025/

Write your questions in the chat-function. and we will try to answer direct! (or save it to the end of the day)

Session 1 – Basics (March 10)

16:00	Introduction
16:15-17:00	Basics of kinship testing and the utility of X-chromosomal markers
17:00-17:10	Short break
17:10-18:00	Software: FamLinkX
18:05-18:40	Exercises
18:40-19:00	Summary

Session 2 – Advanced (March 17)

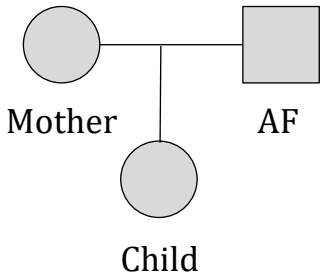
16:00	Introduction
16:15-17:00	Advanced theory
17:00-17:10	Short break
17:10-18:00	Haplotypes and databases
18:05-18:40	Exercises
18:40-19:00	Summary

Solving relationship issues with DNA data

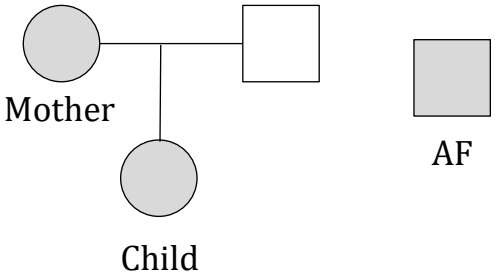
Legal situations: (e.g.) paternity. immigration. missing person identification. criminal acts (incest. human trafficking). investigative leads and more

Example 1 ”Simple” question

H_1: AF is the father of the child



H_2: AF is unrelated to the child

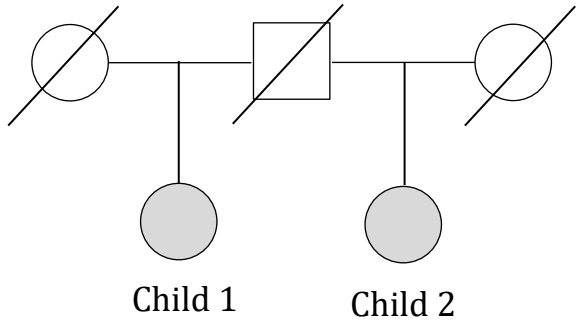


Question: Is AF the biological father of the child?

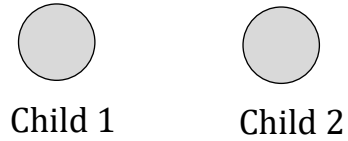
Genetic data: 15-21 autosomal STRs

Example 2 ”More complex” question

H_1: Child 1 and 2 have the same father



H_2: Child 1 and 2 are unrelated

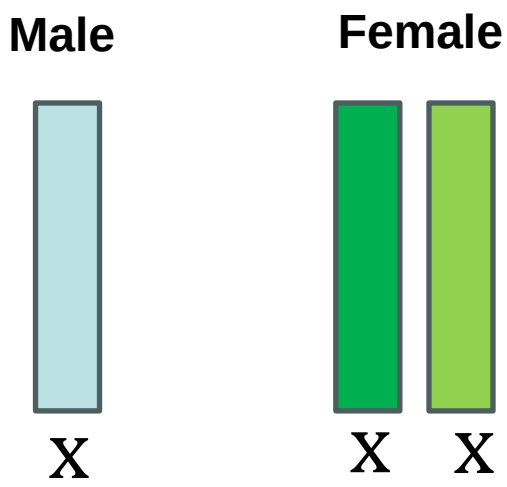


Question: Are child 1 and child 2 paternal half-sibs. or unrelated

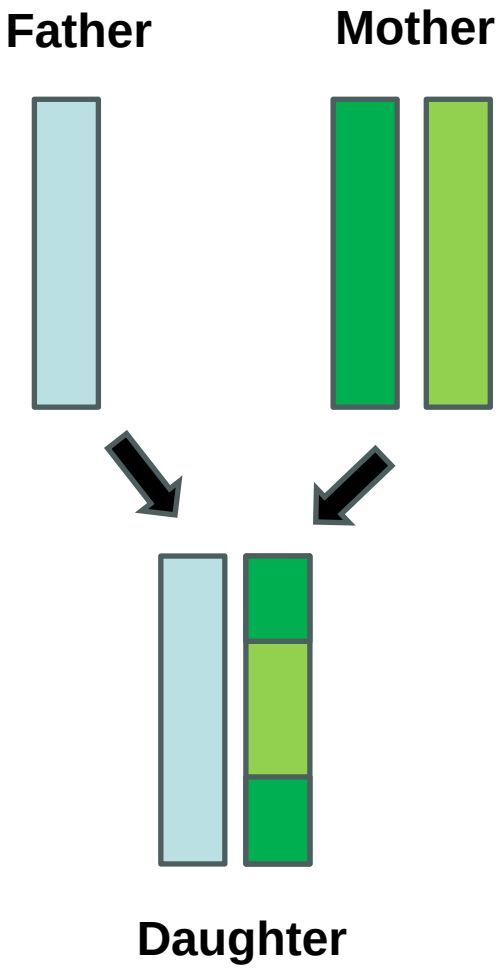
Genetic data: 4-12 X-STRs

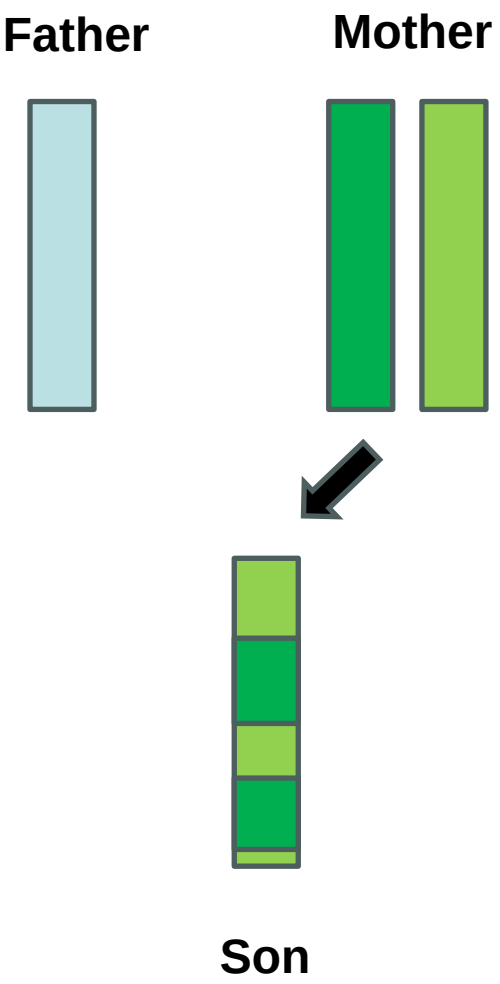
X chromosome in humans

- A female has two X chromosomes
- A male has one X chromosome
- In rare occasions other variations may exist.
  XXY (Klinefelter). X0 (Turner). XXX (Triple X). XYY



X-chromosomal inheritance pattern

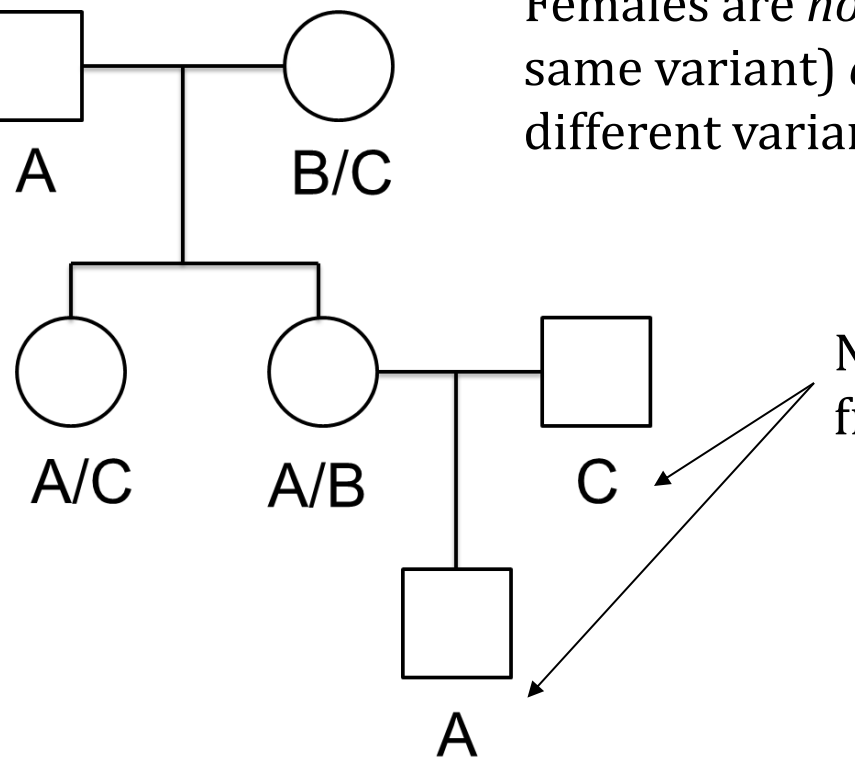




Inheritance pattern (one X locus)

Males are *hemizygous*

Females are *homozygous* (two alleles of the same variant) *or heterozygous* (two alleles of different variants)



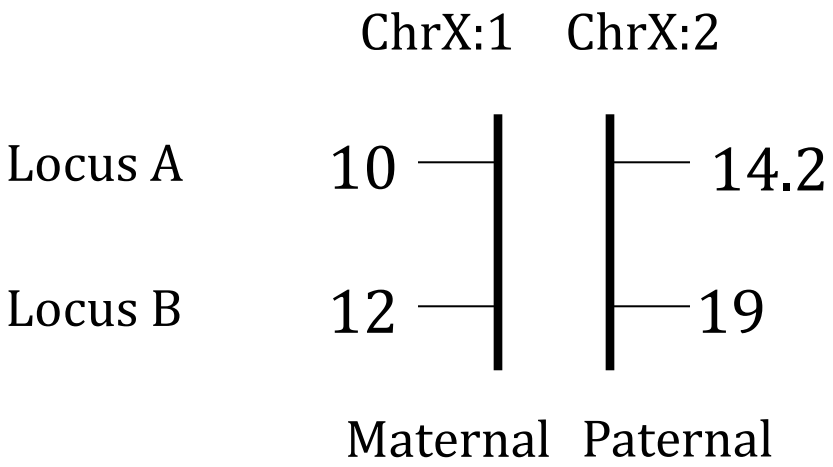
Forced transmission from father to daughter

No transmission from father to son

Basic notations:

Allele. haplotype. genotype. diplotype

Female

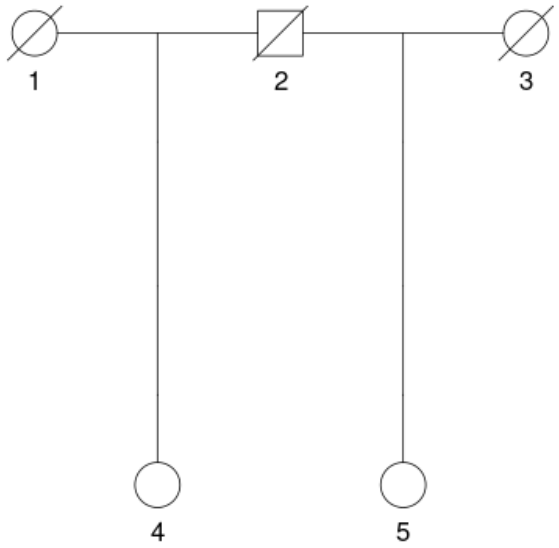


- 10 is an *allele*
- 10/14.2 (or 10.14.2) is a *genotype*
- 10_12 is a *haplotype*
- 10_12/14.2_19 is a *diplotype*
  (or 10_12|14.2_19)
  (or 10|14.2
  12|19)

Inheritance pattern makes X-chromosomal analysis more (or less) informative compared with autosomal DNA analysis

Generally more informative

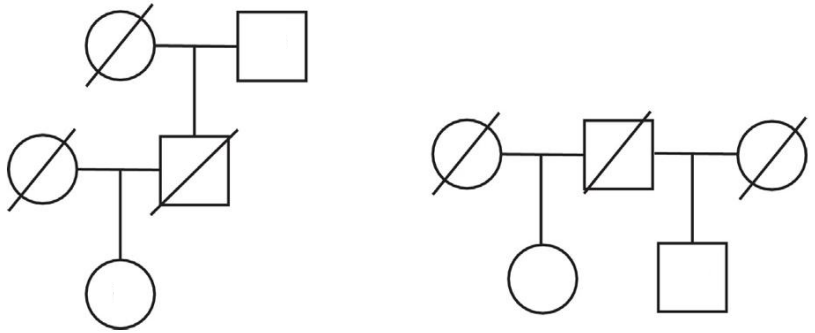
- Paternal half-sisters vs unrelated
- Paternal grandmother/granddaughter vs unrelated
- For many pedigress. the exclusion probability is not null



See Pinto et al.. 2011

Generally less informative

- Father/son vs unrelated
- Paternal grandfather/grandson vs unrelated
- Paternal halfbrothers vs unrelated

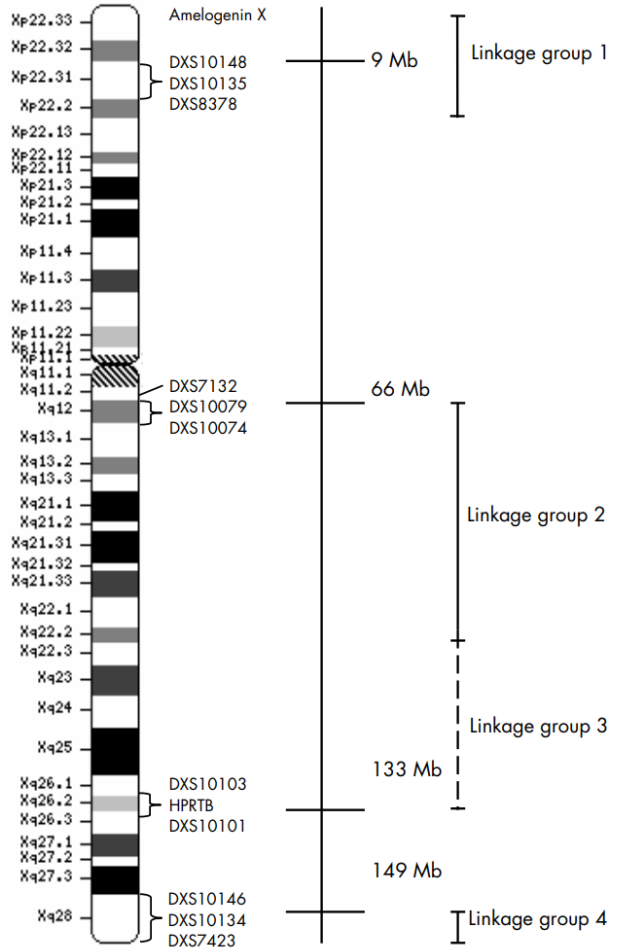


Tillmar et al.. 2017

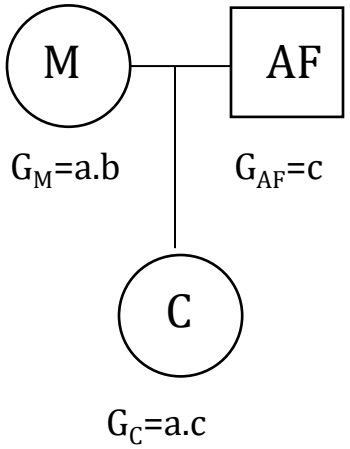
Two common X-chromosomal marker panels

- STRs (short tandem repeats)
- **"X-Decaplex"**
  - 10 X STRs. in genetic linkage but mostly not in linkage disequilibrium (LD. allelic association).
  - Developed by GEP-ISFG (Gusmao et al.. 2009)

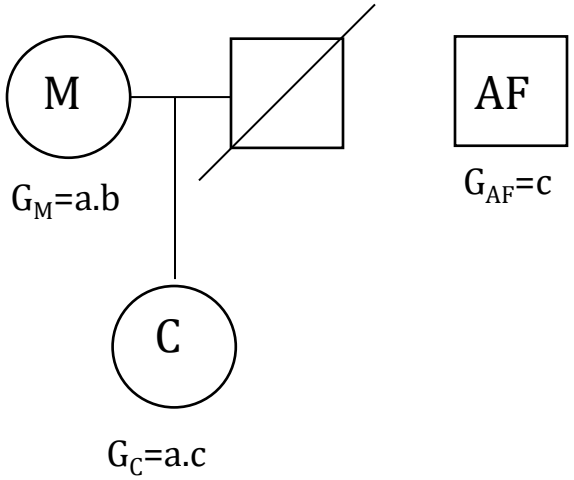
- **Argus X-12QS**
  - 12 X STRs. in four "linkage groups". in genetic linkage but mostly not in linkage disequilibrium (LD. allelic association).
  - Investigator Argus X-12 QS (Qiagen)



Which hypothesis is best supported by observed DNA profiles?



$\Pr(DNA \mid H_1)$

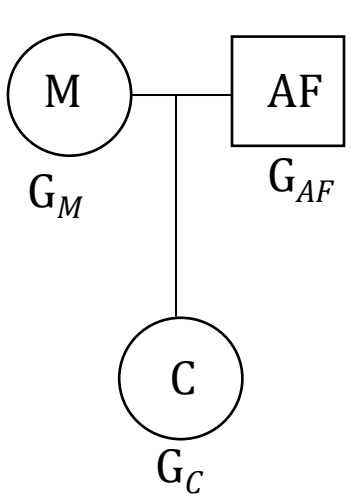


$\Pr(DNA \mid H_2)$

$$\text{LR} = \frac{\Pr(DNA \mid H_1)}{\Pr(DNA \mid H_2)}$$

$$LR = \frac{\Pr(DNA \mid H_1)}{\Pr(DNA \mid H_2)}$$

$$\Pr(DNA \mid H_1)$$



Founders

Child

Genotype probabilities

- Allele frequencies
- Haplotype frequencies

Transmission probabilities

- Mendelian factors (0.5. 1)
- Recombination frequency
- Mutation rate

Genotype/diplotype. allele/haplotype frequencies

- By applying Hardy-Weinberg formulas. we can obtain the needed genotype/diplotype frequencies from allele/haplotype frequencies (assuming HW equilibrium).

Allele:

$$p_i \approx \frac{x_i + 1}{N + 1}$$

p_i The probability to observe allele *i* in the population

x_i Count of allele *i*

N Total number of observed alleles in the population database

Will be covered in Daniel's presentation

Haplotype:

$$p_i \approx \frac{x_i + \lambda\pi_i}{N + \lambda}$$

p_i The probability to observe haplotype *i* in the population

x_i Observed count of haplotype *i*

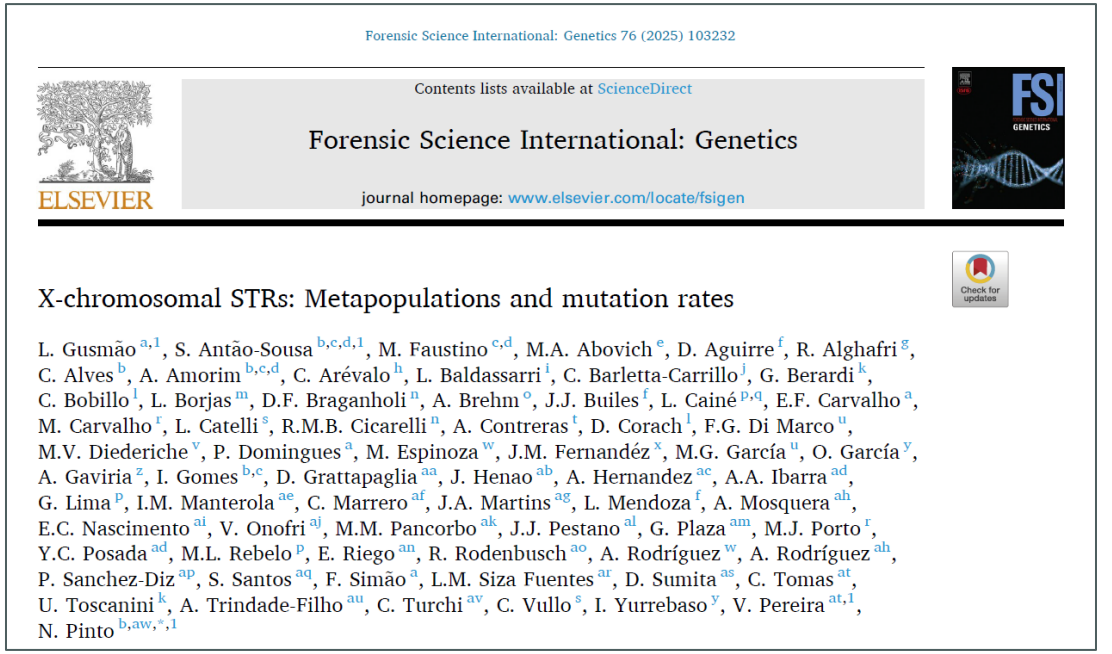
N Total number of observed haplotypes in the population database

π_i Prior probability of haplotype *i* (estimated from allele frequencies)

λ Lambda. the weight given to the prior probability

Allele/haplotype frequencies

- Gusmao et al.. 2025



Forensic Science International: Genetics 76 (2025) 103232

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

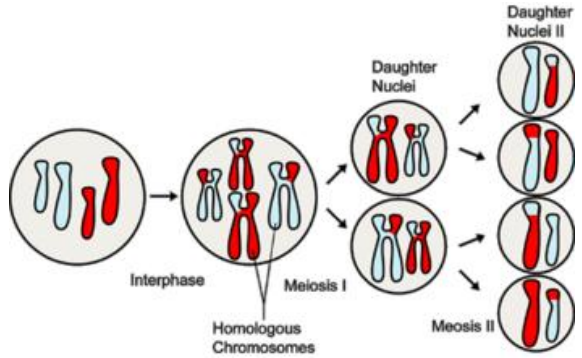
X-chromosomal STRs: Metapopulations and mutation rates

L. Gusmão [a,1], S. Antão-Sousa [b,c,d,1], M. Faustino [c,d], M.A. Abovich [e], D. Aguirre [f], R. Alghafri [g], C. Alves [b], A. Amorim [b,c,d], C. Arévalo [h], L. Baldassarri [i], C. Barletta-Carrillo [j], G. Berardi [k], C. Bobillo [l], L. Borjas [m], D.F. Braganholi [n], A. Brehm [o], J.J. Builes [f], L. Cainé [p,q], E.F. Carvalho [a], M. Carvalho [r], L. Catelli [s], R.M.B. Cicarelli [n], A. Contreras [t], D. Corach [l], F.G. Di Marco [u], M.V. Diederiche [v], P. Domingues [a], M. Espinoza [w], J.M. Fernandéz [x], M.G. García [u], O. García [y], A. Gaviria [z], I. Gomes [b,c], D. Grattapaglia [aa], J. Henao [ab], A. Hernandez [ac], A.A. Ibarra [ad], G. Lima [p], I.M. Manterola [ae], C. Marrero [af], J.A. Martins [ag], L. Mendoza [f], A. Mosquera [ah], E.C. Nascimento [ai], V. Onofri [aj], M.M. Pancorbo [ak], J.J. Pestano [al], G. Plaza [am], M.J. Porto [r], Y.C. Posada [ad], M.L. Rebelo [p], E. Riego [an], R. Rodenbusch [ao], A. Rodríguez [w], A. Rodríguez [ah], P. Sanchez-Diz [ap], S. Santos [aq], F. Simão [a], L.M. Siza Fuentes [ar], D. Sumita [as], C. Tomas [at], U. Toscanini [k], A. Trindade-Filho [au], C. Turchi [av], C. Vullo [s], I. Yurrebaso [y], V. Pereira [at,1], N. Pinto [b,aw,*,1]

- https://famlink.se/fx_databases.html

Linkage and Linkage disequilibrium

- Linkage (or genetic linkage)
  - Can be described as the co-segregation of closely located loci within a family or pedigree.
  - **Effects the transmission probabilities!**

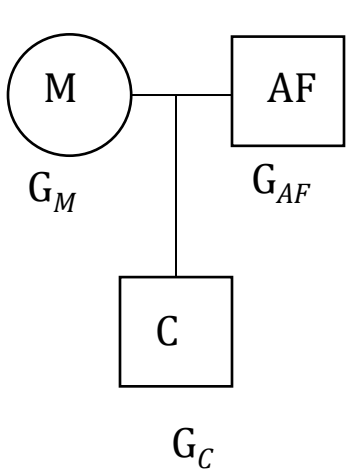


- Linkage disequilibrium (LD)
  - Allelic association.
  - Two alleles (at two different markers) which is observed more often/less often than can be expected.
  - **Effects the founder genotype probabilities. not the transmission probabilities!**
  - Haplotype frequencies rather than allele frequencies must be used

$$LR = \frac{\Pr(DNA \mid H_1)}{\Pr(DNA \mid H_2)}$$

$$\Pr(DNA \mid H_1)$$

Linkage disequilibrium (LD)



Founders

Genotype/diplotype probabilities

Haplotype frequencies (instead of allele frequencies)

Child

Transmission probabilities

Mendelian factor

Recombination frequency/rate

Linkage

Forensic Science International: Genetics 29 (2017) 269–275

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Research Paper

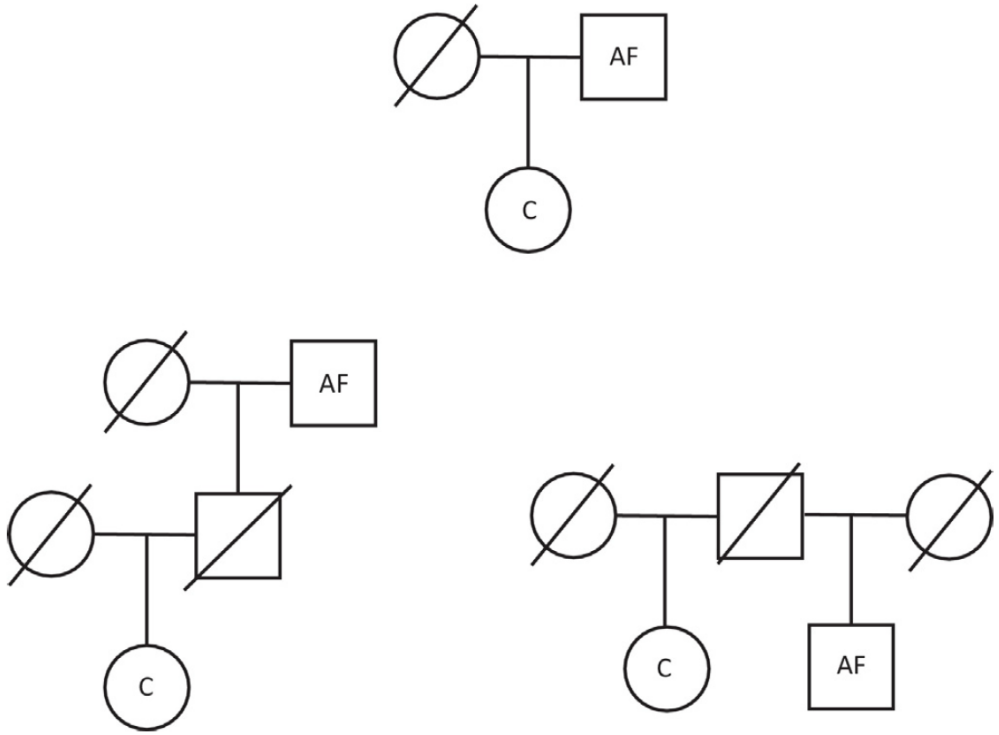
DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis

Andreas O. Tillmar[a,b,*], Daniel Kling[c], John M. Butler[d], Walther Parson[e,f], Mechthild Prinz[g], Peter M. Schneider[h], Thore Egeland[i,1], Leonor Gusmão[j,1]

[a] Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden
[b] Department of Clinical and Experimental Medicine, Faculty of Health Sciences, Linköping University, Linköping, Sweden
[c] Department of Forensic Services, Oslo University Hospital, Oslo, Norway
[d] National Institute of Standards and Technology, Special Programs Office, Gaithersburg, MD, USA
[e] Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria
[f] Forensic Science Program, The Pennsylvania State University, PA, USA
[g] John Jay College of Criminal Justice, New York, USA
[h] Institute of Legal Medicine, Faculty of Medicine, University of Cologne, Germany
[i] Norwegian University of Life Sciences, Oslo, Norway
[j] State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil

Recommendation #1

In paternity cases (duos or trios, with a daughter), X-STR analysis should be used to supplement DNA testing results when the information obtained from standard autosomal markers is inconclusive, such as may be observed in paternity cases with few genetic inconsistencies.

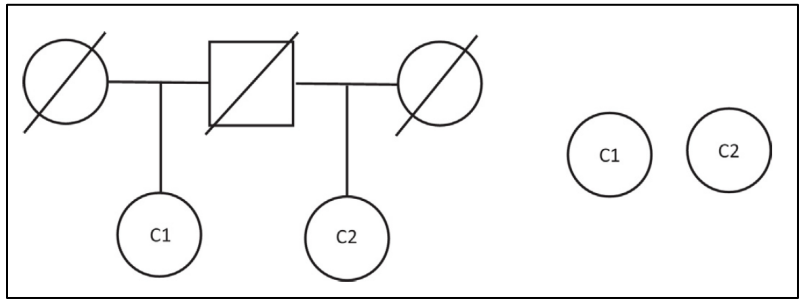


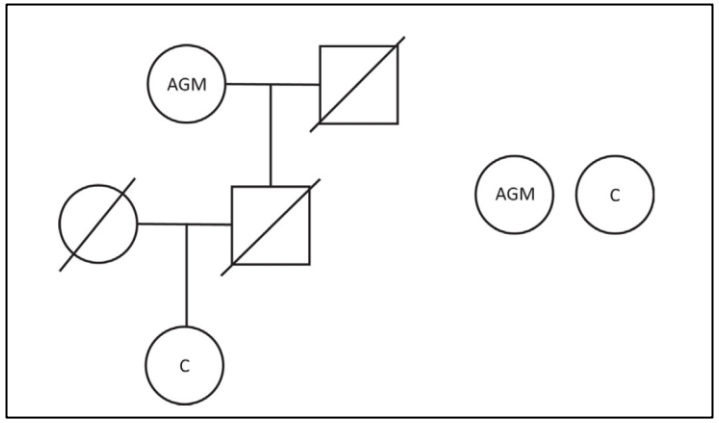
When few inconsistencies exist between the alleged father (AF) and the daughter (C) in a standard paternity duo case (upper), the most likely explanations are either that mutations have occurred or that there is another close relationship between the alleged father and the child.

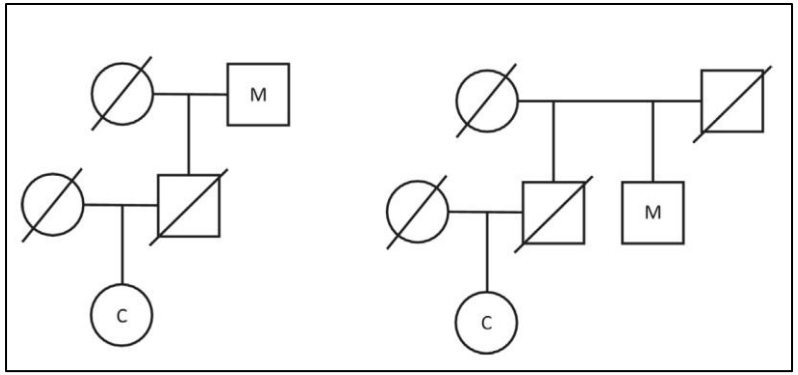
If the true father is the father (with different mothers) (lower right), or son (lower left) of the alleged father, they will not share X-chromosomal alleles identical by descent (IBD) and the X-chromosomal markers can be much more informative than the autosomal markers.

Recommendation #2

X-chromosomal markers should be used in specific kinship cases when the exclusion power does not equal null in contrast to the autosomal markers examined. Important examples include full or paternal half sibling duos involving two females, and paternal grandmother/granddaughter duos. Furthermore, X-chromosomal markers should be used in situations where two alternative hypotheses possess the same likelihood for autosomal markers but are expected to differ when X-chromosomal markers are examined. X-chromosomal analysis may also help to distinguish possible related fathers in incest cases.

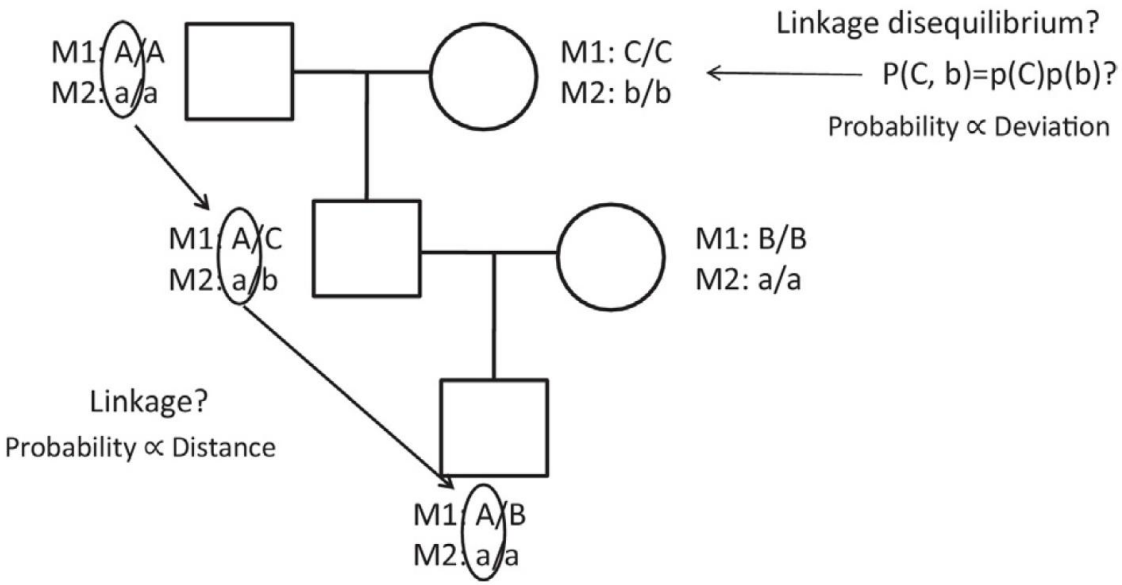






Recommendation #3

Prior to using a X-chromosomal assay or commercial kit, markers should be evaluated to determine whether or not they are linked. Recombination rates should primarily be estimated from family studies or secondarily via mapping functions based on genetic distances. A recombination rate below 0.5 indicates linkage.



Recommendation #4

Linkage should be accounted for when calculating LRs given that the X-chromosomal markers are linked and that linkage will have an impact on the final LR. This also includes accounting for recombination events within a cluster of X-chromosomal markers, known as linkage group.

Box 1.

Consider a maternity case involving a putative mother and a child (male). The data consist of two linked markers (separated with a recombination rate, *r*) on the X chromosome. The putative mother has genotype *a*/*b* at marker 1 and *c*/*d* at marker 2, and the child has genotype *a* at marker 1 and *d* at marker 2. The formula for the LR can then be written as (assuming that *a*, *b*, *c* and *d* are different alleles):

$$LR = \frac{\Pr(DNA|mother\ of\ child)}{\Pr(DNA|unrelated)} = \frac{2 \cdot p_{ac} \cdot p_{bd} \cdot 0.5 \cdot r + 2 \cdot p_{ad} \cdot p_{bc} \cdot 0.5(1-r)}{(2 \cdot p_{ac} \cdot p_{bd} + 2 \cdot p_{ad} \cdot p_{bc})p_{ad}}$$
$$= \frac{p_{ac} \cdot p_{bd} \cdot r + p_{ad} \cdot p_{bc} \cdot (1-r)}{(2 \cdot p_{ac} \cdot p_{bd} + 2 \cdot p_{ad} \cdot p_{bc})p_{ad}}$$

If LE holds (i.e. $p_{xy}=p_x \cdot p_y$) the LR becomes:

$$LR = \frac{2 \cdot p_{ac} \cdot p_{bd} \cdot 0.5 \cdot r + 2 \cdot p_{ad} \cdot p_{bc} \cdot 0.5(1-r)}{(2 \cdot p_{ac} \cdot p_{bd} + 2 \cdot p_{ad} \cdot p_{bc})p_{ad}}$$
$$= \frac{p_a \cdot p_c \cdot p_b \cdot p_d \cdot r + p_a \cdot p_c \cdot p_b \cdot p_d \cdot (1-r)}{(2 \cdot p_a \cdot p_c \cdot p_b \cdot p_d + 2 \cdot p_a \cdot p_c \cdot p_b \cdot p_d)p_a \cdot p_d} = \frac{1}{4 \cdot p_a \cdot p_d}$$

For a maternity case, the recombination rate has no impact on the final LR if LE can be assumed. However if LD exists linkage needs to be accounted for even for a maternity case scenario as derived above.

Recommendation #5

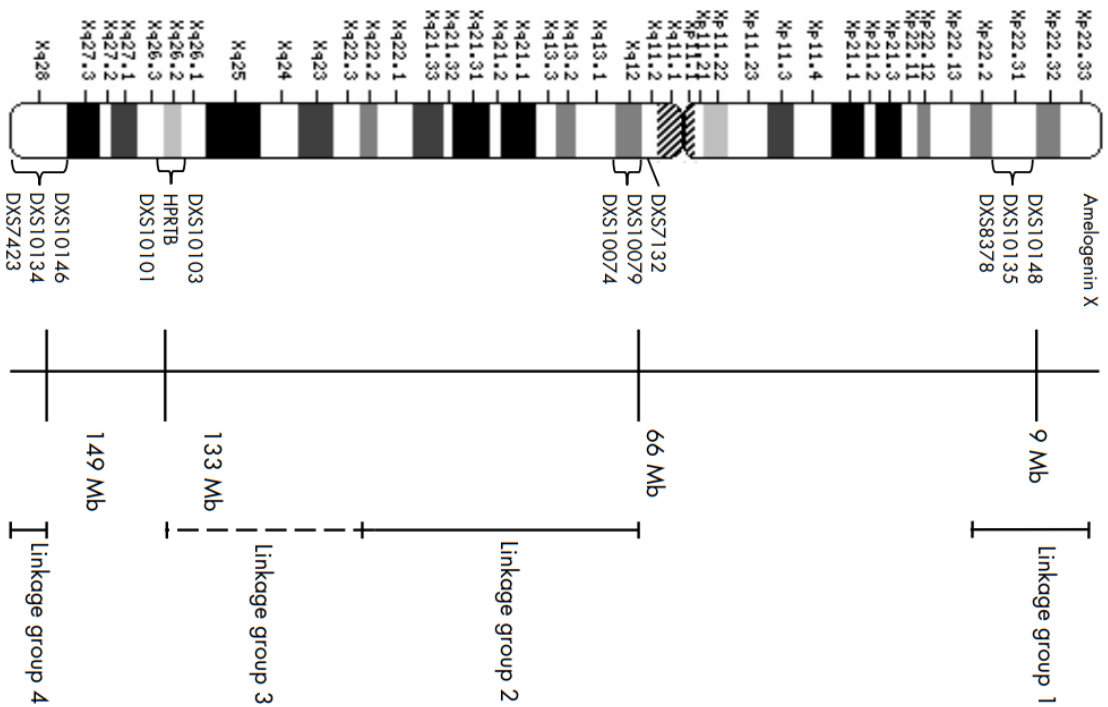
Linkage equilibrium tests should be performed when generating population frequency data for the markers in a X-chromosomal marker multiplex.

Recommendation #6

X-chromosomal markers that are located closely to each other and not in linkage equilibrium should be reported as haplotype frequencies rather than single locus allele frequencies for population databasing.

Recommendation #7

Haplotype frequencies should be used for likelihood calculations when LD exists.



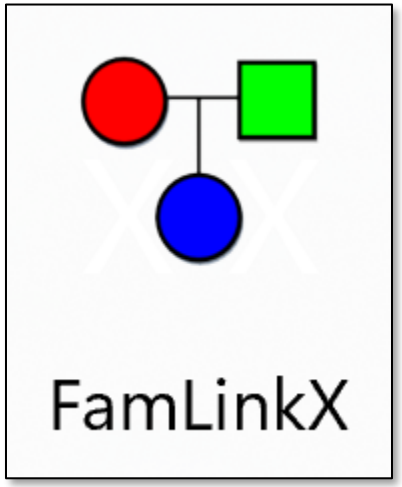
Argus X12 QS

Recommendation #8

Appropriate software should be used when calculating LRs based on X-chromosomal markers in kinship analysis to avoid manual calculation errors. The software should rely on likelihood calculations and should be able to accommodate linkage, linkage disequilibrium and mutations.

Recommendation #9

As for any other software calculating likelihood ratios to evaluate competing kinship scenarios, use of software for X chromosome applications should follow the recommendations from the ***DNA Commission of the ISFG*** on the validation of software programs



M.D. Coble et al.,, DNA Commission of the International Society for Forensic Genetics: recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, Forensic Sci. Int. Genet. 25 (2016) 191–197.

Recommendation #10

Individual autosomal LR and X-chromosomal LR results should only be combined whenever equivalent (and clearly defined) hypotheses are used for both autosomal and X-chromosomal data, and when it is appropriate to assume that substructure and LD between autosomal and X-chromosomal alleles do not play a role.

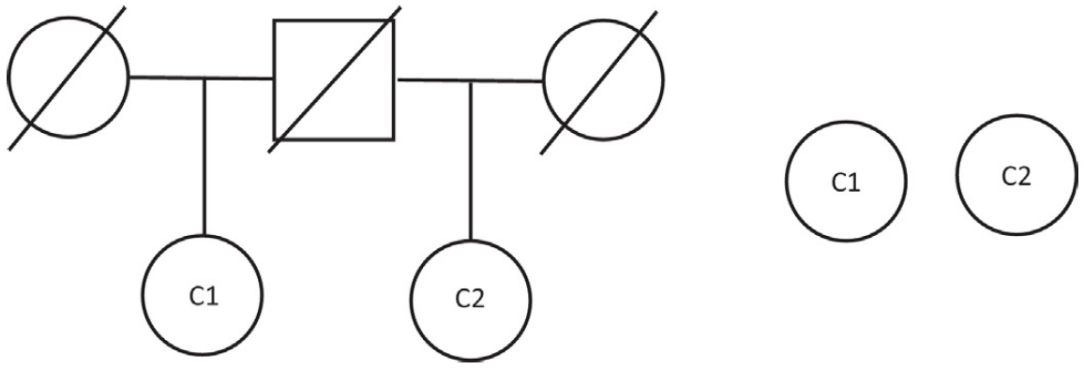
$$LR_{A,X} = LR_A \cdot LR_X$$

- If a population is not stratified, it is not expected to find LD between alleles from markers at different chromosomes and it is possible to multiply the LRs obtained for the autosomal and X-chromosomal markers respectively.
- To combine the LR from autosomal markers with the LR from X-chromosomal markers, the case specific hypotheses must be unambiguously formulated and be equivalent for the autosomal and for the X-chromosomal calculations.

Understanding IBD sharing and kinship coefficients

- Can be used to understand if autosomal or X-Chromosomal DNA markers are more informative to solve a specific relationship

- Is X or autosomal DNA markers more informative?



IBD sharing: k0, k1, k2

- • k0: Probability of sharing 0 alleles identical by descent (IBD)
- • k1: Probability of sharing 1 allele IBD
- • k2: Probability of sharing 2 alleles IBD

- Different relationships have different expected values for k0, k1, and k2.

Kinship Coefficient (Φ)

- The kinship coefficient (Φ) is the probability that two alleles, one from each individual, are IBD.
- For autosomal, it is calculated from k-values using: $\Phi = (1/4)k1 + (1/2)\ k2$.
- For X, it is calculated from k-values using: $\Phi = (1/2)k1 + k2$ or $(1/4)k1 + (1/2)\ k2$.
- **The higher the value of Φ, the more informative can the analysis be!**

Let's use QuickPed (https://magnusdv.shinyapps.io/quickped/) instead of doing estimations by hand!

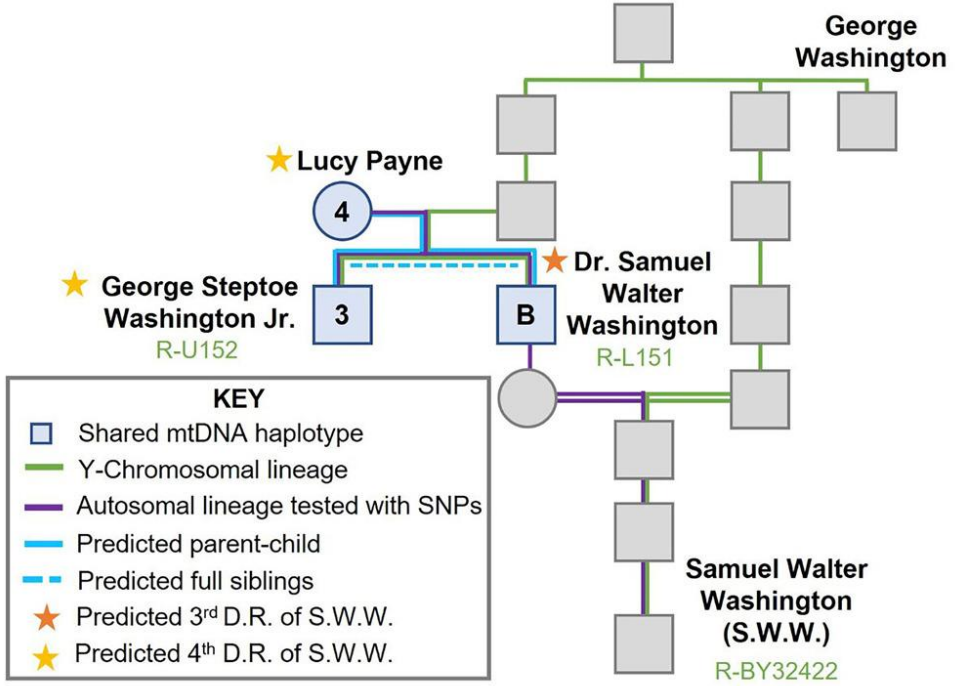
The use of X SNP data in a historic case

ARTICLE · Volume 27, Issue 4, 109353, April 19, 2024 · ***Open Access***

Unearthing who and Y at Harewood Cemetery and inference of George Washington's Y-chromosomal haplotype

Courtney Cavagnino [1,2,10] · Göran Runfeldt [3] · Michael Sager [3] · Roberta Estes [3] · Andreas Tillmar [4,5] · Ellen M. Greytak [6] · Jacqueline Tyler Thomas [1,2] · Elise Anderson [1,7] · Jennifer Daniels-Higginbotham [1,2] · Katelyn Kjelland [1,7] · Kimberly Sturk-Andreaggi [1,2] · Thomas J. Parsons [8] · Timothy P. McMahon [1] · Charla Marshall [1,2,9] Show less

Affiliations & Notes Article Info



The use of X SNP data in a historic case

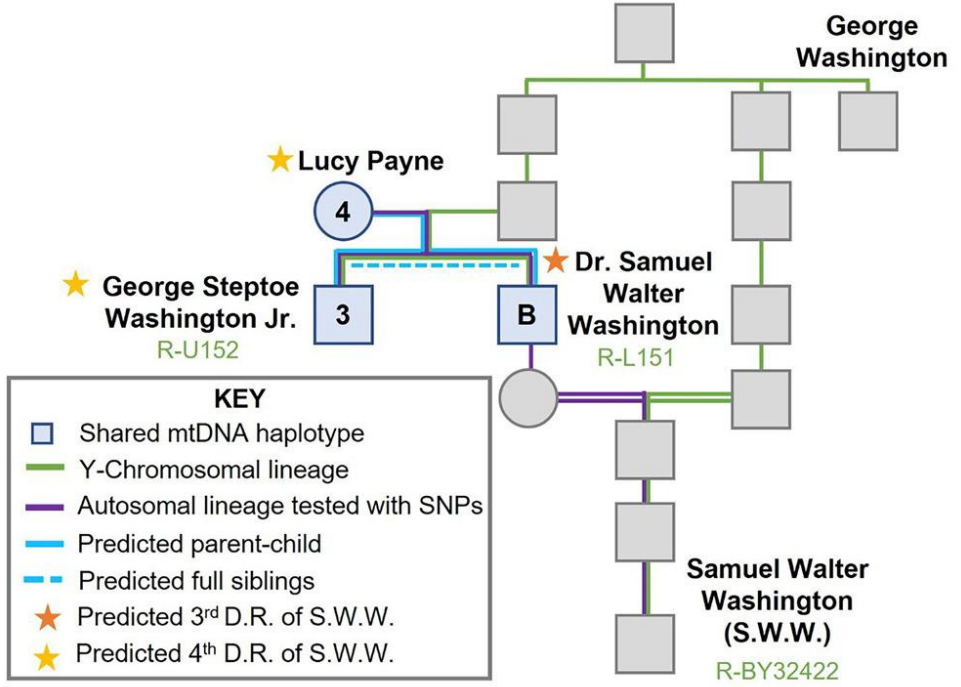
ARTICLE · Volume 27, Issue 4, 109353, April 19, 2024 · ***Open Access***

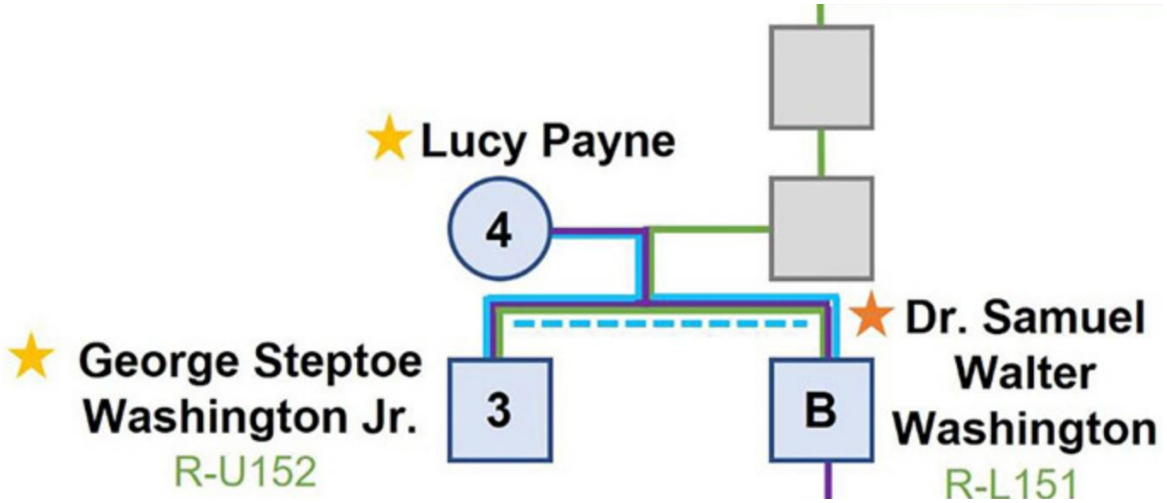
Unearthing who and Y at Harewood Cemetery and inference of George Washington's Y-chromosomal haplotype

Courtney Cavagnino [1,2,10] · Göran Runfeldt [3] · Michael Sager [3] · Roberta Estes [3] · Andreas Tillmar [4,5] · Ellen M. Greytak [6] · Jacqueline Tyler Thomas [1,2] · Elise Anderson [1,7] · Jennifer Daniels-Higginbotham [1,2] · Katelyn Kjelland [1,7] · Kimberly Sturk-Andreaggi [1,2] · Thomas J. Parsons [8] · Timothy P. McMahon [1] · Charla Marshall [1,2,9] Show less

Affiliations & Notes Article Info

- Were the bodies found in the cemetery, "Lucy P", "George SW" and "Samuel WW"?

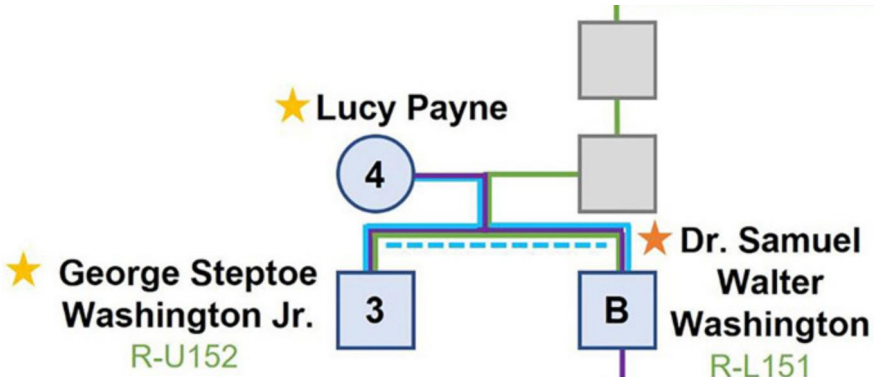


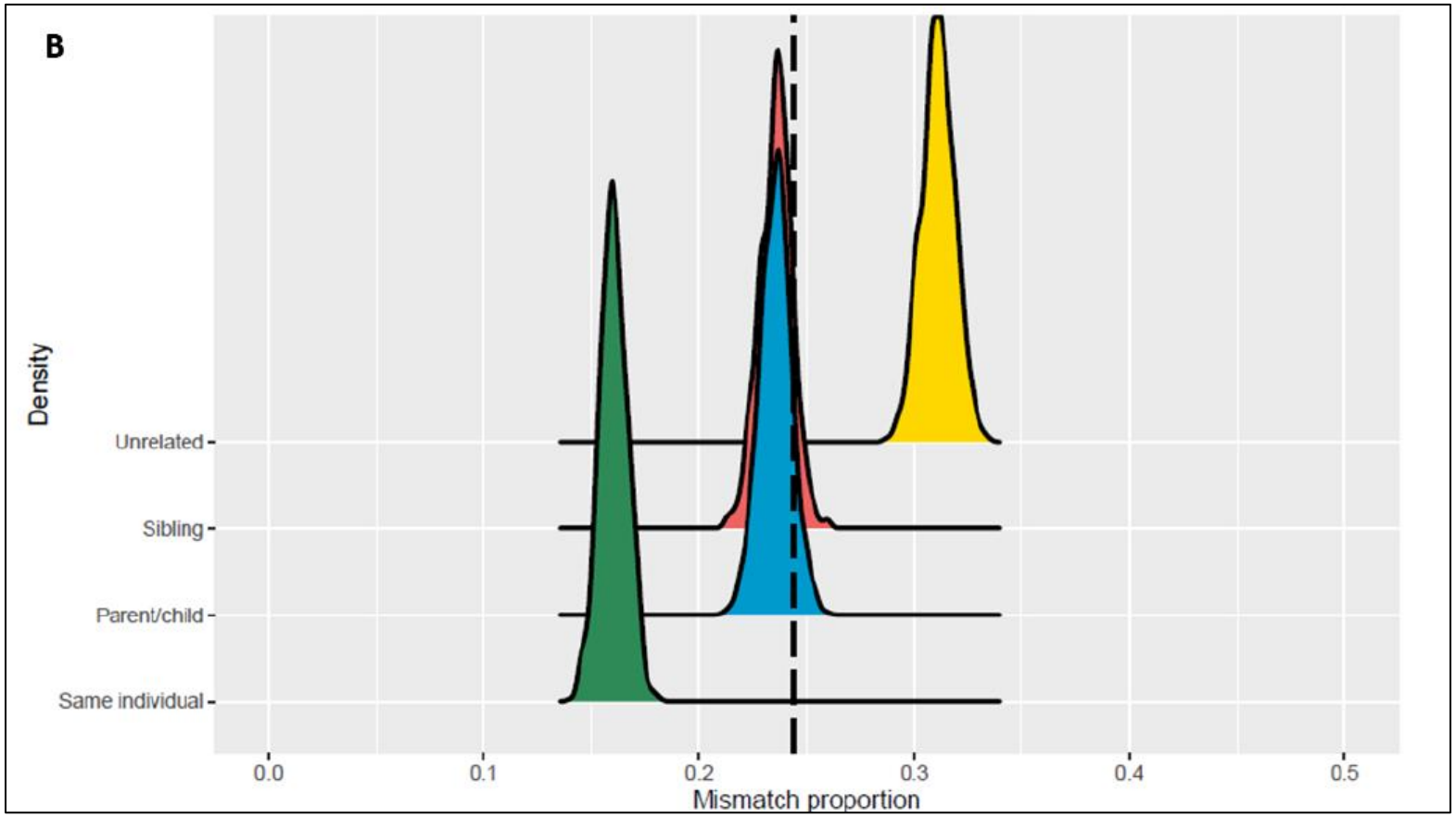


- Can DNA be used to show that 4 is the mother of 3 and B, and 3 and B are full siblings?

- Poor DNA samples
- Approx **1X coverage** (=unknown homozygous/heterozygous status for the diploid DNA markers)
- Autosomal SNPs (95K) shows **first-degree relationships** based on probabilistic genotyping.
- First-degree parent/child **or** first-degree full siblings?
- Can we use X data? (YES!)

4 and 3 – Autosomal DNA markers





Forced haploid genome analysis

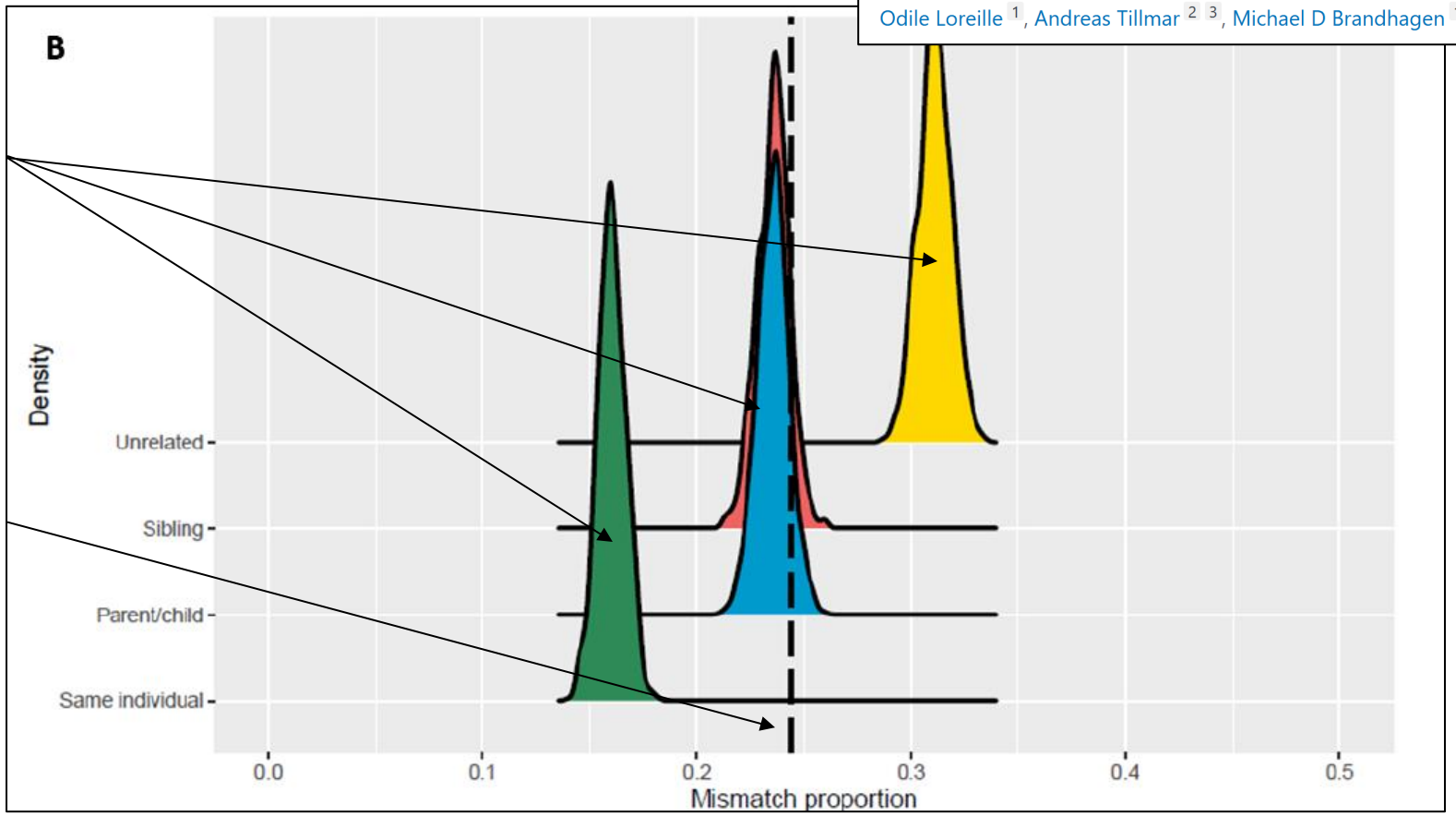
Genes (Basel). 2022 Jan 23;13(2):202. doi: 10.3390/genes13020202.

Improved DNA Extraction and Illumina Sequencing of DNA Recovered from Aged Rootless Hair Shafts Found in Relics Associated with the Romanov Family

Odile Loreille [1], Andreas Tillmar [2 3], Michael D Brandhagen [1], Linda Otterstatter [4], Jodi A Irwin [1]

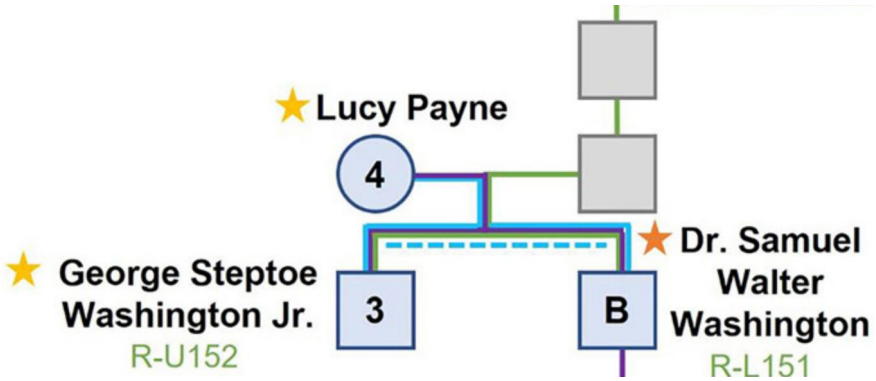
The colored areas are expected proportions for different relationship (based on simulations)

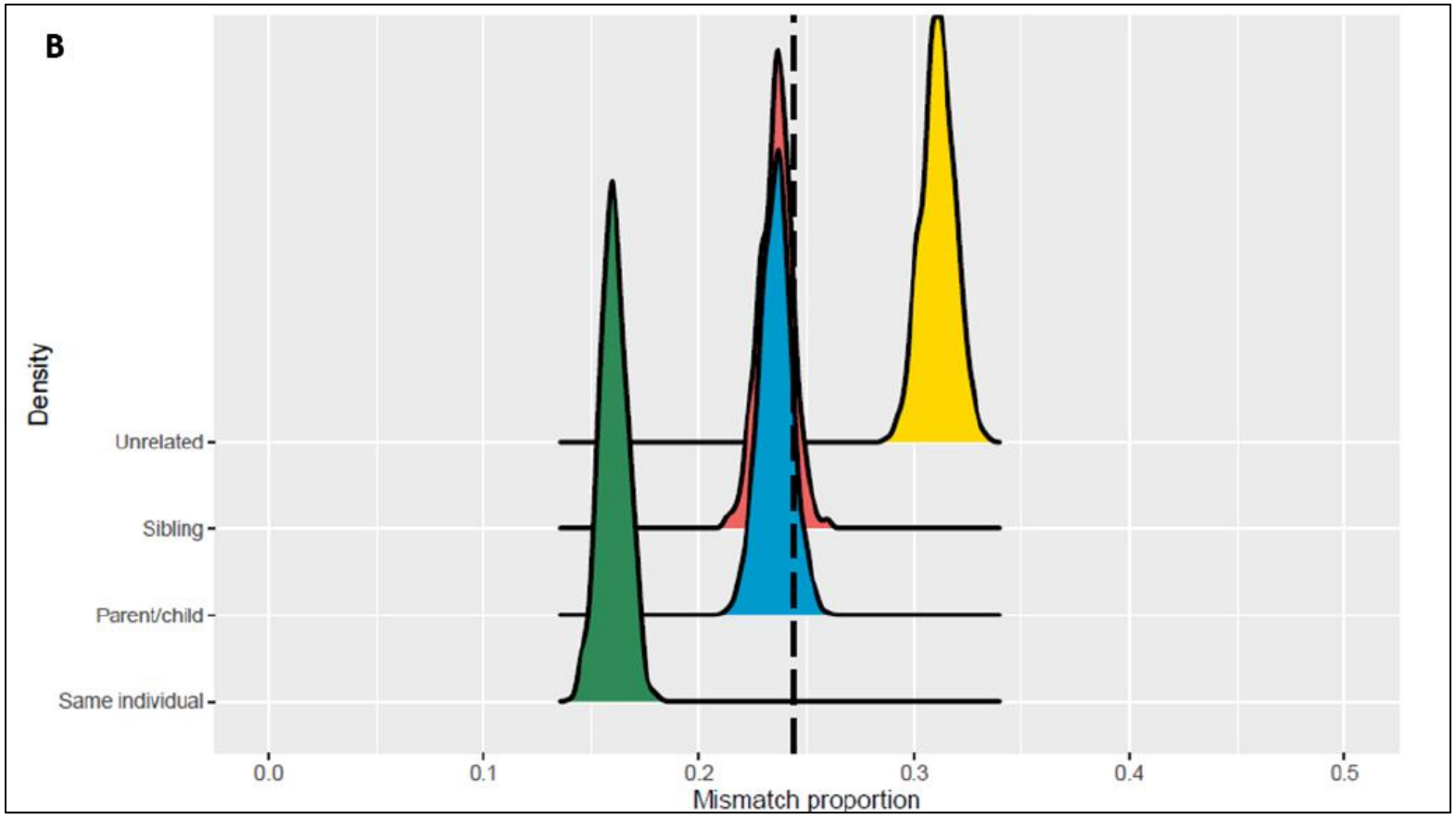
The dashed line is the estimated proportion in the specified comparison



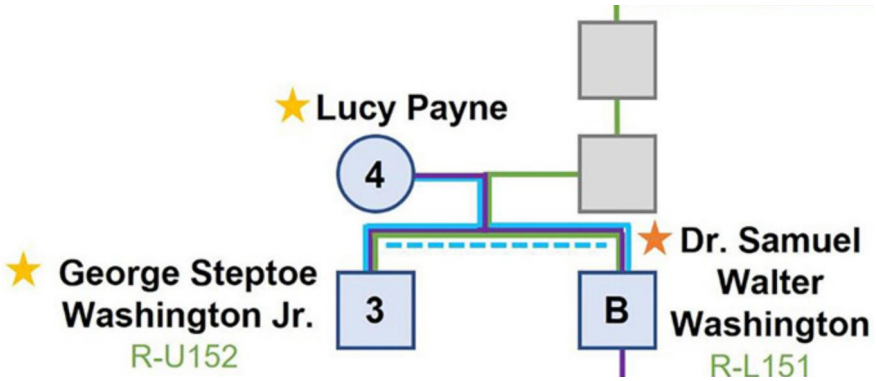
Mismatch proportion is measured by estimating non-identical alleles between forced haploid genomes. **Mismatch proportion** is reversely proportional to the kinship coefficient.

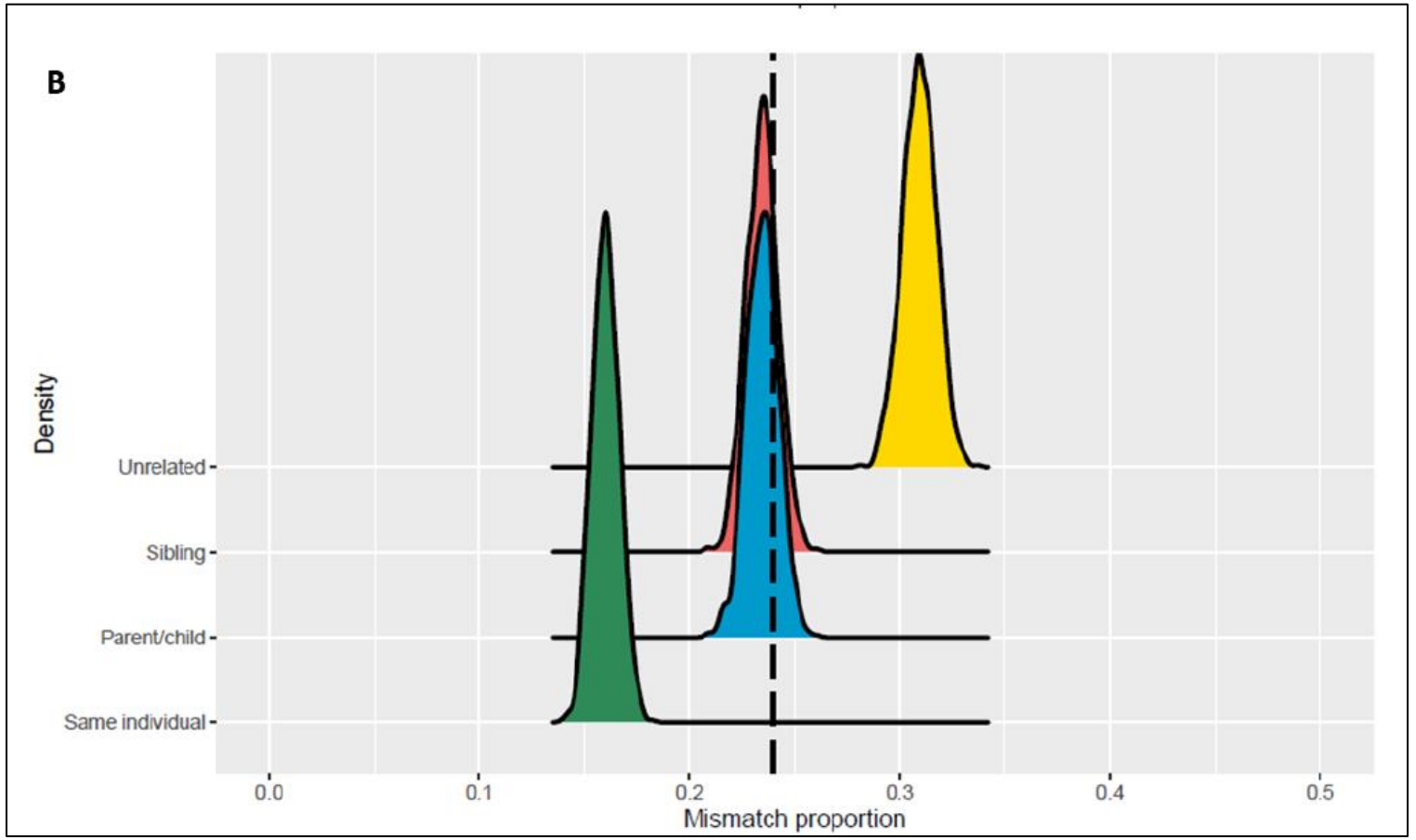
4 and 3 – Autosomal DNA markers



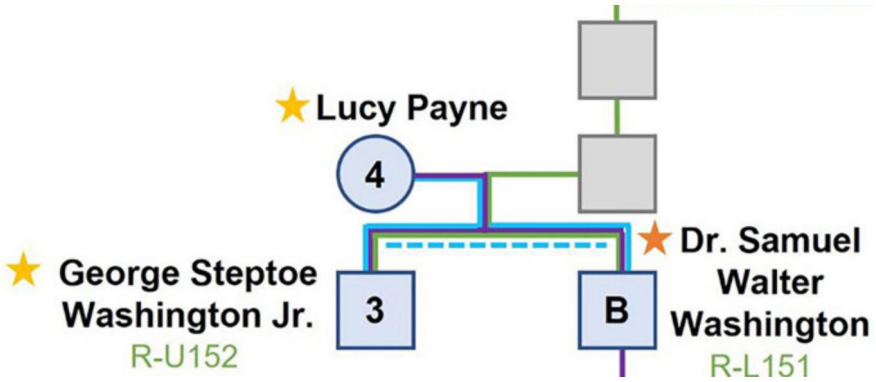


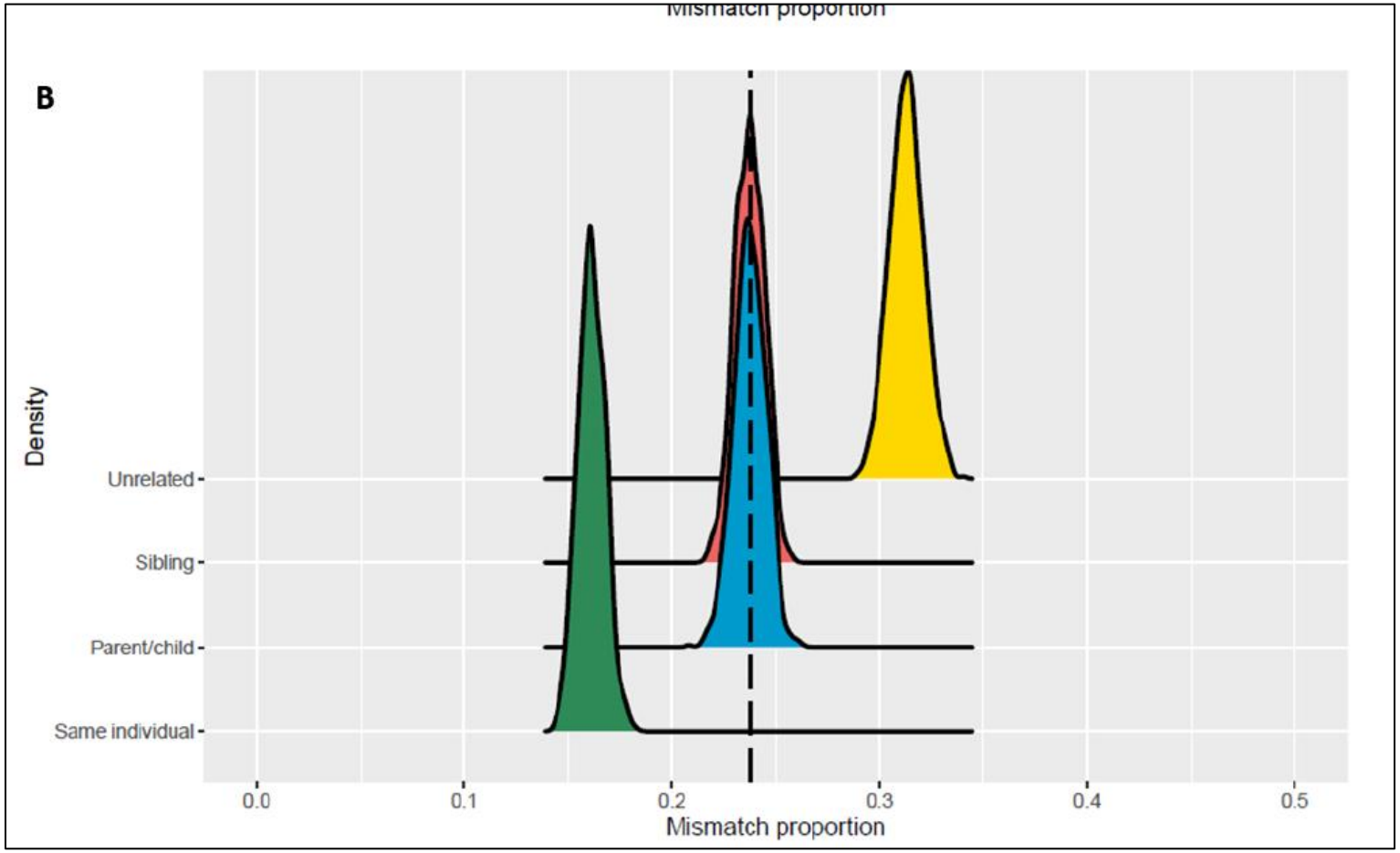
3 and B – Autosomal DNA markers



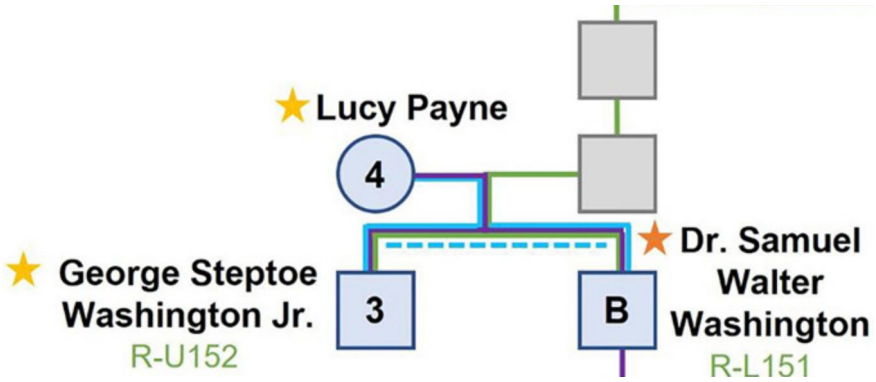


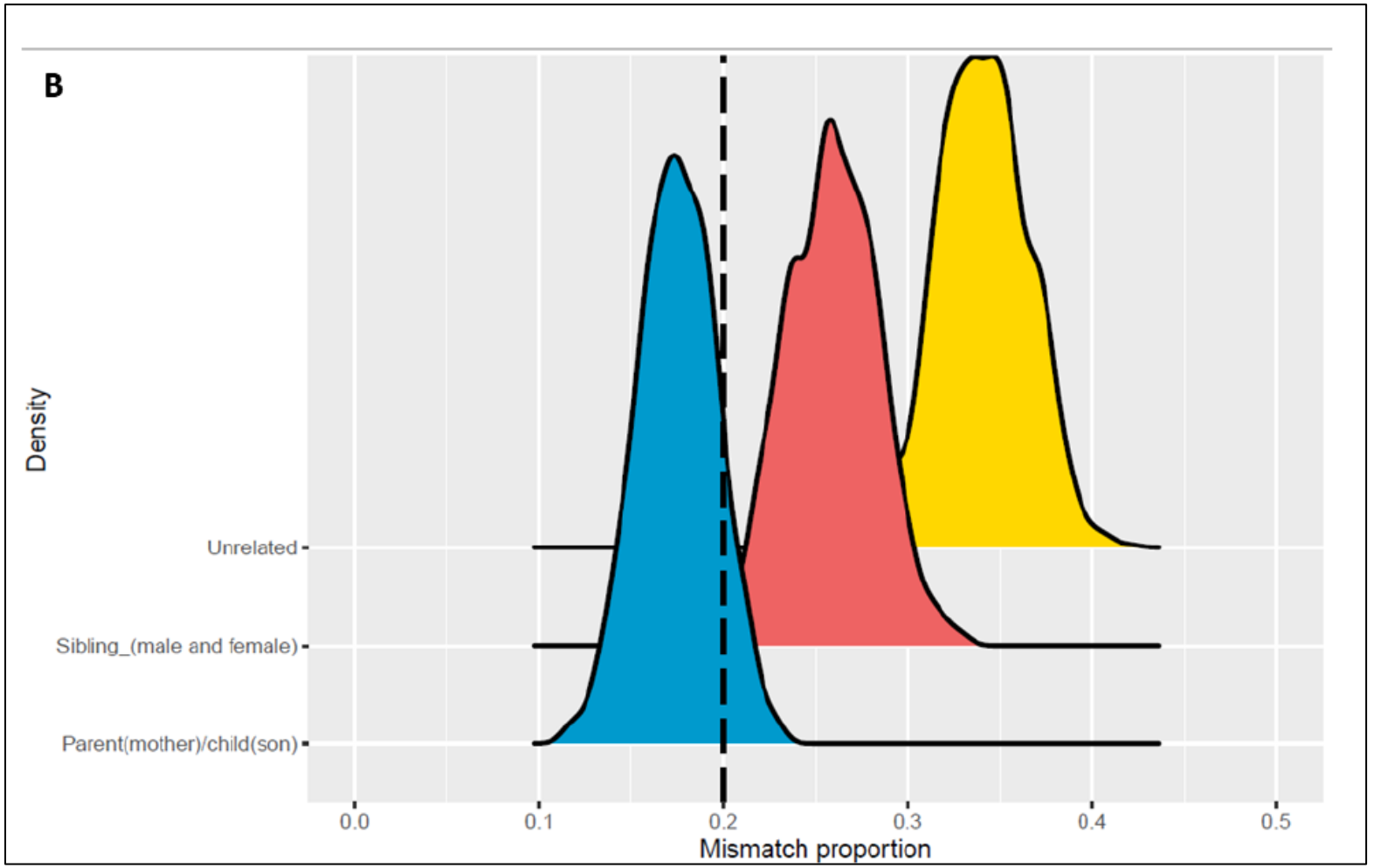
4 and B – Autosomal DNA markers



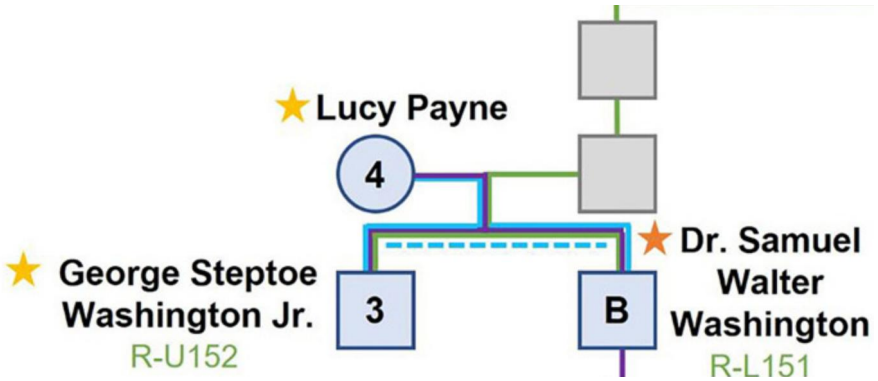


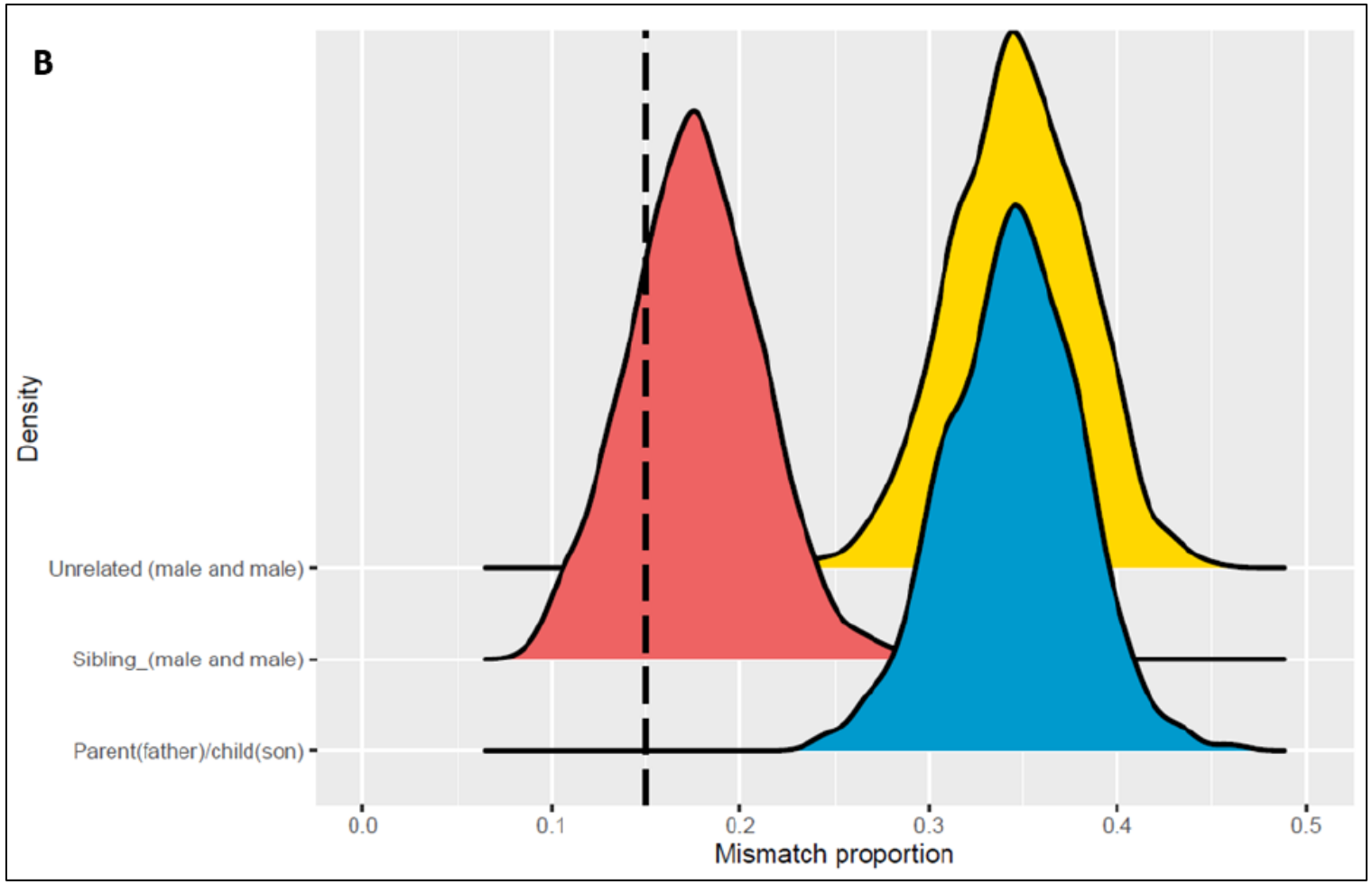
4 and 3 – X markers



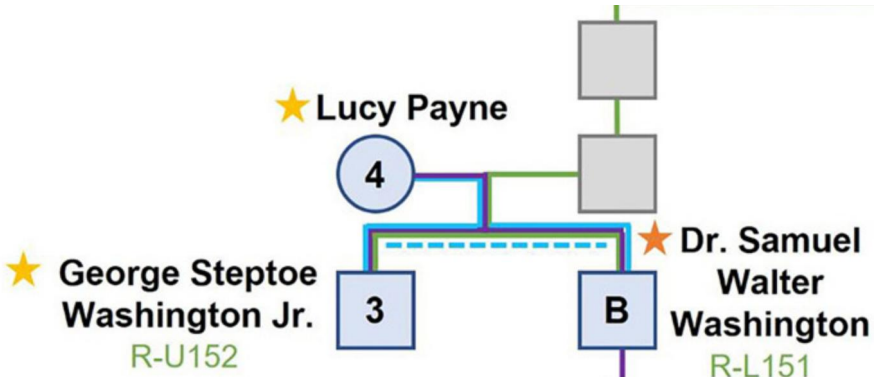


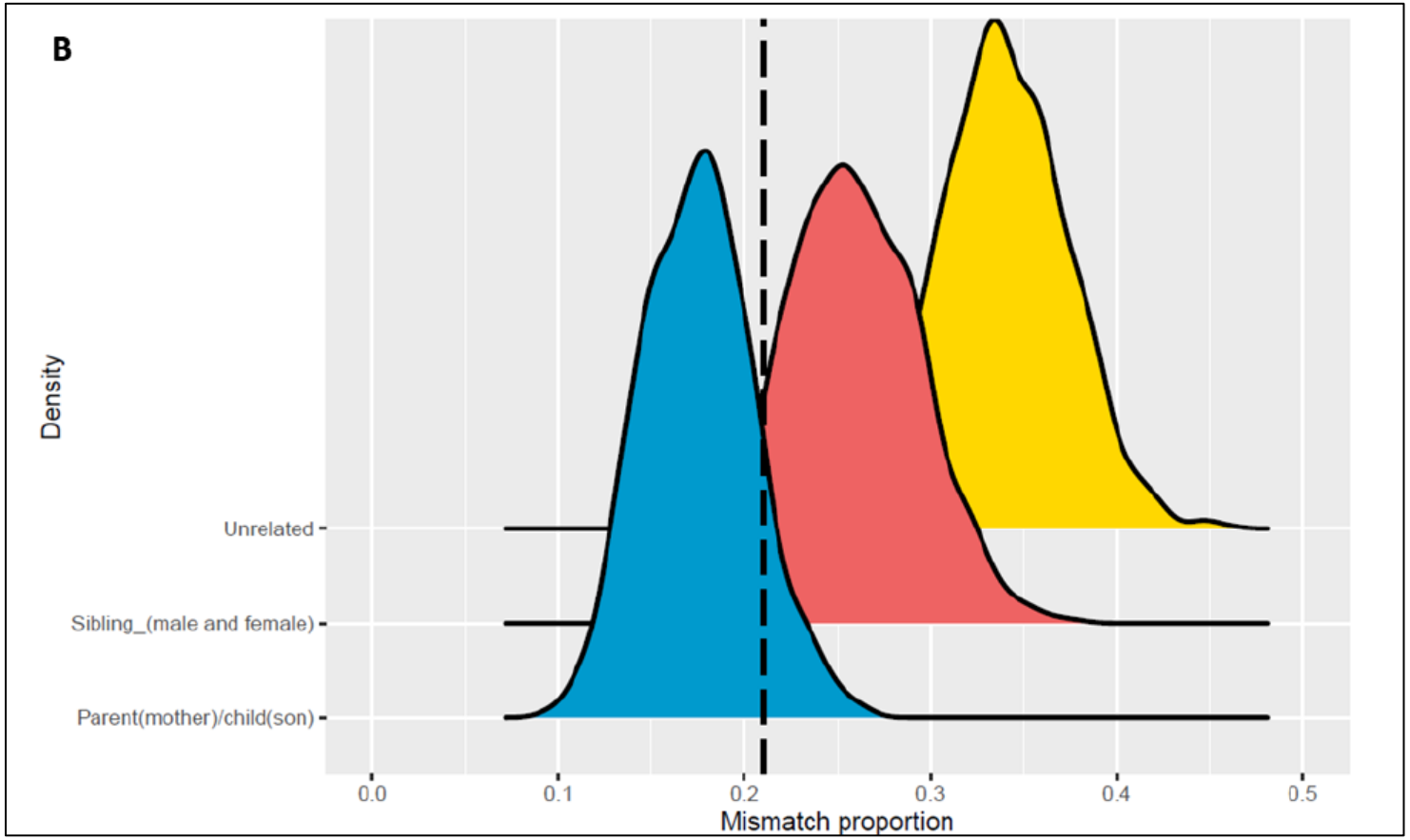
3 and B – X markers





4 and B – X markers





X-chromosomal markers in Forensic Genetics

GHEP 2025 Virtual workshop series.

March ~~10,17~~ and **24th**

Daniel Kling and Andreas Tillmar