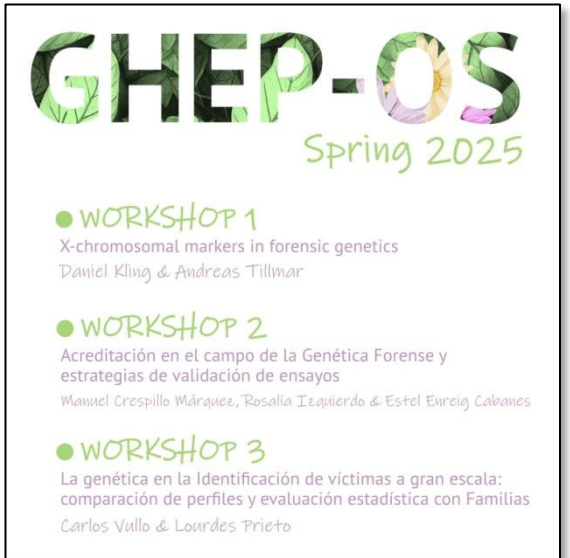
Case examples and advanced topics

DANIEL KLING & ANDREAS TILLMAR

Daniel.kling@rmv.se

Andreas.tillmar@rmv.se



Teachers

Daniel Kling, PhD

- Forensic Expert
- National Board of Forensic Medicine, Sweden
- Worked in the field for almost 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 almost 20 years
- Technical leadership mixed with R&D
- Applied biostatistics, relationship inference, population genetics, genetic genealogy

Disclaimer!

Points of view are those of the presenters and do not necessarily represent the official position or policies of the National Board of Forensic Medicine or ISFG. Certain commercial software, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Topics

1. Case examples
2. Pseudo-counts
3. Anepleudies
4. SNPs and expanded marker panels (Simulations)

Case examples

Curious cases of X

See https://link.springer.com/article/10.1007/s00414-017-1612-8 for further cases!

International Journal of Legal Medicine
March 2018, Volume 132, Issue 2, pp 361–371 | Cite as

Curiosities of X chromosomal markers and haplotypes

Authors | Authors and affiliations

Daniel Kling

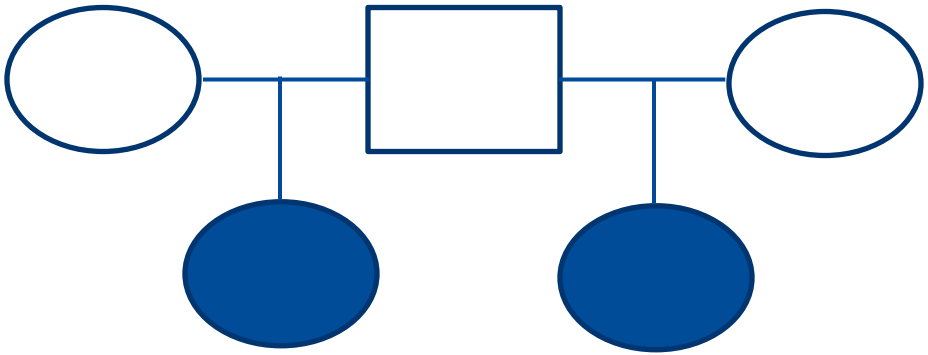
Original Article
First Online: 26 May 2017

Abstract

Recent progress in forensic genetics has introduced a number of closely located short tandem repeat (STR) markers on the X chromosome. Inevitably, dependencies arise that have to be accounted for. This paper will in detail explore the complex statistical interpretation of X-chromosomal STR markers, focusing on likelihood calculations. Specifically, we will investigate how the phase uncertainty of haplotypes comes into play in the statistical evaluations and what curious effects this phenomenon can have. The starting point is the different real cases where the weight of evidence has provided unexpected results that require further investigation in order to be fully understood. We will touch upon subjects such as association between alleles, recombinations, and mutations. The aim of this study is to facilitate a better understanding of the interaction between the concepts in addition to provide an understanding why good estimates of haplotype frequencies are crucial. The individual subjects have been discussed in other fields, whereas this study will focus on forensic applications where few studies have been conducted relating to the understanding of how these concepts interact.

Curious case of X (1)

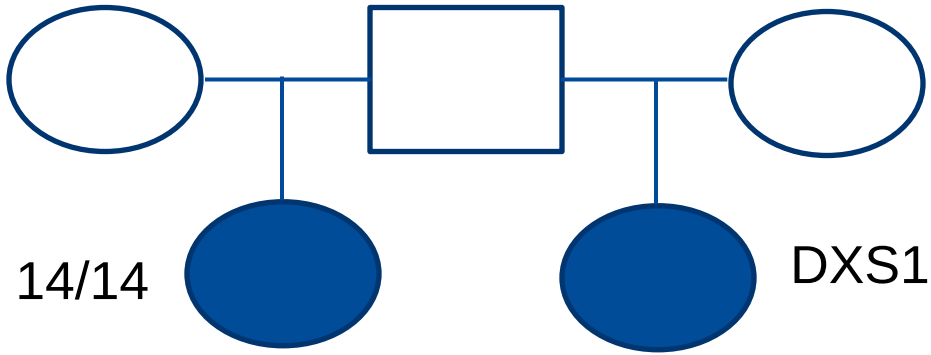
H1: Paternal half sisters
H2: Unrelated



IBD1=1

Curious case of X (1)

H1: Paternal half sisters
H2: Unrelated



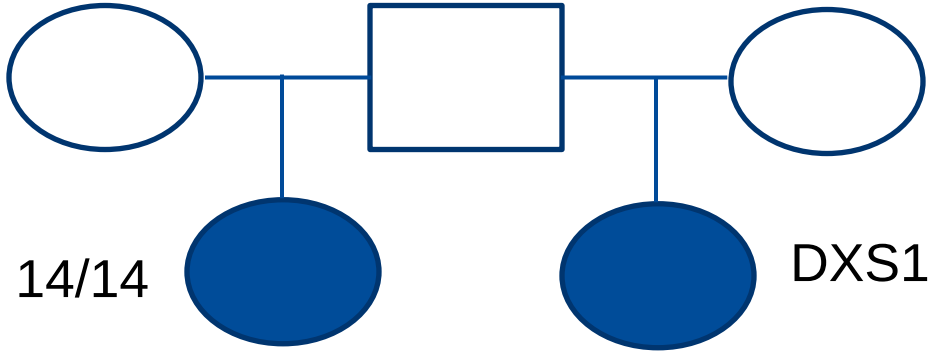
DXS10135: 14/14

DXS10135: 14/14

Curious case of X (1)

H1: Paternal half sisters
H2: Unrelated

14 14 14



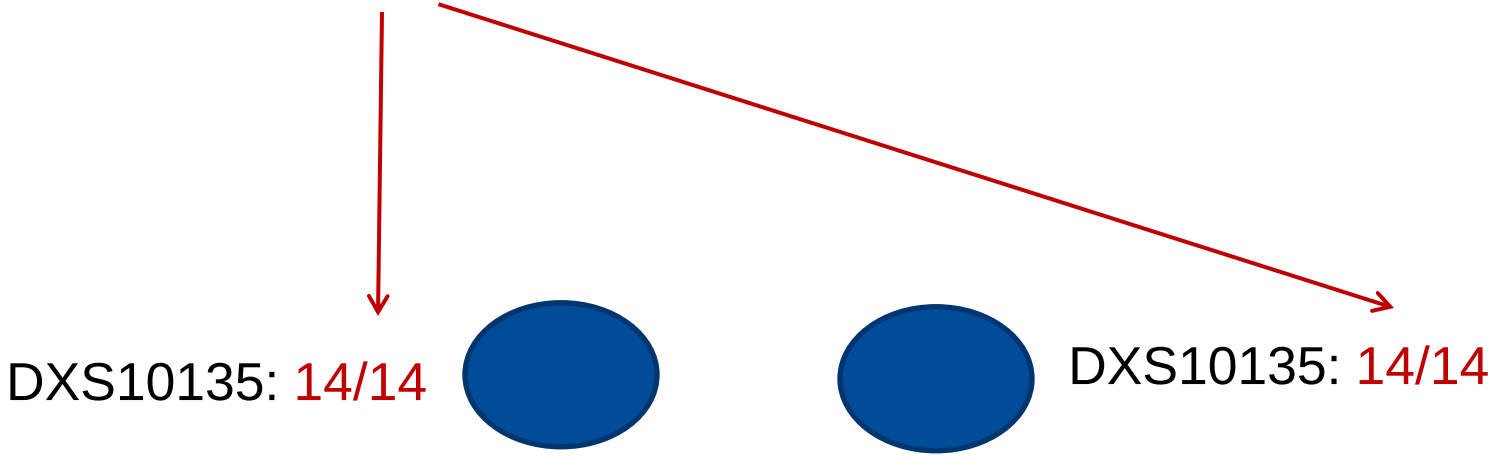
DXS10135: 14/14

DXS10135: 14/14

$LR=f_{14}^{3}/\ldots$

Curious case of X (1)

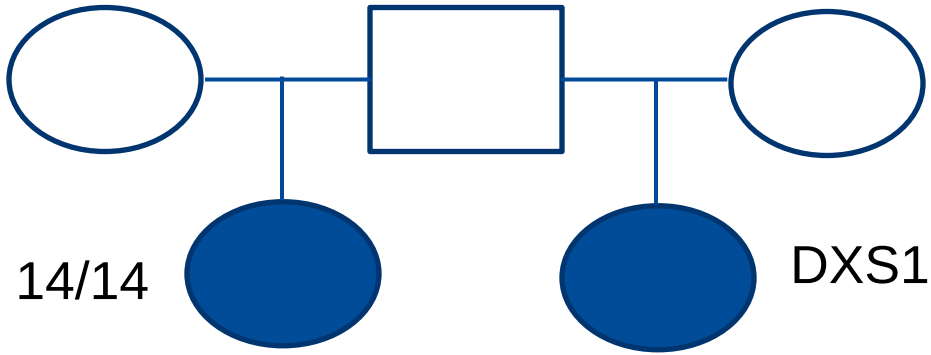
H1: Paternal half sisters
H2: Unrelated



$LR = f_{14}^{3} / f_{14}^{4} = \ldots$

Curious case of X (1)

H1: Paternal half sisters
H2: Unrelated



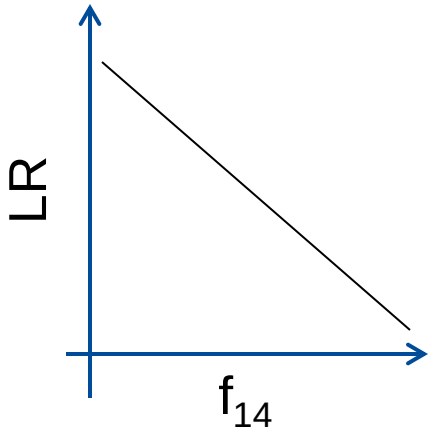
DXS10135: 14/14

DXS10135: 14/14

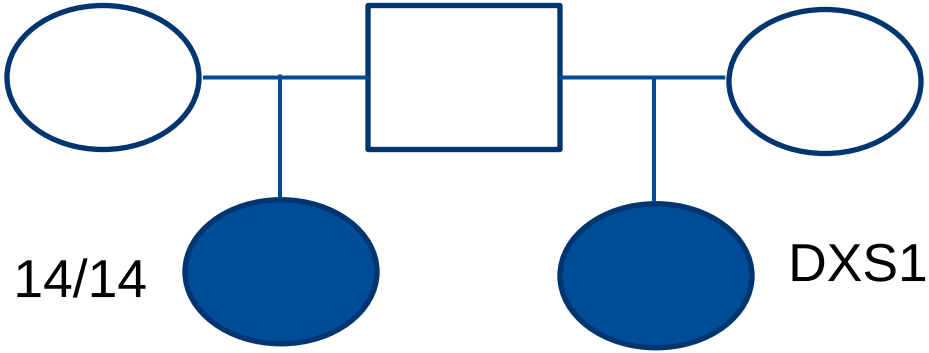
$LR=f_{14}^{3}/f_{14}^{4}=1/f_{14}$

Curious case of X (1)

H1: Paternal half sisters
H2: Unrelated



$LR=1/f_{14}$



DXS10135: 14/14

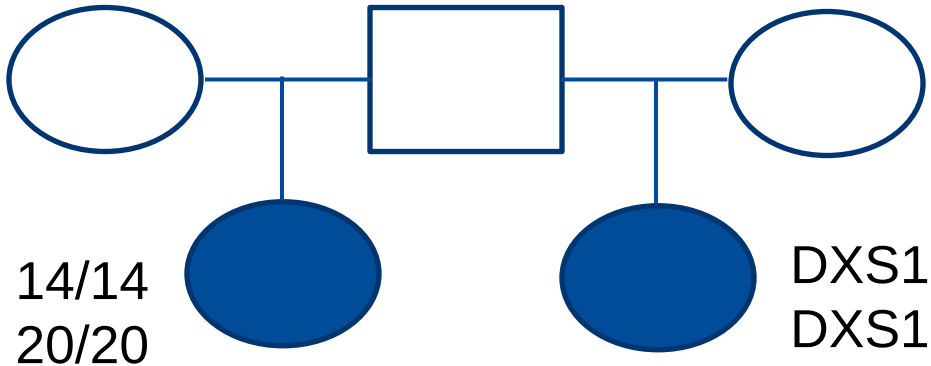
DXS10135: 14/14

Conclusion: If 14 is rare, the LR is high

Curious case of X (1)

H1: Paternal half sisters
H2: Unrelated

$LR_{LE}=1/f_{14}*1/f_{20}$
$LR_{LD}=1/f_{14\_20}$



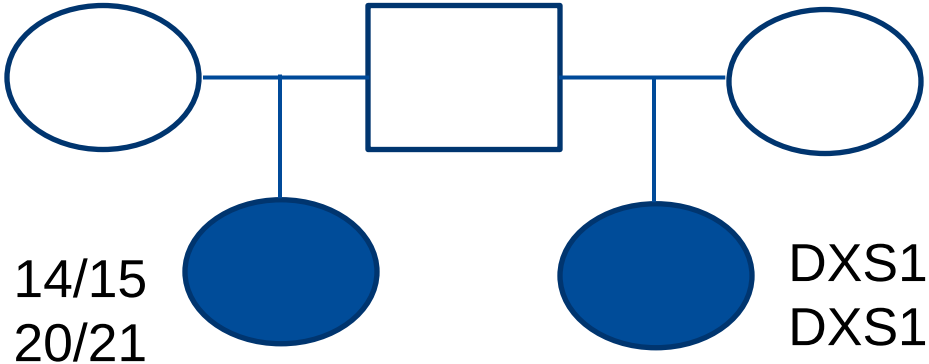
DXS10135: 14/14
DXS10124: 20/20

DXS10135: 14/14
DXS10124: 20/20

Curious case of X (1)

H1: Paternal half sisters
H2: Unrelated

$LR_{LE}=1/4f_{14}*1/4f_{20}$
$LR_{LD}=???$



DXS10135: 14/15
DXS10124: 20/21

DXS10135: 14/16
DXS10124: 20/22

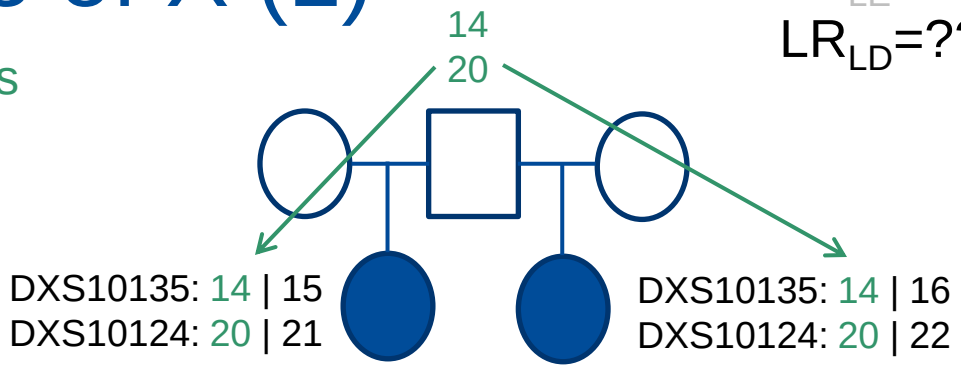
Curious case of X (1)

$LR_{LE}=1/4f_{14}*1/4f_{20}$

$LR_{LD}=???$

H1: Paternal half sisters

H2: Unrelated



$LR_{LD}=f_{14\_20}*f_{15\_21}*f_{16\_22}/\ldots$

Curious case of X (1)

$LR_{LE}=1/4f_{14}*1/4f_{20}$

$LR_{LD}=???$

H1: Paternal half sisters

H2: Unrelated

DXS10135: 14 | 15
DXS10124: 20 | 21
or
…

DXS10135: 14 | 16
DXS10124: 20 | 22
or
…

$LR_{LD}=f_{14\_20}*f_{15\_21}*f_{16\_22}/[(2f_{14\_20}f_{15\_21}+\ldots)*(2f_{14\_20}f_{16\_22}+\ldots]$

Curious case of X (1)

$LR_{LE}=1/4f_{14}*1/4f_{20}$

$LR_{LD}=???$

H1: Paternal half sisters

H2: Unrelated

DXS10135: 14 | 15
DXS10124: 20 | 21
or
DXS10135: 15 | 14
DXS10124: 20 | 21

DXS10135: 14 | 16
DXS10124: 20 | 22
or
DXS10135: 16 | 14
DXS10124: 20 | 22

$$LR_{LD}=f_{14\_20}*f_{15\_21}*f_{16\_22}/[(2f_{14\_20}f_{15\_21}+2f_{14\_21}f_{15\_20})*(2f_{14\_20}f_{16\_22}+2f_{16\_20}f_{14\_22})]$$

Curious case of X (1)

$LR_{LE}=1/4f_{14}*1/4f_{20}$

$LR_{LD}=???$

H1: Paternal half sisters

H2: Unrelated

DXS10135: 14 | 15
DXS10124: 20 | 21
or
DXS10135: 15 | 14
DXS10124: 20 | 21

DXS10135: 14 | 16
DXS10124: 20 | 22
or
DXS10135: 16 | 14
DXS10124: 20 | 22

$$LR_{LD}=f_{14\_20}*f_{15\_21}*f_{16\_22}/[(2f_{14\_20}f_{15\_21}+2f_{14\_21}f_{15\_20})*(2f_{14\_20}f_{16\_22}+2f_{16\_20}f_{14\_22})]$$

Conclusion: Since the phase (haplotypes) is not certain, we need to add both possibilities

Curious case of X (1)

$LR_{LE}=1/4f_{14}*1/4f_{20}$

$LR_{LD}=???$

H1: Paternal half sisters

H2: Unrelated

DXS10135: 14 | 15
DXS10124: 20 | 21
or
DXS10135: 15 | 14
DXS10124: 20 | 21

DXS10135: 14 | 16
DXS10124: 20 | 22
or
DXS10135: 16 | 14
DXS10124: 20 | 22

$$LR_{LD}=f_{14\_20}*f_{15\_21}*f_{16\_22}/[(2f_{14\_20}f_{15\_21}+2f_{14\_21}f_{15\_20})*(2f_{14\_20}f_{16\_22}+2f_{16\_20}f_{14\_22})]$$

Assumption: 14_20 is rare, 15_20 and 16_20 are common

Curious case of X (1)

$LR_{LE}=1/4f_{14}*1/4f_{20}$

$LR_{LD}=???$

H1: Paternal half sisters

H2: Unrelated

$$LR_{LD}=f_{14\_20}*f_{15\_21}*f_{16\_22}/[(2f_{14\_21}f_{15\_20})*(2f_{16\_20}f_{14\_22})]$$

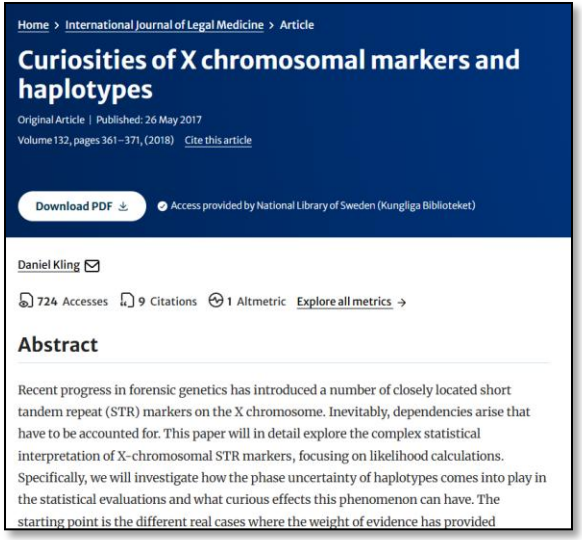
What happens if 14_20 is extremely rare???

Curious case of X (1)

H1: Paternal half sisters
H2: Unrelated

Table 2. Genotype data for three STR markers located in LG2 of the Argus X12 kit.

Marker	G1	G2
DXS7132	12,13	13,14
DXS10079	19,19	18,19
DXS10074	14,18	18,18



Home › International Journal of Legal Medicine › Article

Curiosities of X chromosomal markers and haplotypes

Original Article | Published: 26 May 2017

Volume 132, pages 361–371, (2018) Cite this article

Download PDF

Access provided by National Library of Sweden (Kungliga Biblioteket)

Daniel Kling

724 Accesses 9 Citations 1 Altmetric Explore all metrics

Abstract

Recent progress in forensic genetics has introduced a number of closely located short tandem repeat (STR) markers on the X chromosome. Inevitably, dependencies arise that have to be accounted for. This paper will in detail explore the complex statistical interpretation of X-chromosomal STR markers, focusing on likelihood calculations. Specifically, we will investigate how the phase uncertainty of haplotypes comes into play in the statistical evaluations and what curious effects this phenomenon can have. The starting point is the different real cases where the weight of evidence has provided

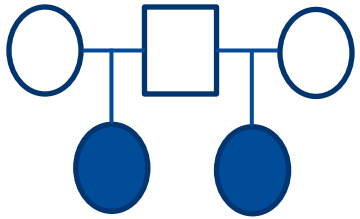
Curious case of X (1)

H1: Paternal half sisters (genotypes G1 and G2)
H2: Unrelated

Table 2. Genotype data for three STR markers located in LG2 of the Argus X12 kit.

Marker	G1	G2
DXS7132	12,13	13,14
DXS10079	19,19	18,19
DXS10074	14,18	18,18

DXS7132: 13
DXS10079: 19
DXS10074: 18



”Forced” paternal haplotype Maternal haplotypes

$$LR = \frac{L_1}{L_2} = \frac{H_{13,19,18} \cdot H_{12,19,14} \cdot H_{14,18,18}}{4\left(H_{12,19,14}H_{13,19,18} + H_{12,19,18}H_{13,19,14}\right) \cdot \left(H_{13,18,18}H_{14,19,18} + H_{13,19,18}H_{14,18,18}\right)}$$

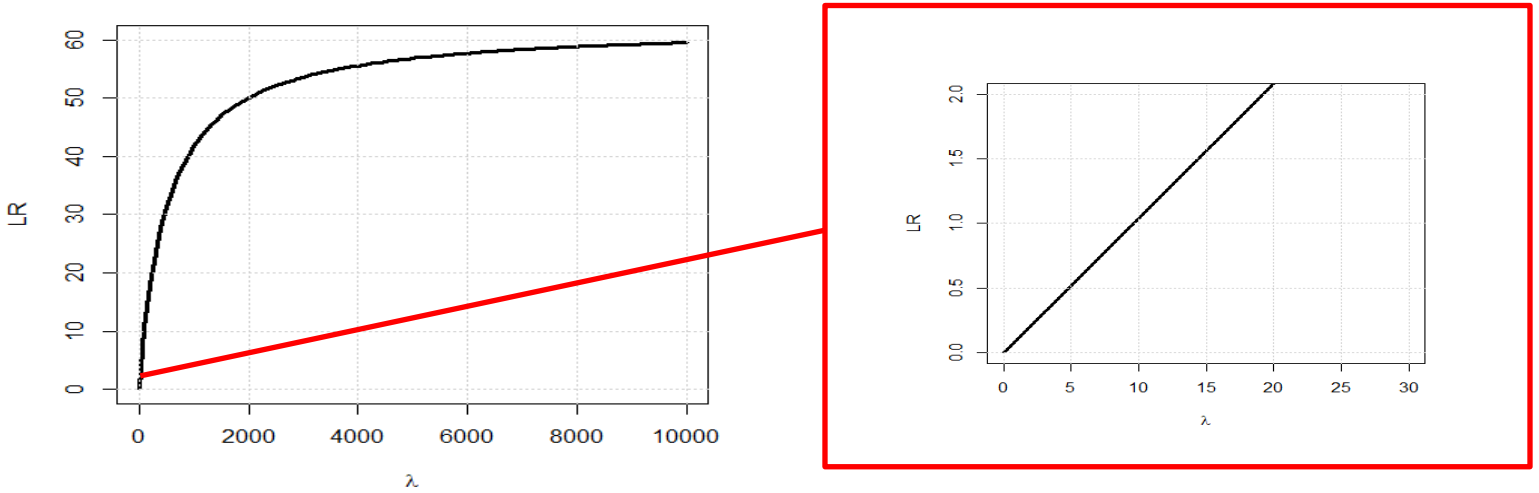
Two different haplotype setups Two different haplotype setups

Curious case of X (1)

H1: Paternal half sisters (genotypes G1 and G2)

H2: Unrelated

Interpretation: LR switches from below 1 to above at lambda=10.

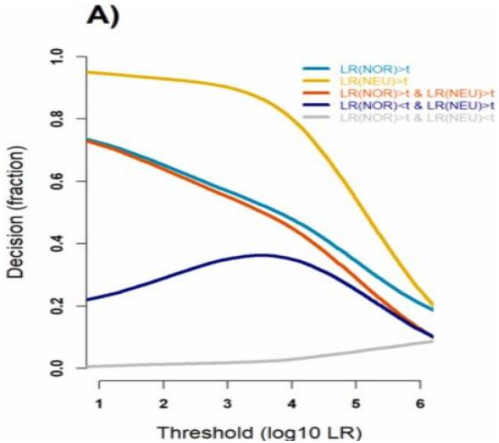


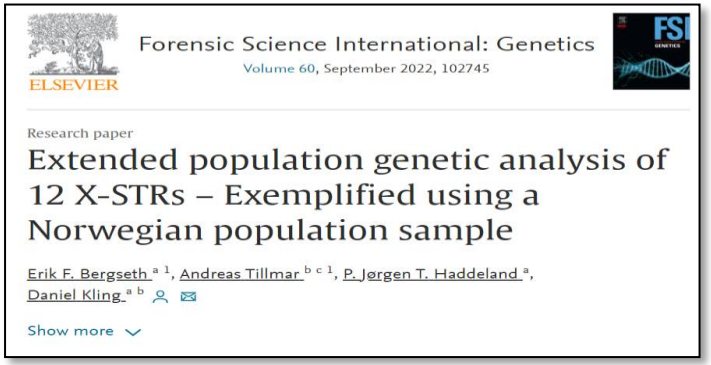
Answer lies in the combination of rare shared paternal haplotype and phase uncertainty

Curious case of X (1)

H1: Paternal half sisters (genotypes G1 and G2)
H2: Unrelated

How often does it happen??





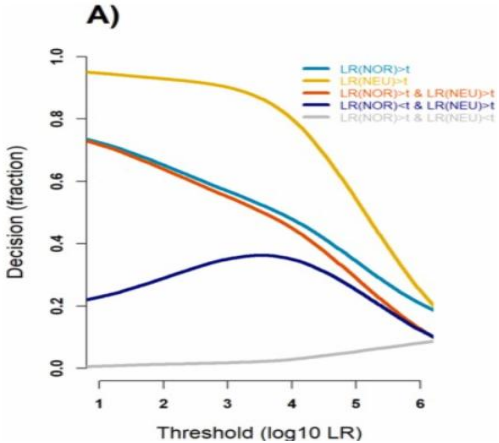
We simulate data in an extended Northern European (NEU) database, N=2624

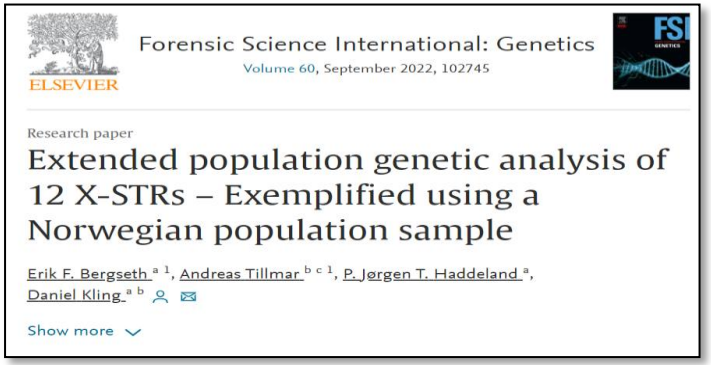
Compute LR using a smaller Norwegian (NOR) Database, N=631.

Curious case of X (1)

H1: Paternal half sisters (genotypes G1 and G2)
H2: Unrelated

How often does it happen??




Forensic Science International: Genetics
Volume 60, September 2022, 102745
ELSEVIER
Research paper
Extended population genetic analysis of 12 X-STRs – Exemplified using a Norwegian population sample
Erik F. Bergseth [a 1], Andreas Tillmar [b c 1], P. Jørgen T. Haddeland [a], Daniel Kling [a b]
Show more

We simulate data in an extended Northern European (NEU) database, N=2624

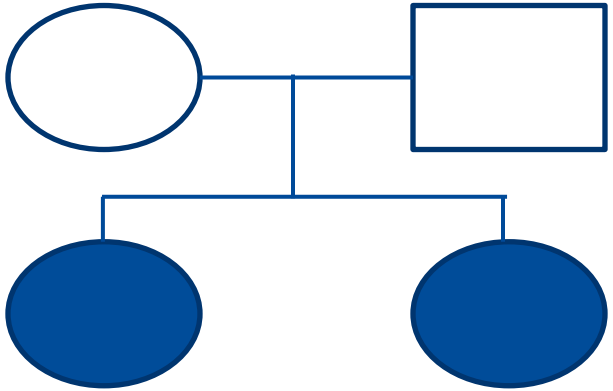
Compute LR using a smaller Norwegian (NOR) Database, N=631.

Curious case of X (2)

H1: Two females are full sisters
H2: The two females are maternal half sisters

Table 3. Excerpt of data from one of the linkage groups in the Argus X12 kit.

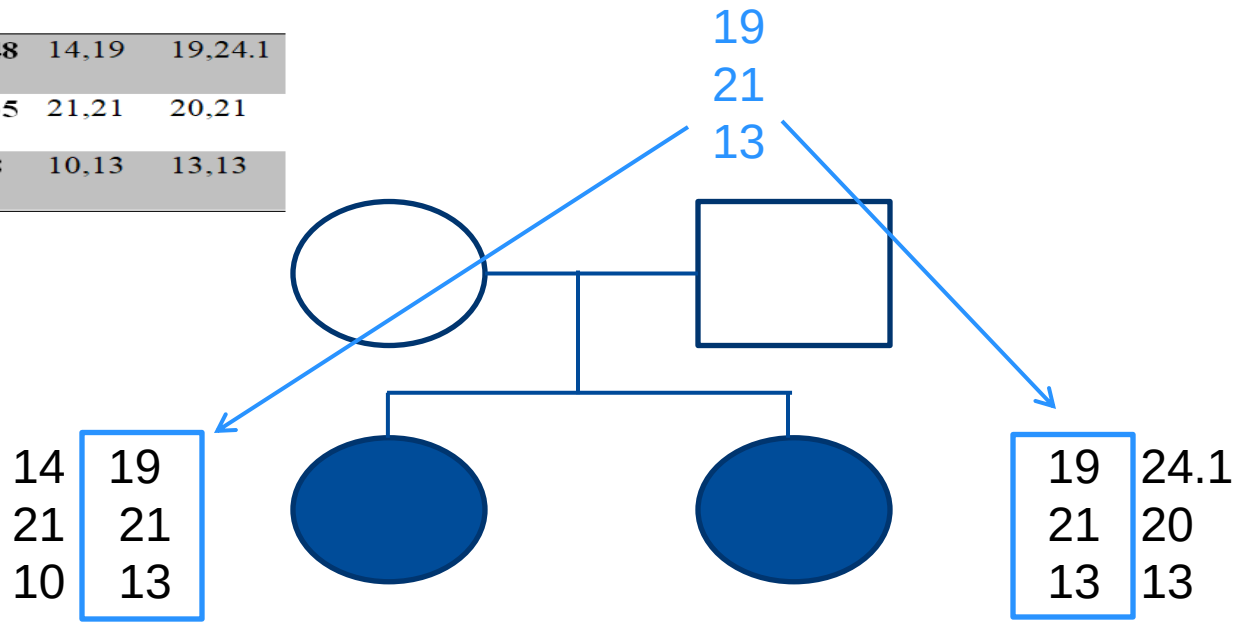
Marker	G1	G2
DXS10148	14,19	19,24.1
DXS10135	21,21	20,21
DXS8378	10,13	13,13



Curious case of X (2)

Table 3. Excerpt of data from one of the linkage groups in the Argus X12 kit.

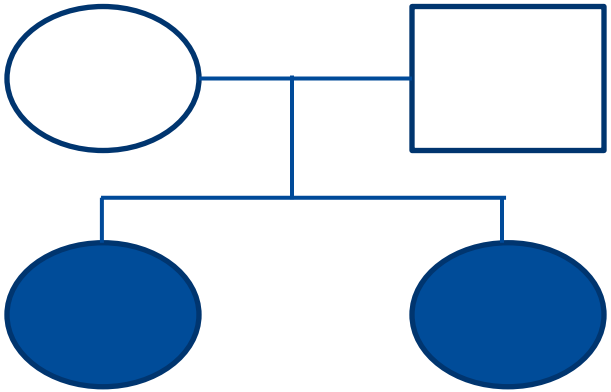
Marker	G1	G2
DXS10148	14,19	19,24.1
DXS10135	21,21	20,21
DXS8378	10,13	13,13



Curious case of X (2)

Table 3. Excerpt of data from one of the linkage groups in the Argus X12 kit.

Marker	G1	G2
DXS10148	14,19	19,24.1
DXS10135	21,21	20,21
DXS8378	10,13	13,13



14	19
21	21
10	13

Requires mutation

19	24.1
20	21
13	13

Curious case of X (2)

Table 3. Excerpt of data from one of the linkage groups in the Argus X12 kit.

Marker	G1	G2
DXS10148	14,19	19,24.1
DXS10135	21,21	20,21
DXS8378	10,13	13,13

No mutations

Paternal haplotype requires a single mutation

$$LR = \frac{L_1}{L_2} = \frac{0.5\left[(1-\mu)^6 F_{19,21,13} F_{14,21,10} F_{24.1,20,13} + 0.5\mu(1-\mu)^5 F_{19,20,13} F_{14,21,10} F_{24.1,21,13}\right]}{}$$

Curious case of X (2)

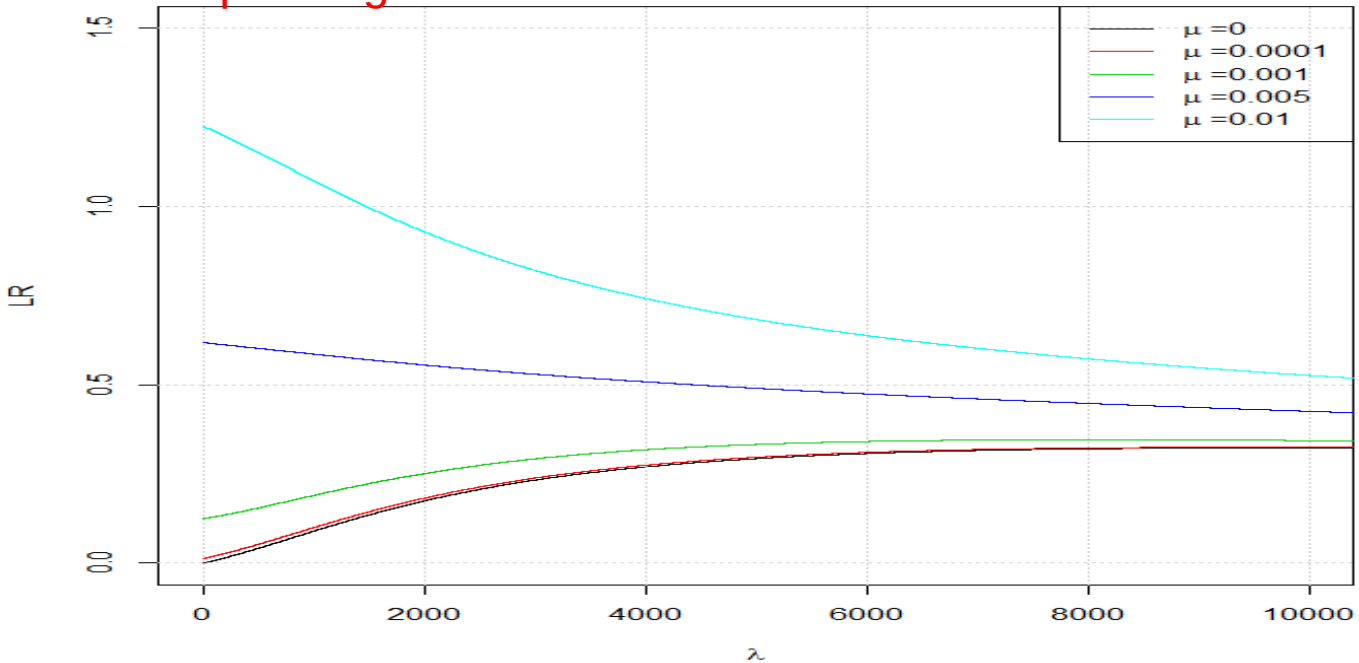
Table 3. Excerpt of data from one of the linkage groups in the Argus X12 kit.

Marker	G1	G2
DXS10148	14,19	19,24.1
DXS10135	21,21	20,21
DXS8378	10,13	13,13

$$LR = \frac{L_1}{L_2} = \frac{0.5\left[(1-\mu)^6 F_{19,21,13}F_{14,21,10}F_{24.1,20,13} + 0.5\mu(1-\mu)^5 F_{19,20,13}F_{14,21,10}F_{24.1,21,13}\right]}{F_{14,21,10}F_{19,21,13}\left[F_{24.1,20,13} + \left(2F_{19,20,13}F_{24.1,21,13} + 2F_{19,21,13}F_{24.1,20,13}\right)\right] + F_{14,21,13}F_{19,21,10}\left[\left(2F_{19,20,13}F_{24.1,21,13} + 2F_{19,21,13}F_{24.1,20,13}\right)\right]}$$

Maternal half sisters

Interpretation: LR depends on mutation rate and behaves differently depending on lambda



Curious case of X (3)

- Real example from a paternity case

Marker	Alleged father	Child
DXS7132	14	14/14
DXS10079	20	15/21
DXS10103	16	16/18

Possible mutation

Curious case of X (3)

- Real example from a paternity case

Marker	Alleged father	Child
DXS7132	14	14/14
DXS10079	20	15/21
DXS10103	16	16/18

Possible mutation

$$LR = \frac{L_1}{L_2} = \frac{F_{14,20,16}\mu(20 > 21)[\ ]}{F_{14,20,16}[]}$$

Paternal haplotype will ”cancel out”

Curious case of X (3)

- Real example from a paternity case

Marker	Alleged father	Child
DXS7132	14	14/14
DXS10079	20	15/21
DXS10103	16	16/18

Possible mutation

$$LR = \frac{L_1}{L_2} = \frac{F_{14,20,16}\,\mu(20 > 21)[\ ]}{F_{14,20,16}[]}$$

Paternal haplotype will ”cancel out”

Curious case of X (3)

- Real example from a paternity case

Marker	Alleged father	Child
DXS7132	14	14/14
DXS10079	20	15/21
DXS10103	16	16/18

$$LR = \frac{L_1}{L_2} = \frac{\mu(20 > 21)\left[F_{14,15,18} + F_{14,15,16}\right]}{\left[2F_{14,21,16}F_{14,15,18} + 2F_{14,21,18}F_{14,15,16}\right]}$$

Two alternative haplotype setups for the child

Haplotype	Observations in Sweden	Observations in Spain
[14 20 16]	14	0
[14 21 16]	6	1
[14 15 18]	0	1
[14 21 18]	6	1
[14 15 16]	5	0

Curious case of X (3)

- Real example from a paternity case

Marker	Alleged father	Child
DXS7132	14	14/14
DXS10079	20	15/21
DXS10103	16	16/18

$$LR_{swe} = \frac{\mu(20>21)F_{14,15,16}}{2F_{14,21,18}F_{14,15,16}} \approx \frac{\mu(20>21)}{2F_{14,21,18}}$$

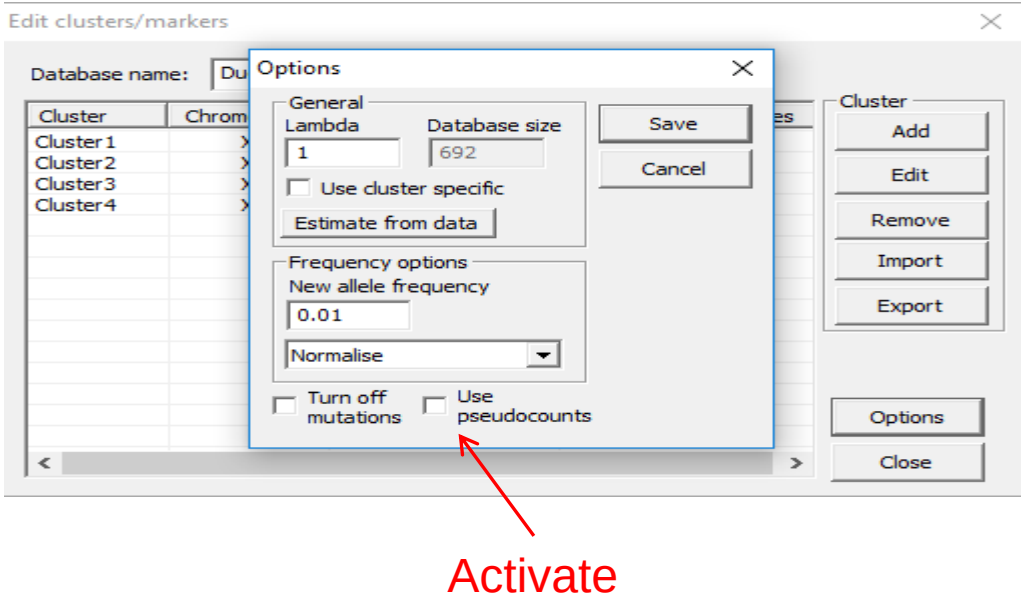
$$LR_{spa} = \frac{\mu(20>21)F_{14,15,18}}{2F_{14,21,16}F_{14,15,18}} \approx \frac{\mu(20>21)}{2F_{14,21,16}}$$

Haplotype	Observations in Sweden	Observations in Spain
[14 20 16]	14	0
[14 21 16]	6	1
[14 15 18]	0	1
[14 21 18]	6	1
[14 15 16]	5	0

Pseudo-counts

New frequency model

- We may use pseudocounts



New frequency model

- We may use pseudocounts

$$F_i = \frac{(c_i + d_i) + \lambda \pi_i}{C + D + \lambda}$$

D describes the total number of observations in a case
d_i is the number of ”weighted” observations for haplotype *i.*

Suggested by Balding for autosomal markers

New frequency model

- Example – Paternal half sisters (revisited)

Table 2. Genotype data for three STR markers located in LG2 of the Argus X12 kit.

Marker	G1	G2
DXS7132	12,13	13,14
DXS10079	19,19	18,19
DXS10074	14,18	18,18

D=4 (there are four observations, two for each female)

Haplotype	d_i
12 19 14	0.5
12 19 18	0.5
13 19 14	0.5
13 19 18	0.5
13 18 18	0.5
13 19 18	0.5
14 18 18	0.5
14 19 18	0.5

New frequency model

- Example – Paternal half sisters (revisited)

Table 2. Genotype data for three STR markers located in LG2 of the Argus X12 kit.

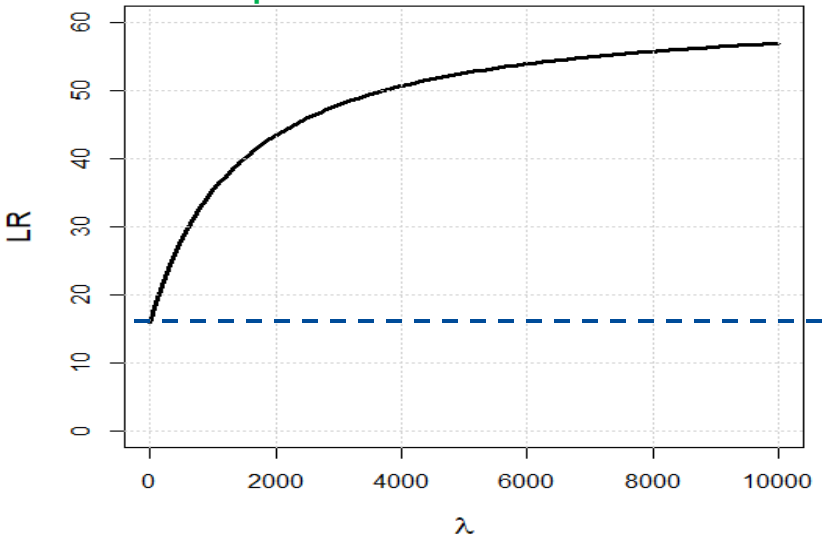
Marker	G1	G2
DXS7132	12,13	13,14
DXS10079	19,19	18,19
DXS10074	14,18	18,18

D=4 (there are four observations, two for each female)

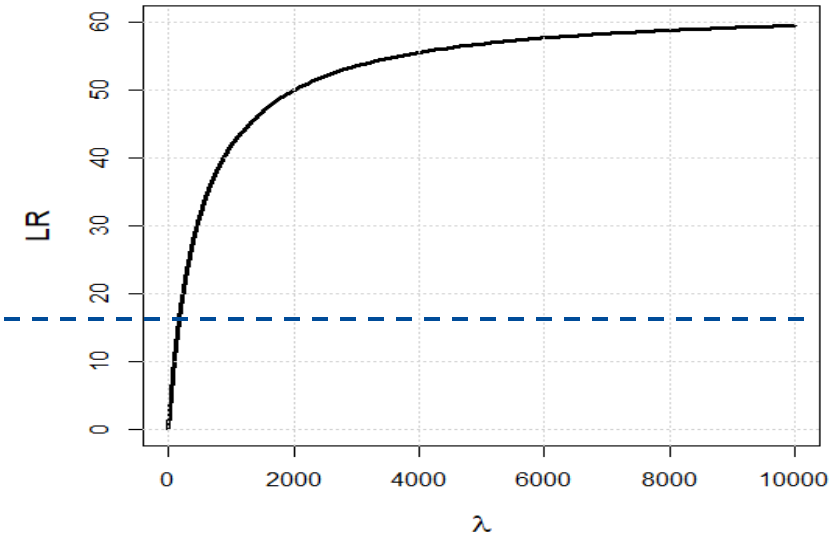
Haplotype	d_i
12 19 14	0.5
12 19 18	0.5
13 19 14	0.5
13 19 18	0.5
13 18 18	0.5
13 19 18	0.5
14 18 18	0.5
14 19 18	0.5

New frequency model

With pseudocounts



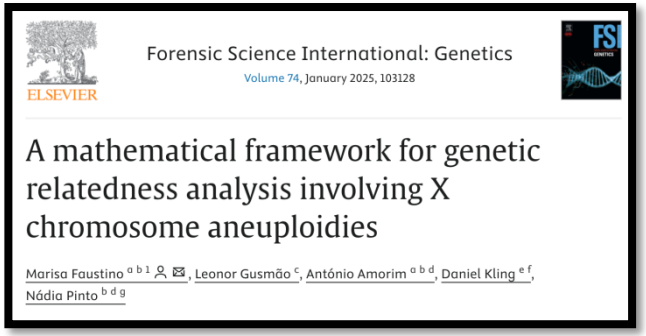
Without pseudocounts



New frequency model

- Model does not change the haplotype frequencies in the database
- Adjusts the frequencies on-the-fly
- More intuitive results
- Mathematically sound?
- Good for smaller databases (say <1000 haplotypes)

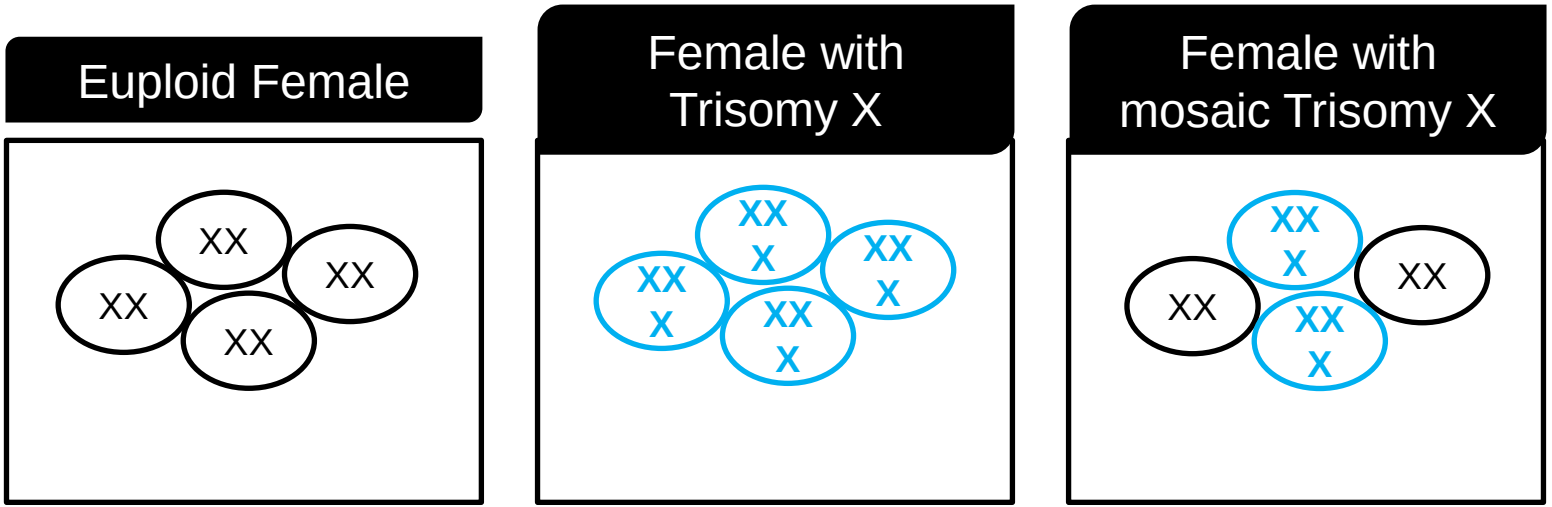
Anepleudies



Big thanks to Marisa Faustino for sharing slides!

Aneuploidy

- Loss or gain of one or more chromosomes



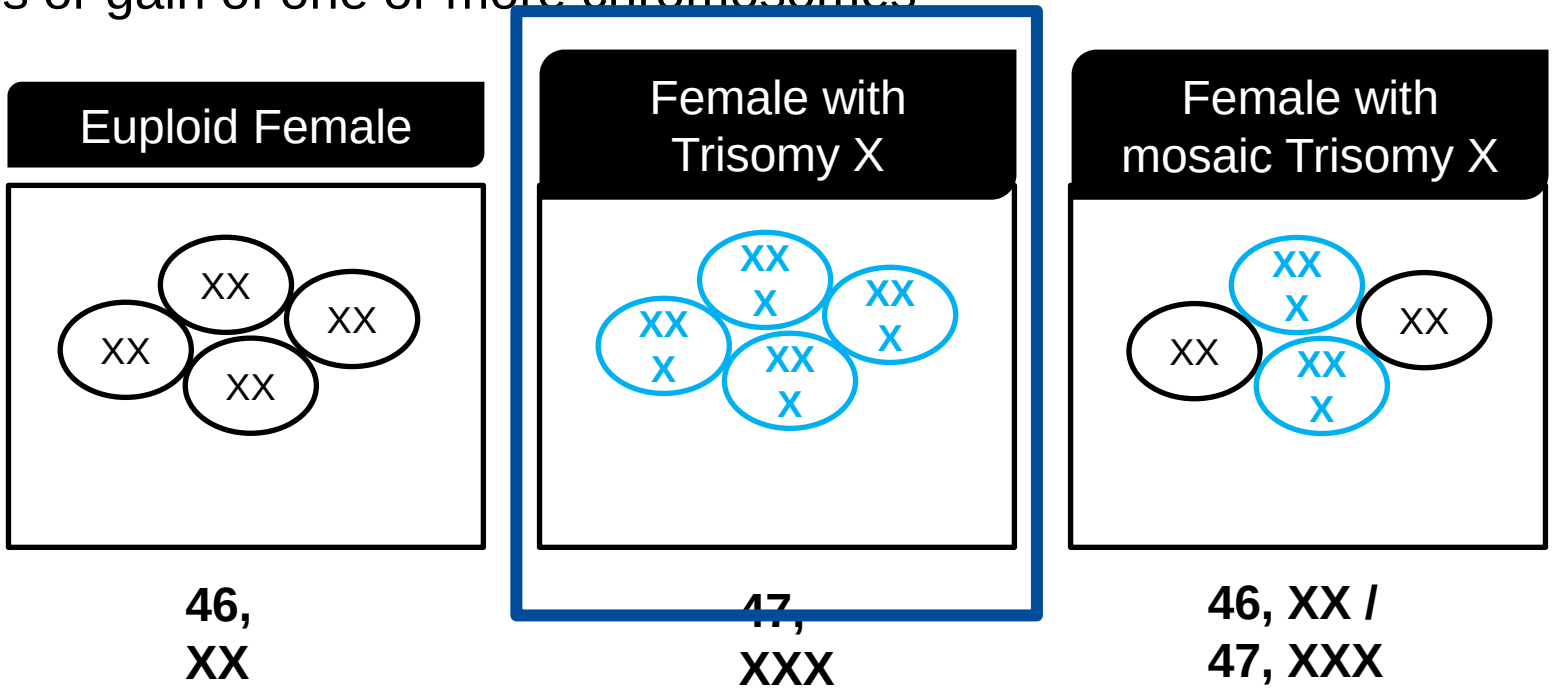
46, XX

47, XXX

46, XX / 47, XXX

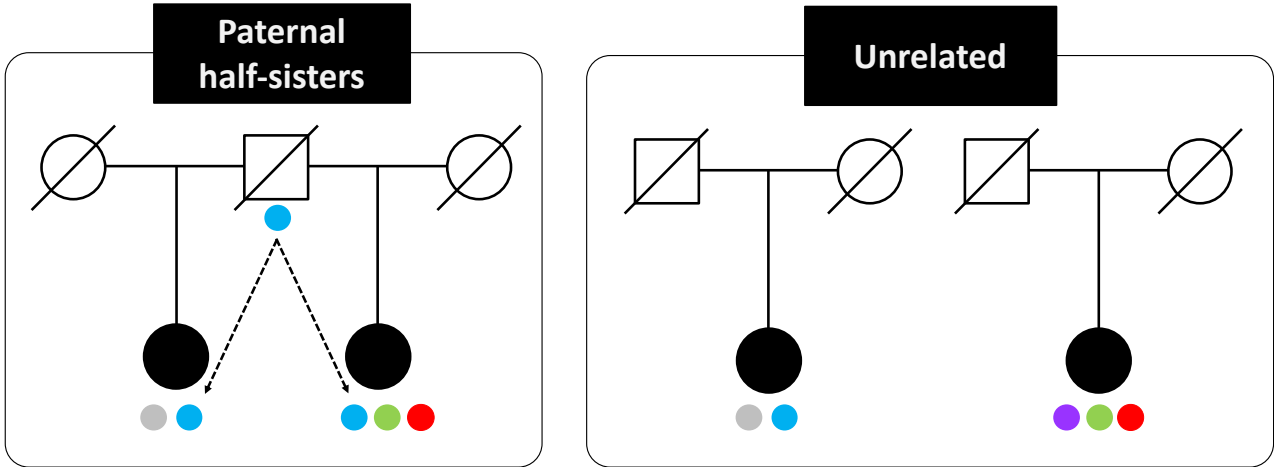
Aneuploidy

- Loss or gain of one or more chromosomes



Kinship Investigation Problem

Hypotheses



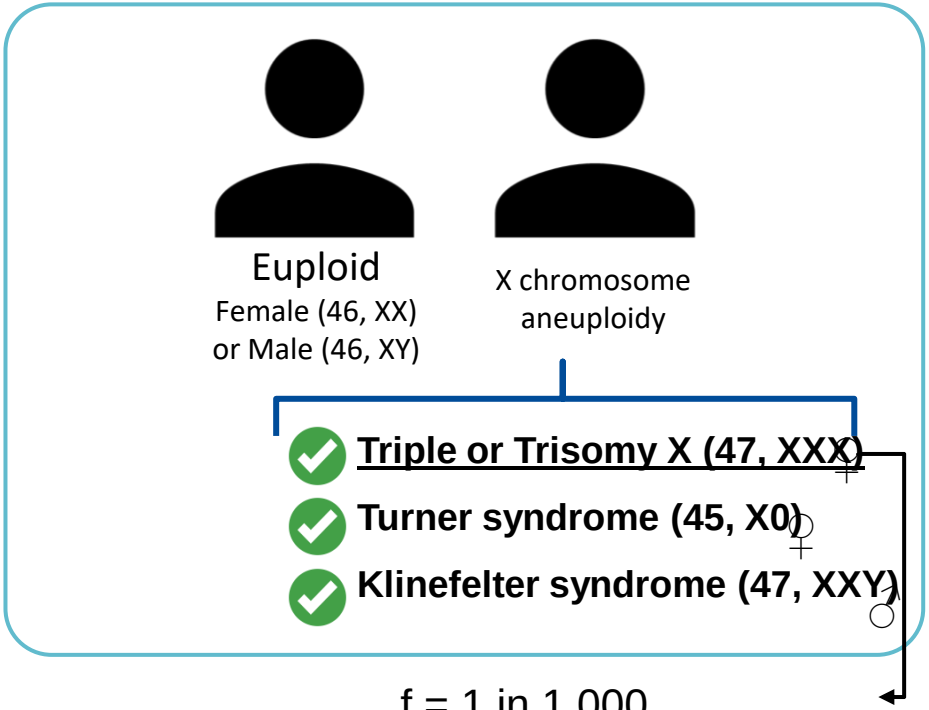
LR = ?

Objective

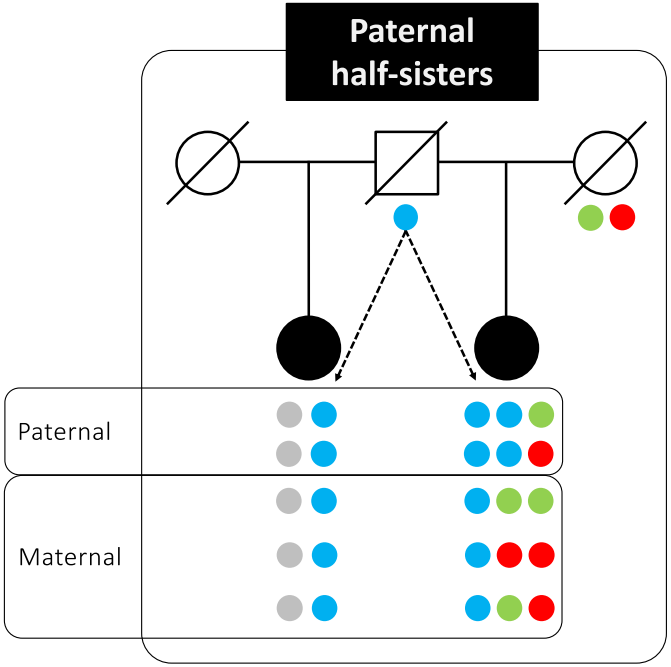
❖ Establish the mathematical framework to weight the DNA evidence of **independent X chromosome markers** in kinship analyses, between 2 **non-inbred** individuals when:

Assumptions:

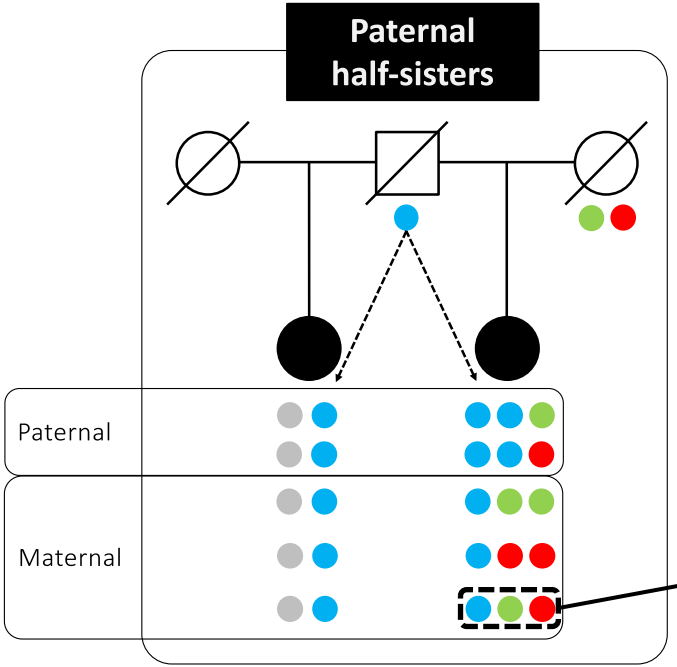
- Parents without aneuploidies
- No allelic mutations
- Full codominance



Kinship Investigation



Kinship Investigation



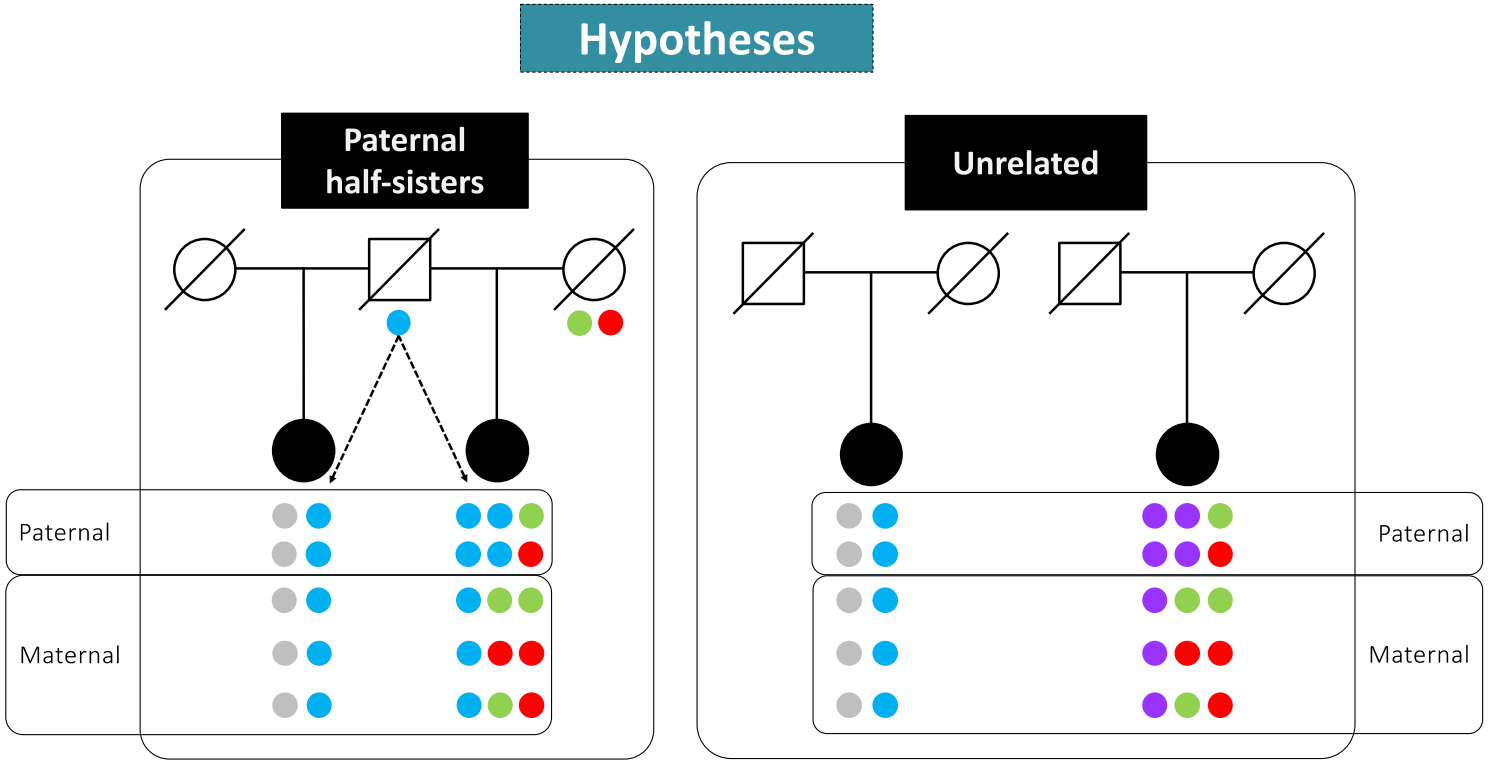
Key points (Trisomy X):

- **Paternal extra X chr:**
  1 pair of IBD alleles (max 2 peaks)

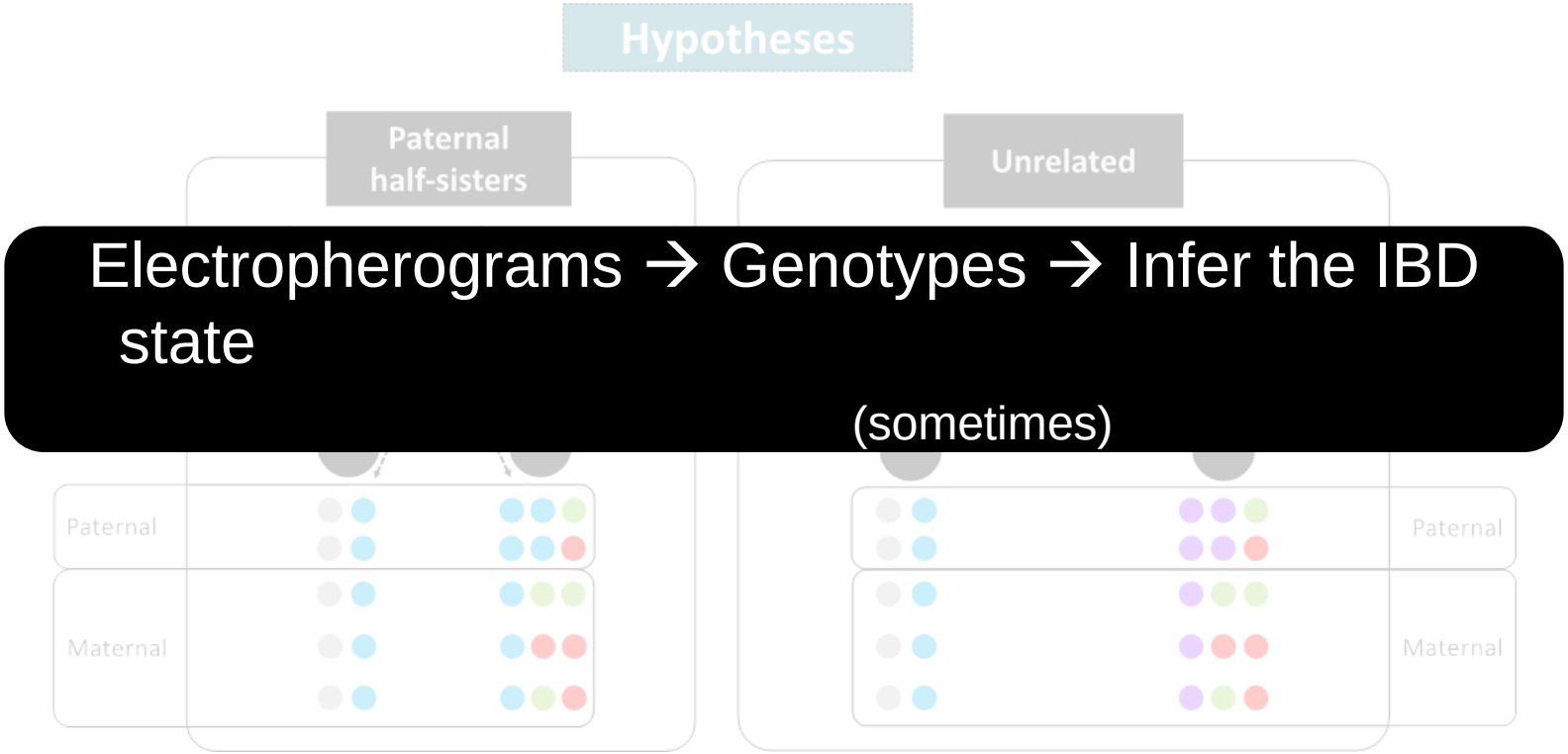
- **Maternal extra X chr:**
  1 pair of IBD or non-IBD alleles (max 3 peaks)

- **3 non-IBD alleles:**
  maternal extra X chr

Kinship Investigation



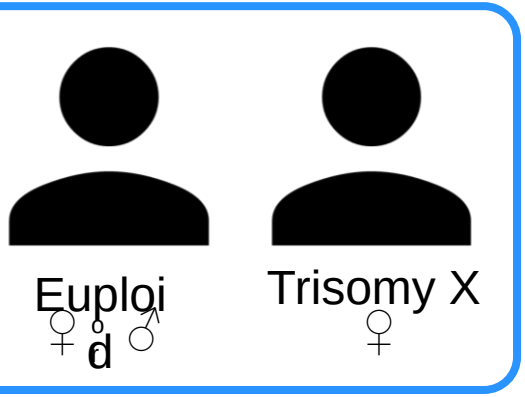
Kinship Investigation

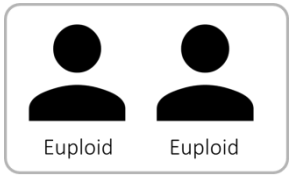


Electropherograms → Genotypes → Infer the IBD state

(sometimes)

Kinship Evaluation



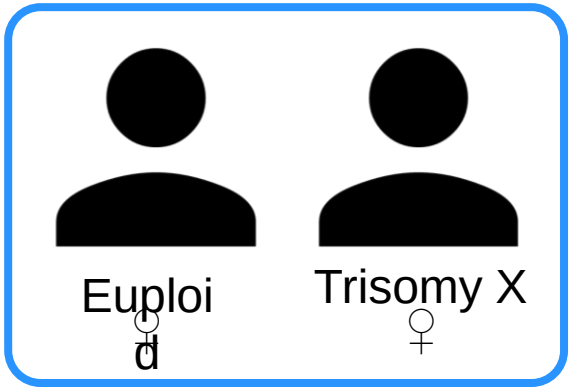


$$LR = \frac{P(genotypes|H_1)}{P(genotypes|H_2)}$$

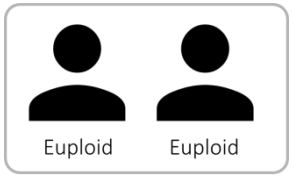
$$LR = \frac{P(extra\ X\ mat)\ P(genotypes|H_1, m) + P(extra\ X\ pat)\ P(gentoypes|H_1, \overline{m})}{P(extra\ X\ mat)\ P(genotypes|H_2, m) + P(extra\ X\ pat)\ P(gentoypes|H_2, \overline{m})}$$

LR : Likelihood ratio
H : Hypothesis i
m : maternal extra X
$\overline{m}$: paternal extra X

Kinship Evaluation



?

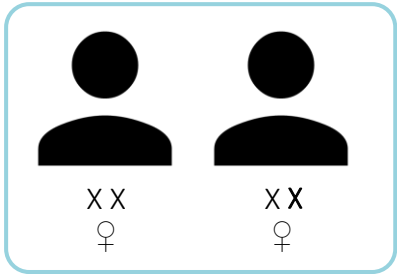


$$LR = \frac{P(genotypes|H_1)}{P(genotypes|H_2)}$$

$$LR = \frac{P(extra\ X\ mat)\ P(genotypes|H_1, m) + P(extra\ X\ pat)\ P(gentoypes|H_1, \overline{m})}{P(extra\ X\ mat)\ P(genotypes|H_2, m) + P(extra\ X\ pat)\ P(gentoypes|H_2, \overline{m})}$$

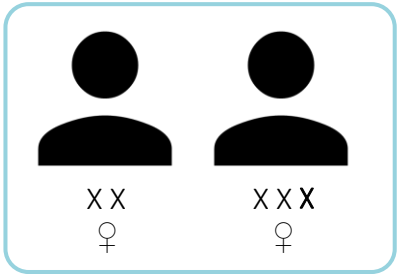
LR : Likelihood ratio
H : Hypothesis i
m : maternal extra X
$\overline{\text{m}}$: paternal extra X

Genotypic configurations



9 possibilities

- aa, aa
- aa, bb
- aa, ab
- ab, aa
- aa, bc
- bc, aa
- ab, ab
- ab, ac
- ab, cd

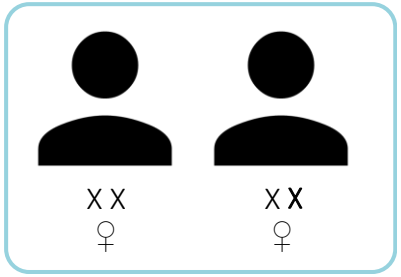


16 possibilities

- aa, aaa
- aa, aab
- aa, abb
- aa, abc
- aa, bbb
- aa, bbc
- aa, bcd
- ab, aab
- ab, abc
- ab, aaa
- ab, aac
- ab, acc
- ab, acd
- ab, ccc
- ab, ccd
- ab, cde

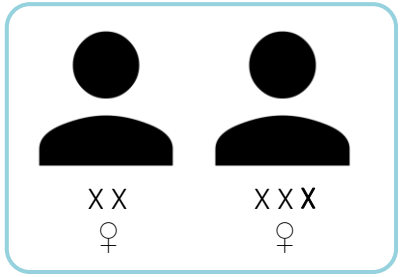
Pinto et al., (2011)

Genotypic configurations



9 possibilities

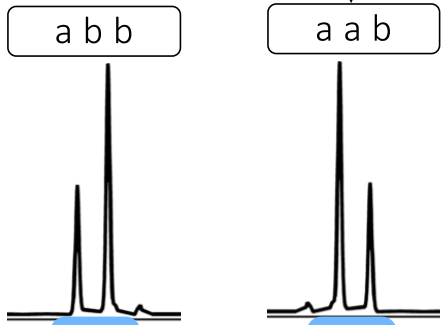
- *aa, aa*
- *aa, bb*
- *aa, ab*
- *ab, aa*
- *aa, bc*
- *bc, aa*
- *ab, ab*
- *ab, ac*
- *ab, cd*



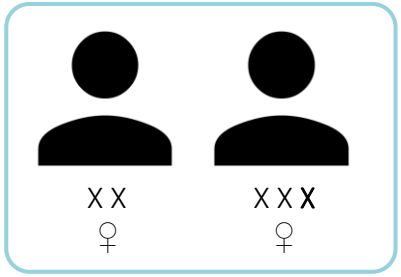
16 possibilities

Pinto et al., (2011)

- *aa, aaa*
- *aa, aab*
- *aa, abb*
- *aa, abc*
- *aa, bbb*
- *aa, bbc*
- *aa, bcd*
- *ab, aab*
- *ab, abc*
- *ab, aaa*
- *ab, aac*
- *ab, acc*
- *ab, acd*
- *ab, ccc*
- *ab, ccd*
- *ab, cde*



Joint Genotypic Probabilities



- aa, aaa
- aa, aab
- aa, abb
- aa, abc
- aa, bbb
- aa, bbc
- aa, bcd
- ab, aab
- ab, abc
- ab, aaa
- ab, aac
- ab, acc
- ab, acd
- ab, ccc
- ab, ccd
- ab, cde

P (IBD arrangement 1)

$$P(aa, abc) = 6\varphi_1 {f_a}^3 f_b f_c + 2\varphi_3 {f_a}^2 f_b f_c$$

frequency of allele a

6 ways to arrange the alleles considering IBD arrangement 1

(

...

IBD arrangements

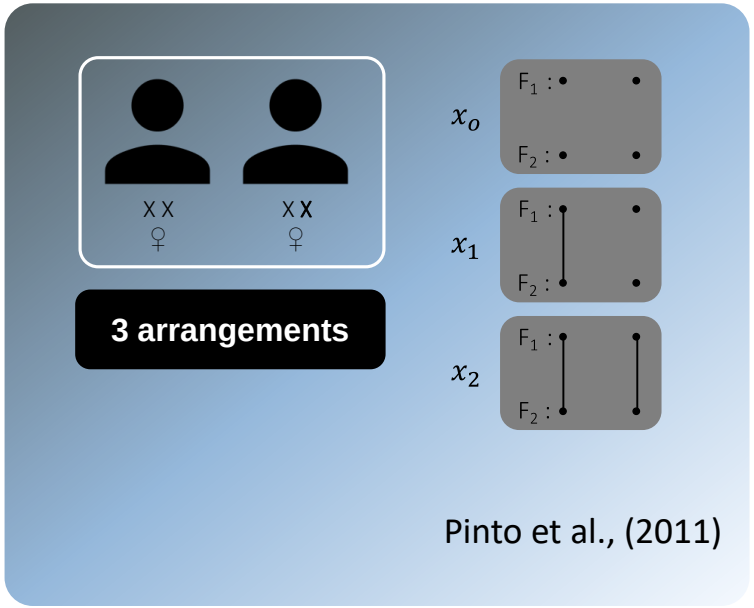
The IBD arrangements represent the different possibilities of individuals sharing pairs of IBD alleles per marker

IBD arrangements

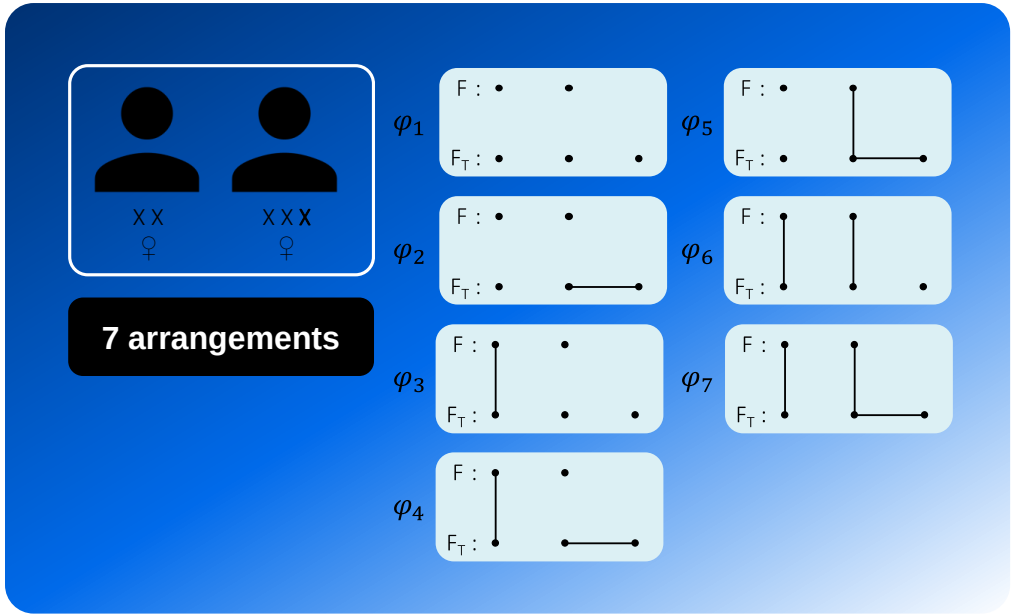
The IBD arrangements represent the different possibilities of individuals sharing pairs of IBD alleles per marker

- **non-inbred**

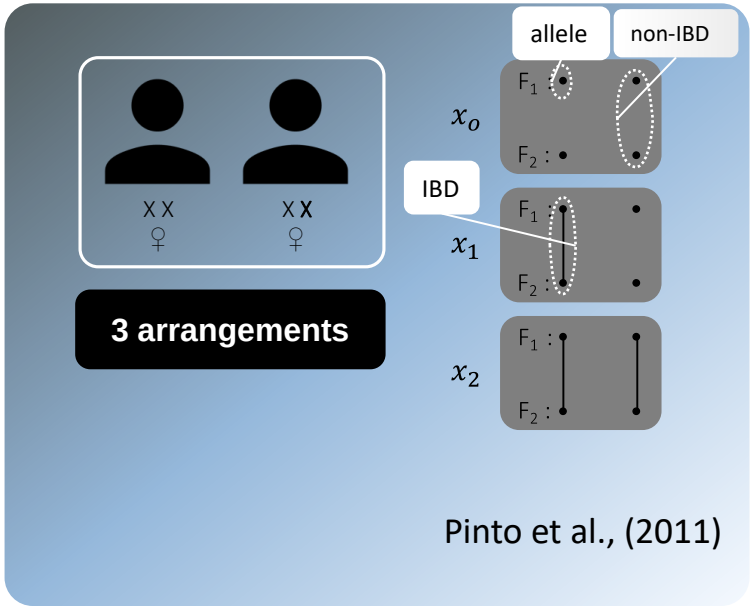
IBD arrangements



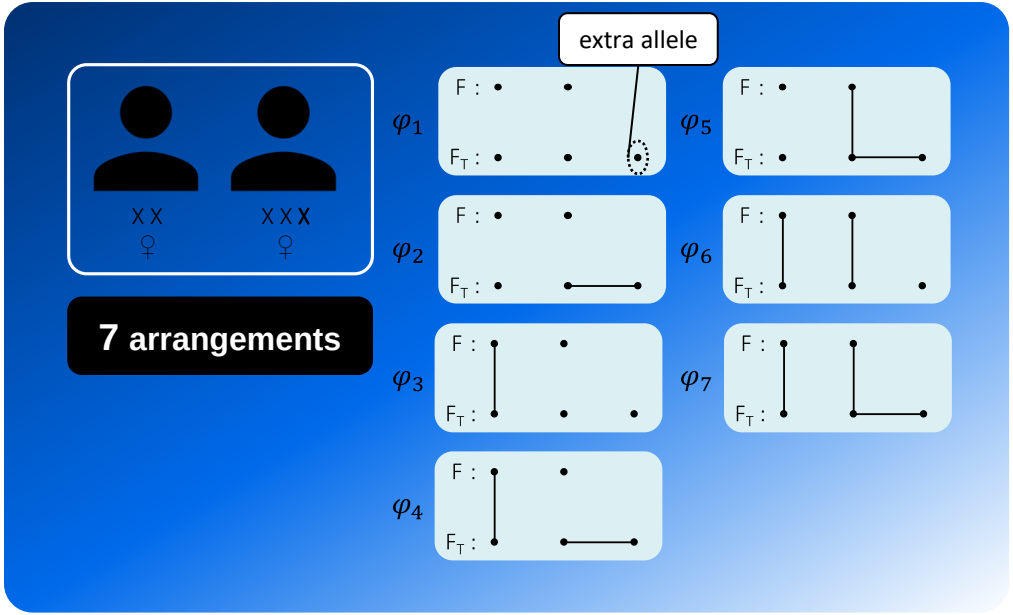
Pinto et al., (2011)



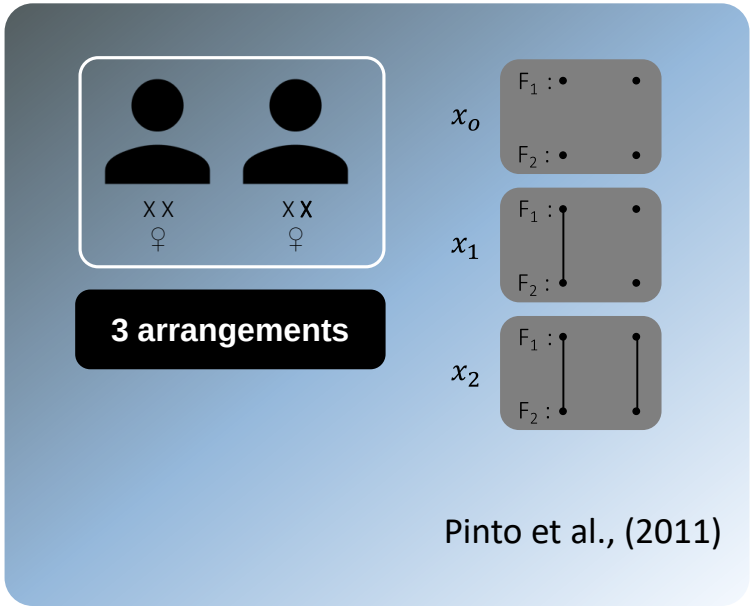
IBD arrangements



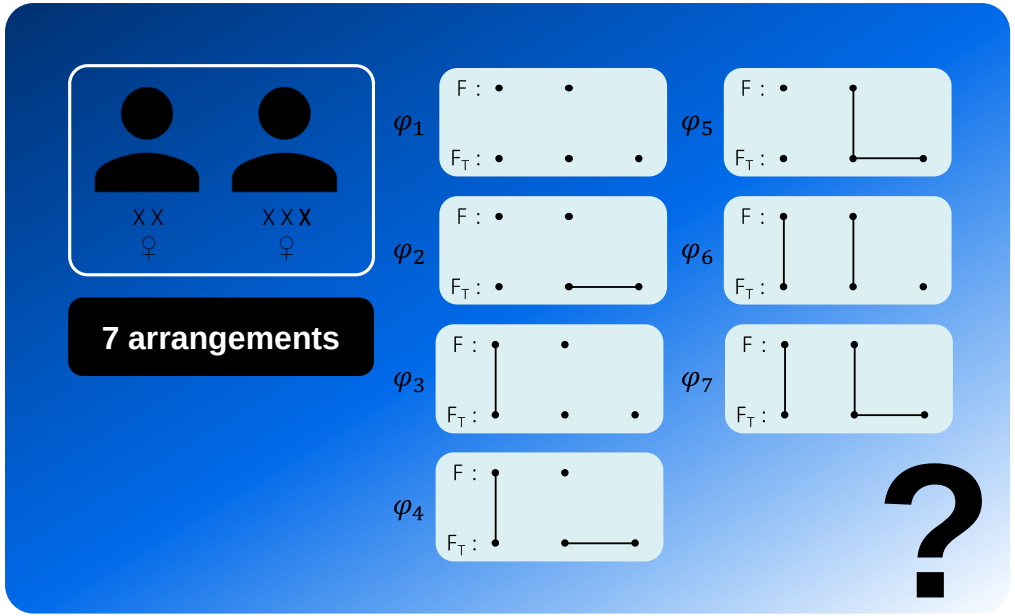
Pinto et al., (2011)



IBD arrangements

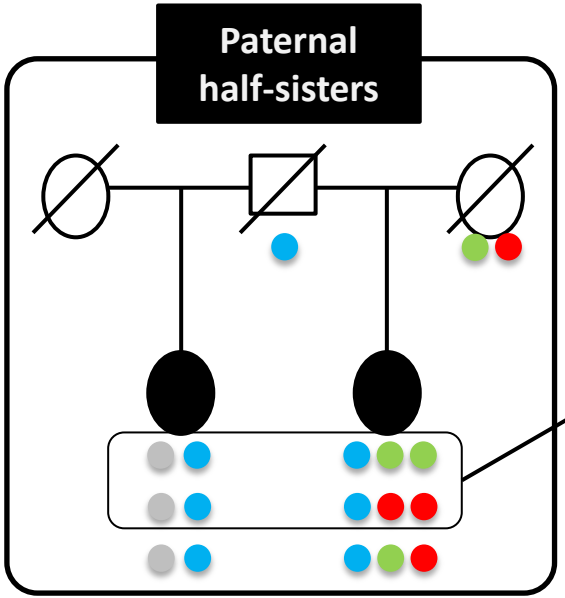


Pinto et al., (2011)



IBD arrangements probabilities

When the extra X chromosome is maternal:

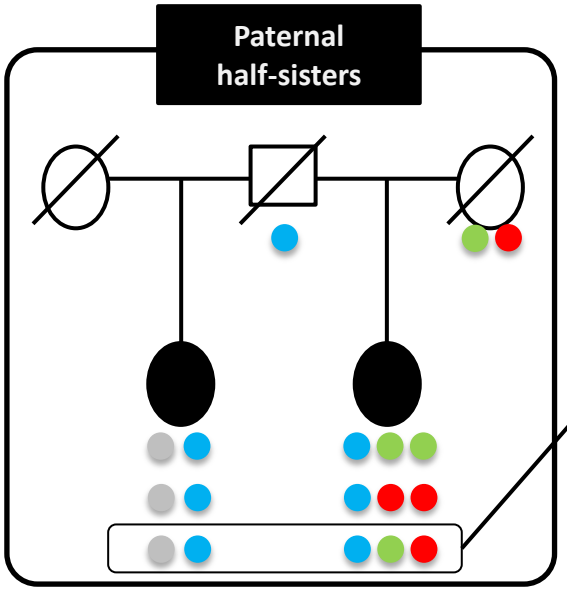


φ_1	F, F_T	0
φ_2	F, F_T	
φ_3	F, F_T	$1 - P(x)$
φ_4	F, F_T	$P(x)$
φ_5	F, F_T	0
φ_6	F, F_T	
φ_7	F, F_T	

$P(x)$: Probability of the maternal alleles being IBD

IBD arrangements probabilities

When the extra X chromosome is maternal:

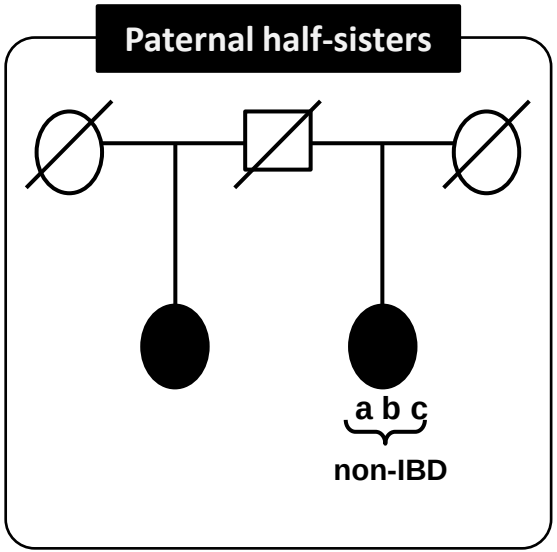


φ_1	F, F_T	0
φ_2	F, F_T	0
φ_3	F, F_T	$1 - P(x)$
φ_4	F, F_T	$P(x)$
φ_5	F, F_T	0
φ_6	F, F_T	0
φ_7	F, F_T	0

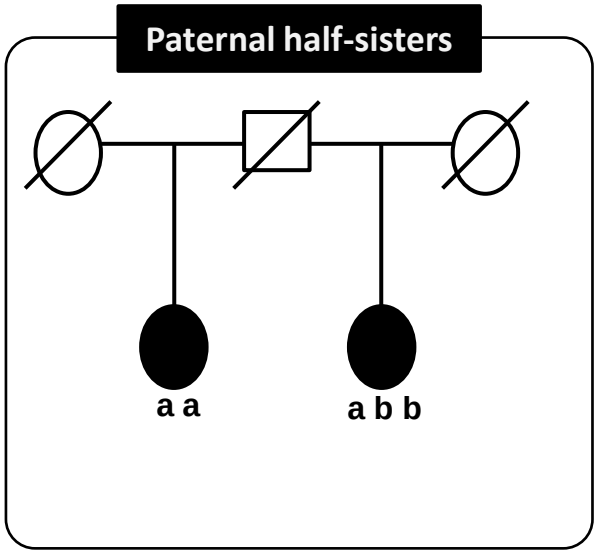
$P(x)$: Probability of the maternal alleles being IBD

P(x) inference

When the extra X chromosome is maternal:

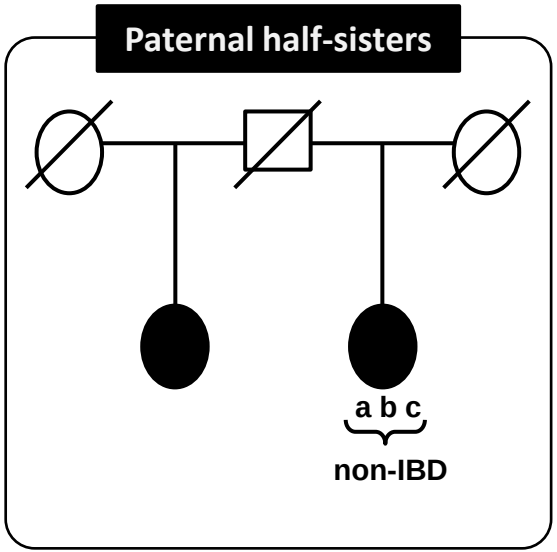


$$P(x) = 0$$

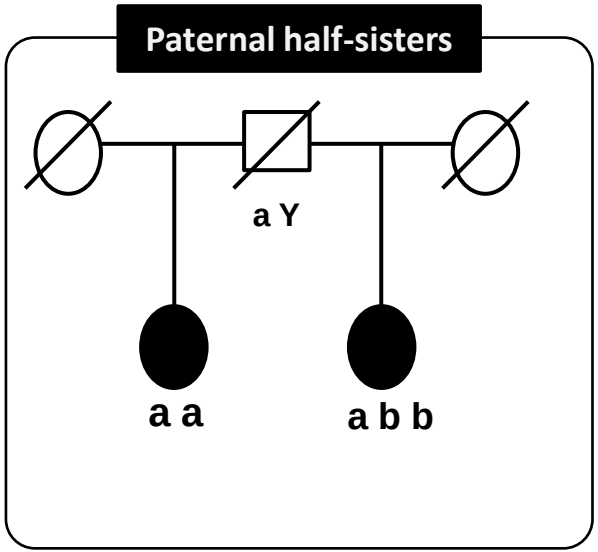


P(x) inference

When the extra X chromosome is maternal:

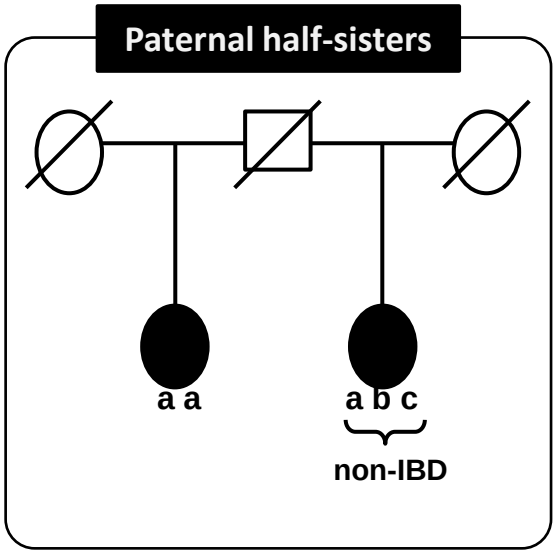


$$P(x) = 0$$

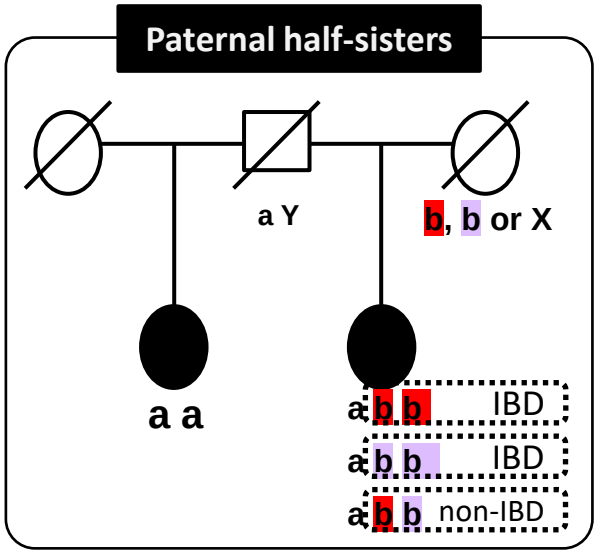


P(x) inference

When the extra X chromosome is maternal:



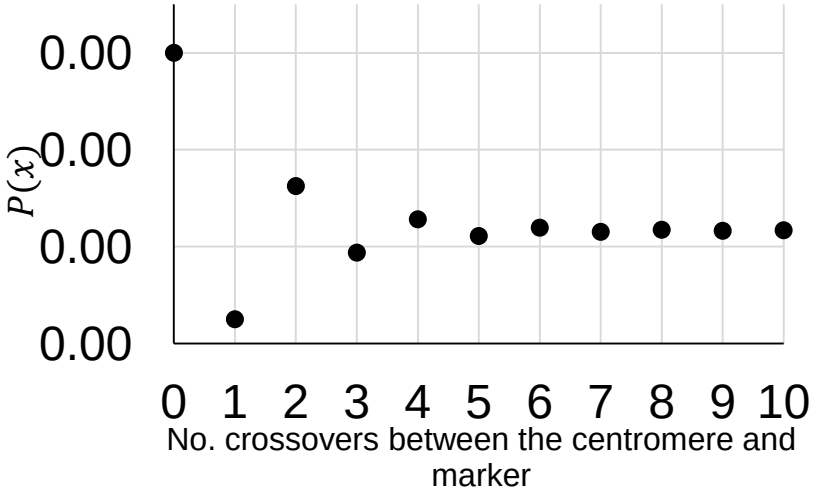
$P(x) = 0$



$P(x) = ?$

P(x) estimation

- Graph made from an equation adapted from Côté and Edwards, (1975)
- Considering P(meiosis I errors) = 0.63 (Thomas et al. (2001))



Depends on:

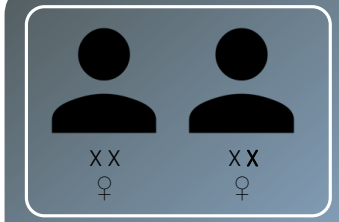
- Number of crossovers between the centromere and marker

$P(x)$:

- ~0.37 for pericentromeric markers
- ~0.33 for non-pericentromeric markers

$P(x)$: Probability of the maternal alleles inherited being IBD

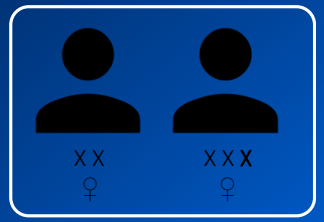
IBD arrangements probabilities



	Pat half-sisters	Unrelated
x_o	0	0
x_1	1	0
x_2	0	1

Pinto et al., (2011)

- Dependent on kinship
- Independent of genotypes

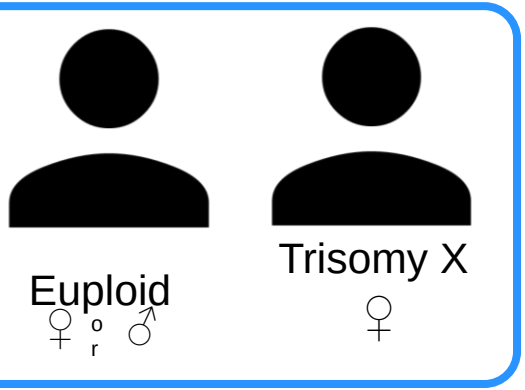


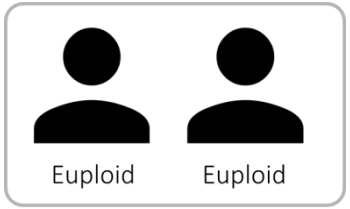
Maternal extra X chromosome

	Pat half-sisters	Unrelated
φ_1	0	$1 - P(x)$
φ_2		$P(x)$
φ_3	$1 - P(x)$	0
φ_4	$P(x)$	
φ_5	0	
φ_6		
φ_7		

- Dependent on the kinship & P(x) value & genotypes

Kinship Evaluation





$$LR = \frac{P(genotypes|H_1)}{P(genotypes|H_2)}$$

$$LR = \frac{P(extra\ X\ mat)\ P(genotypes|H_1, m) + P(extra\ X\ pat)\ P(gentoypes|H_1, \overline{m})}{P(extra\ X\ mat)\ P(genotypes|H_2, m) + P(extra\ X\ pat)\ P(gentoypes|H_2, \overline{m})}$$

LR : Likelihood ratio
H : Hypothesis i
m : maternal extra X
$\overline{m}$: paternal extra X

ELSEVIER
Forensic Science International: Genetics
Volume 74, January 2025, 103128

A mathematical framework for genetic relatedness analysis involving X chromosome aneuploidies

Marisa Faustino [a b 1], Leonor Gusmão [c], António Amorim [a b d], Daniel Kling [e f], Nádia Pinto [b d g]

H_1: The females are related as paternal half-sisters
H_2: The females are unrelated

Maternal meiotic error inferred as F_T shows 3 different alleles in DXS10135: 17,18,19

IBD probabilities (Table 1A):
H_1: $\varphi^m_{1,2,5-7} = 0$; $\varphi^m_3 = 1 - P(x|H_1, G)$; $\varphi^m_4 = P(x|H_1, G)$
H_2: $\varphi^m_1 = 1 - P(x|H_2, G)$; $\varphi^m_2 = P(x|H_1, G)$; $\varphi^m_{5-7} = 0$

		M_1: DXS10135	M_2: DXS10073	M_3: DXS10101	M_4: DXS10146
$G_{j=1,..4}$		F: 17,17 F_T: 17,18,19 F: aa; F_T: abc	F: 20,20 F_T: 20,20,20 F: aa; F_T: aaa	F: 28,31 F_T: 28,28,28 F: ab; F_T: aaa	F: 26,30 F_T: 26,26,30 F: ab; F_T: aab
H_1	Table 2A	$P(x\|H_1, G_1) = 0$ $\varphi^m_4 = 0$ $\varphi^m_3 = 1$	$P(x\|H_1, G_2) = z$ $\varphi^m_3 = 1 - z$ $\varphi^m_4 = z$	$P(x\|H_1, G_3) = z$ $\varphi^m_3 = 1 - z$ $\varphi^m_4 = z$	$P(x\|H_1, G_4) = z$ $\varphi^m_3 = 1 - z$ $\varphi^m_4 = z$
	Table 3A	$P(G_1\|H_1) =$ $= 2f^2_{17}f_{18}f_{19}$	$P(G_2\|H_1) =$ $= (1-z)f^4_{20} + zf^3_{20}$	$P(G_3\|H_1) =$ $= (1-z)f^3_{28}f_{31}$ $+ zf^2_{28}f_{31}$	$P(G_4\|H_1) =$ $= (1-z)(2f^2_{26}f^2_{30}$ $+ f^3_{26}f_{30}) + zf^2_{26}f_{30}$
H_2	Table 2A	$P(x\|H_2, G_1) = 0$ $\varphi^m_1 = 1$ $\varphi^m_2 = 0$	$P(x\|H_2, G_2) = z$ $\varphi^m_1 = 1 - z$ $\varphi^m_2 = z$	$P(x\|H_2, G_3) = z$ $\varphi^m_1 = 1 - z$ $\varphi^m_2 = z$	$P(x\|H_2, G_4) = z$ $\varphi^m_1 = 1 - z$ $\varphi^m_2 = z$
	Table 3A	$P(G_1\|H_2) =$ $=6f_{17}{}^3f_{18}f_{19}$	$P(G_2\|H_2) =$ $= (1-z)f^5_{20} + zf^4_{20}$	$P(G_3\|H_2) =$ $= 2(1-z)f^4_{28}f_{31}$ $+ 2zf^3_{28}f_{31}$	$P(G_4\|H_2) =$ $= 6(1-z)f^3_{26}f^2_{30}$ $+ 2zf^2_{26}f^2_{30}$
$LR_{j=1,..4}$		$\frac{P(G_1\|H_1)}{P(G_1\|H_2)}$	$\frac{P(G_2\|H_1)}{P(G_2\|H_2)}$	$\frac{P(G_3\|H_1)}{P(G_3\|H_2)}$	$\frac{P(G_4\|H_1)}{P(G_4\|H_2)}$
LR_{Total}		$\prod_{j=1}^{4} LR_j = 1.13 \times 10^5$			

Expanded marker panels

FORCE

- Different kits

STRs

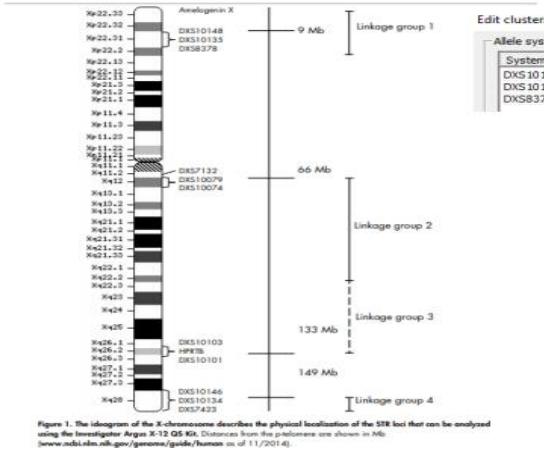
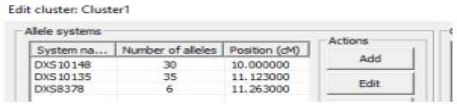


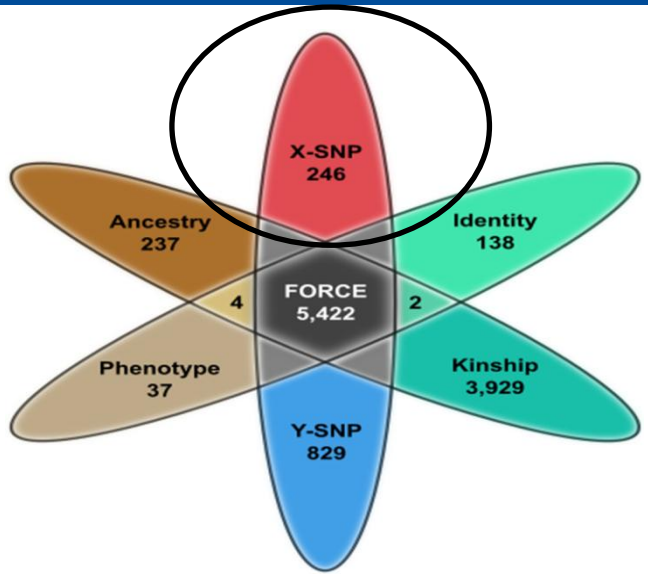
Figure 1. The ideogram of the X-chromosome describes the physical localization of the STR loci that can be analyzed using the Investigator Argus X-12 QS Kit. Distances from the p-telomere are shown in Mb (www.ncbi.nlm.nih.gov/genome/guide/human as of 11/2014).



1. Many alleles
2. Mutations
3. LD

FORCE

	Marker	Genetic position (cM)
2	rs4892897	7.582
3	rs1637781	7.805
4	rs5983084	8.5874
5	rs6642174	8.8134
6	rs6641753	9.1834
7	rs6641574	9.4251
8	rs2058865	9.7825
9	rs5962087	11.3138
10	rs5915796	12.2585
11	rs5916138	12.71
12	rs5915672	13.1805
13	rs6529997	13.9103
14	rs1637788	15.1505
15	rs4240138	15.6932
16	rs2108400	16.1319
17	rs5933710	17.0167
18	rs845444	17.6324
19	rs768568	17.9229
20	rs929217	18.1746



Open Access Article

The FORCE Panel: An All-in-One SNP Marker Set for Confirming Investigative Genetic Genealogy Leads and for General Forensic Applications

by Andreas Tillmar [1,2,*], Kimberly Sturk-Andreaggi [3,4,5], Jennifer Daniels-Higginbotham [3,4], Jacqueline Tyler Thomas [3,4] and Charla Marshall [3,4,6,*]

[1] Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, SE-587 58 Linköping, Sweden
[2] Department of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences, Linköping University, SE-582 25 Linköping, Sweden
[3] Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), Dover Air Force Base, Dover, DE 19902, USA
[4] SNA International, LLC, Contractor Supporting the AFMES-AFDIL, Alexandria, VA 22314, USA
[5] Department of Immunology, Genetics and Pathology, Uppsala University, SE-751 08 Uppsala, Sweden

X-SNPs are chosen to be have no significant LD

Kintelligence (ForenSeq)

- Includes 106 X-SNPs

RESEARCH ARTICLE · Volume 71, 103055, July 2024 · Open Access · Download Full Issue

Developmental validation of the ForenSeq® Kintelligence kit, MiSeq FGx® sequencing system and ForenSeq Universal Analysis Software

Joana Antunes [a] · Paulina Walichiewicz [a] · Elmira Forouzmand [a] · ... · Swathi A. Kumar [a] · Bruce Budowle [b,c] · Kathryn M. Stephens [a,*]

... Show more

Affiliations & Notes · Article Info

Download PDF · Cite · Share · Set Alert · Get Rights · Reprints · Previous article · Next article

Show Outline

Highlights

- ForenSeq Kintelligence kit, Universal Analysis Software, and MiSeq FGx system uses targeted Next-Generation Sequencing.
- 10,230 forensic SNP-loci in a multiplex reaction of short amplicons: kinship, identity, hair and eye color, and biogeographical ancestry.
- Short amplicons for more data from low input and degraded samples.
- SWGDAM developmental validation studies included PCR conditions, sensitivity, stability, MPS, mixtures, case-type samples, and species specificity.

Abstract

Forensic Investigative Genetic Genealogy, a recent sub discipline of forensic genomics, leverages the high throughput and sensitivity of detection of next generation sequencing and established genetic and genealogical approaches to support the identification of human remains from missing persons investigations and investigative lead generation in violent crimes. To facilitate forensic DNA evidence analysis, the ForenSeq® Kintelligence multiplex, consisting of 10,230 SNPs, was developed. Design of the ForenSeq Kintelligence Kit, the MiSeq FGx® Sequencing System and the ForenSeq Universal Analysis Software is described. Developmental validation in accordance with SWGDAM guidelines and forensic quality assurance standards, using single source samples, is reported for the end-to-end workflow from library preparation to data interpretation. Performance metrics support the conclusion that more genetic information can be obtained from challenging samples compared to other commercially available forensic targeted DNA assays developed for capillary electrophoresis (CE) or other current next generation sequencing (NGS) kits due to the higher number of markers, the overall shorter amplicon sizes (97.8% <250 bp), and kit design. Data indicate that the multiplex is robust and fit for purpose for a wide range of quantity and quality samples. The ForenSeq Kintelligence Kit and the Universal Analysis Software allow transfer of the genetic component of forensic investigative genetic genealogy to the operational forensic laboratory.

Figures (6) · Figure Viewer

Show all figures

Article metrics

9 Citations · 17 Captures

PLUMX · View details

Forensic Sci Int Genet. 2022 Nov:61:102769. doi: 10.1016/j.fsigen.2022.102769. Epub 2022 Aug 27.

Fast and accurate kinship estimation using sparse SNPs in relatively large database searches

June Snedecor [1], Tim Fennell [2], Seth Stadick [3], Nils Homer [4], Joana Antunes [5], Kathryn Stephens [6], Cydne Holt [7]

Affiliations + expand

PMID: 36087514 DOI: 10.1016/j.fsigen.2022.102769

Free article

Abstract

Forensic genetic genealogy (FGG) has primarily relied upon dense single nucleotide polymorphism (SNP) profiles from forensic samples or unidentified human remains queried against online genealogy database(s) of known profiles generated with SNP microarrays or from whole genome sequencing (WGS). In these queries, SNPs are compared to database samples by locating contiguous stretches of shared SNP alleles that allow for detection of genomic segments that are identical by descent (IBD) among biological relatives (kinship). This segment-based approach, while robust for detecting distant relationships, generally requires DNA quantity and/or quality that are sometimes not available in forensic casework samples. By focusing on SNPs with maximal discriminatory power and using an algorithm designed for a sparser SNP set than those from microarray typing, performance similar to segment matching was reached even in difficult casework samples. This algorithm locates shared segments using kinship coefficients in "windows" across the genome. The windowed kinship algorithm is a modification of the PC-AiR and PC-Relate tools for genetic relatedness inference, referred to here as the "whole genome kinship" approach, that control for the presence of unknown or unspecified population substructure. Simulated and empirical data in this study, using DNA profiles

FULL TEXT LINKS · ELSEVIER OPEN ACCESS

ACTIONS · Cite · Collections

SHARE

PAGE NAVIGATION · Title & authors · Abstract · Similar articles · Cited by · MeSH terms

MPSplex

- Includes 29 SNPs (tri-allelic)

RESEARCH ARTICLE · Volume 46, 102232, May 2020 · Open Access · Download Full Issue

A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel

C. Phillips · J. Amigo · A.O. Tillmar · ... · Y. Wong · T.J. Parsons · M.V. Lareu · Show more

Affiliations & Notes · Article Info

Download PDF · Cite · Share · Set Alert · Get Rights · Reprints · Previous article · Next article

Show Outline

Highlights

- 271,934 tri-allelic SNPs were identified in the 1000 Genomes Phase III variant catalog and data has been compiled in Mendeley Data for free access.
- From this extensive dataset 8,705 SNPs had heterozygosity values above 0.5 - the maximum value of perfect binary SNPs (0.5:0.5 allele frequencies).
- A large-scale forensic identification multiplex was constructed for MPS, comprising 1,241 autosomal plus 29 X tri-allelic SNPs.
- Approximately 5 % of tri-allelic SNPs selected for the large-scale MPS panel gave three-genotype patterns in one individual or discordant genotypes.
- The need for caution and detailed scrutiny of multiple-allele variant data is highlighted when designing future forensic SNP panels.

Abstract

In a directed search of 1000 Genomes Phase III variation data, 271,934 tri-allelic single nucleotide polymorphisms (SNPs) were identified amongst the genotypes of 2,504 individuals from 26 populations. The majority of tri-allelic SNPs have three nucleotide substitution-based alleles at the same position, while a much smaller proportion, which we did not compile, have a nucleotide insertion/deletion plus substitution alleles. SNPs with three alleles have higher discrimination power than binary loci but keep the some characteristic of optimum amplification of the fragmented DNA found in highly degraded forensic samples. Although most of the tri-allelic SNPs identified had one or two alleles at low frequencies, often single observations, we present a full compilation of the genome positions, rs-numbers and genotypes of all tri-allelic SNPs detected by the 1000 Genomes project from the more detailed analyses it applied to Phase III sequence data. A total of 8,705 tri-allelic SNPs had overall heterozygosities (averaged across all 1000 Genomes populations) higher than the binary SNP maximum value of 0.5. Of these, 1,637 displayed the highest average heterozygosity values of 0.6–0.666. The most informative tri-allelic SNPs we identified were used to construct a large-scale

Figures (9) · Figure Viewer

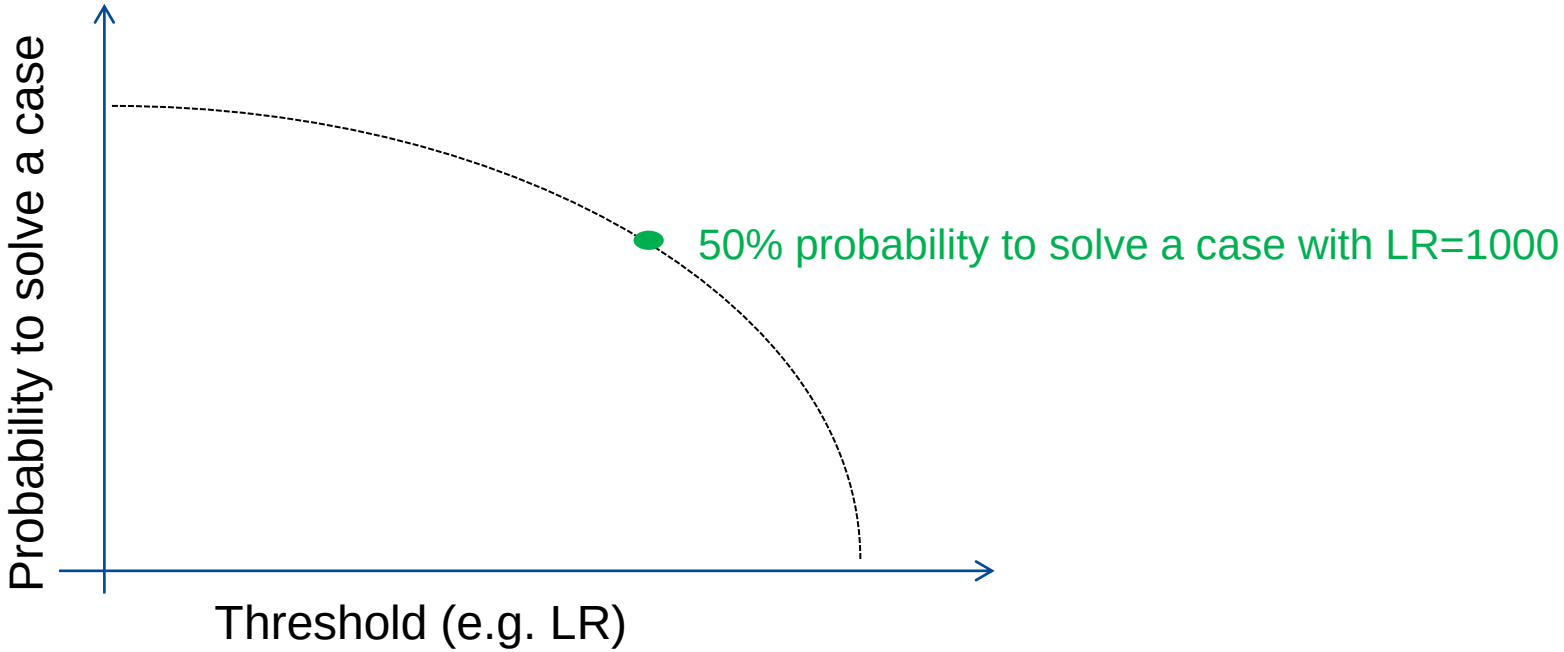
Show all figures

Article metrics

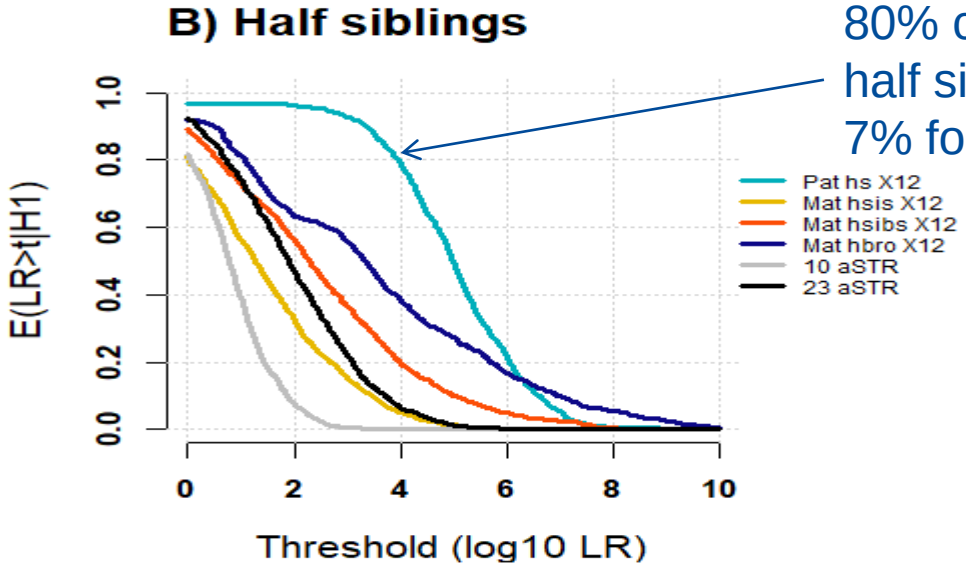
46 Citations · 67 Captures

PLUMX · View details

Exceedance plots



Utility of simulations (Argus X12)

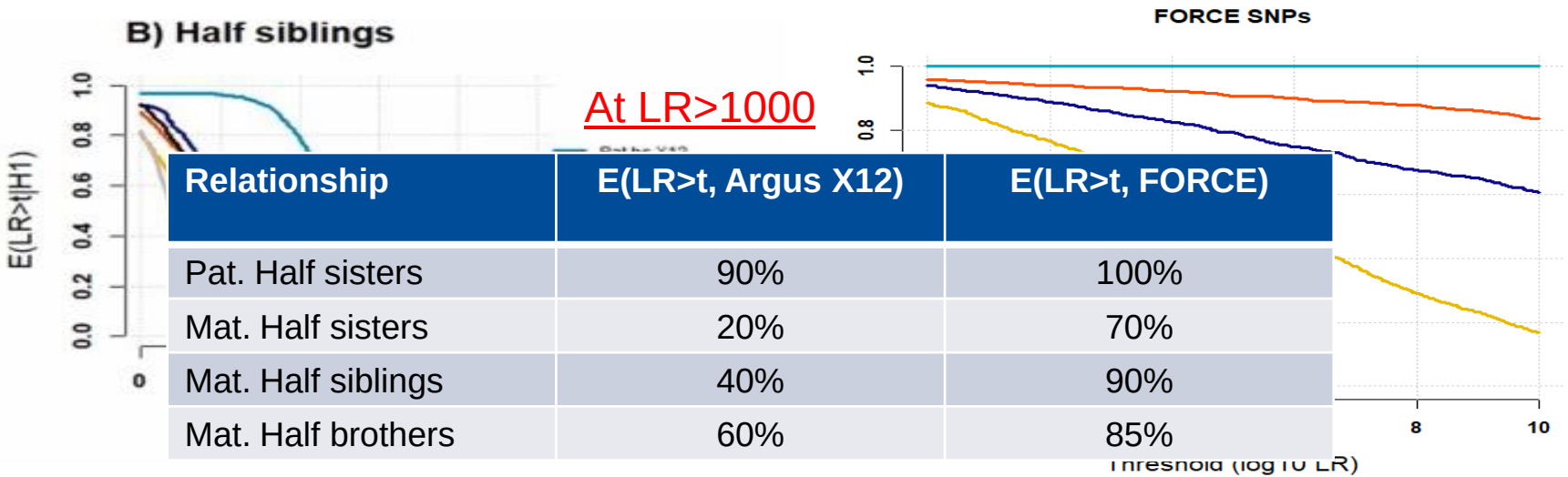


80% chance to solve paternal half siblings with X12 but only 6-7% for 23 aSTR if LR=10,000

Probability that we can solve a case with the given threshold

FORCE X-SNPs

242 SNPs (only linkage, no LD)



At LR>1000

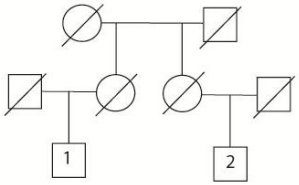
Relationship	E(LR>t, Argus X12)	E(LR>t, FORCE)
Pat. Half sisters	90%	100%
Mat. Half sisters	20%	70%
Mat. Half siblings	40%	90%
Mat. Half brothers	60%	85%

From Bergseth et al. 2022

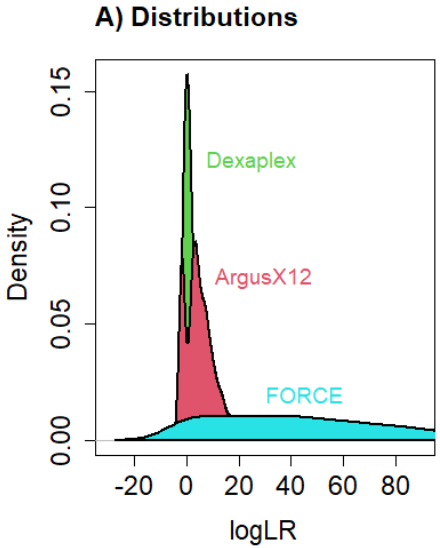
Power increased substantially

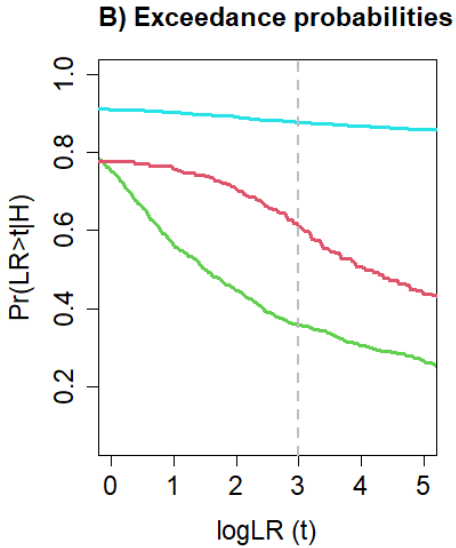
FORCE X-SNPs

Maternal cousins



IBD0=0.625
IBD1=0.375

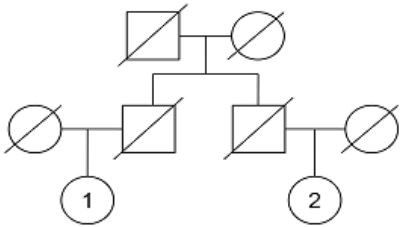




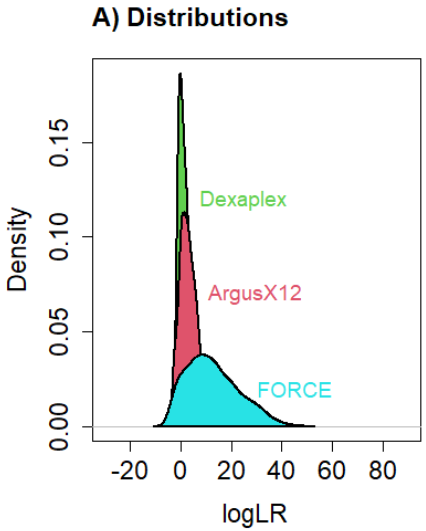
E(LR>1000, Argus X12)	E(LR>1000, FORCE)	E(LR>1000, Decaplex)
61%	88%	36%

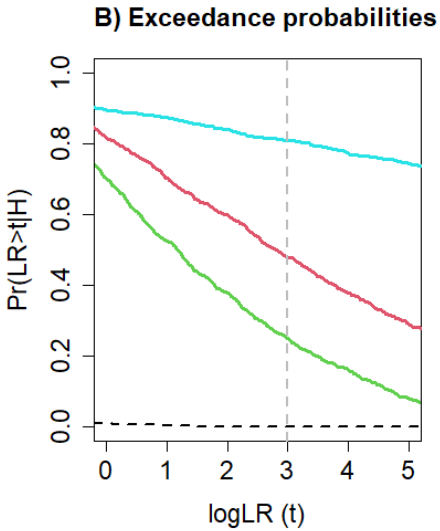
FORCE X-SNPs

Paternal cousins



IBD0=0.5
IBD1=0.5





E(LR>1000, Argus X12)	E(LR>1000, FORCE)	E(LR>1000, Decaplex)
48%	81%	25%

Advanced topics

DANIEL KLING & ANDREAS TILLMAR

Daniel.kling@rmv.se

Andreas.tillmar@rmv.se