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 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

X-Decaplex
ArgusX12

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-chromosom
17:00-17:10	Short break
17:10-18:00	Software: FamLinkX
18:05-18:40	Exercises
18:40-19:00	Summary

Session 3 – Applications and examples (March 24)

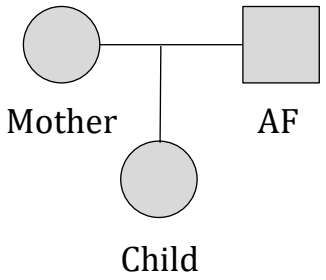
16:00	Introduction
16:15-17:00	Summary of theory and some more advanced topics
17:00-17:10	Short break
17:10-18:00	Examples
18:05-18:40	Exercises
18:40-19:00	Summary

Solving relationship issues with DNA data

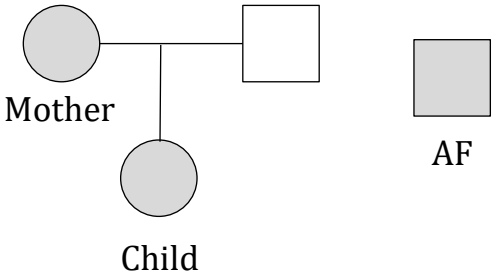
Legal situations: (e.g.) paternity. immigration. missing person identification. criminal acts (incest. human trafficking). investigative leads and more

Example 1 ”Simple” question

H_1: AF is the father of the child



H_2: AF is unrelated to the child

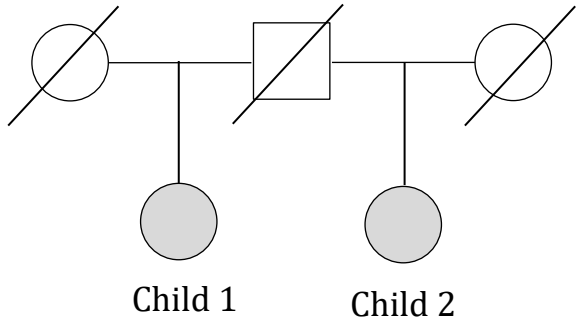


Question: Is AF the biological father of the child?

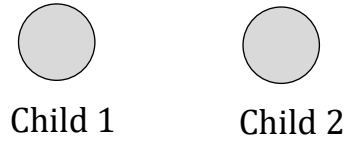
Genetic data: 15-21 autosomal STRs

Example 2 ”More complex” question

H_1: Child 1 and 2 have the same father



H_2: Child 1 and 2 are unrelated

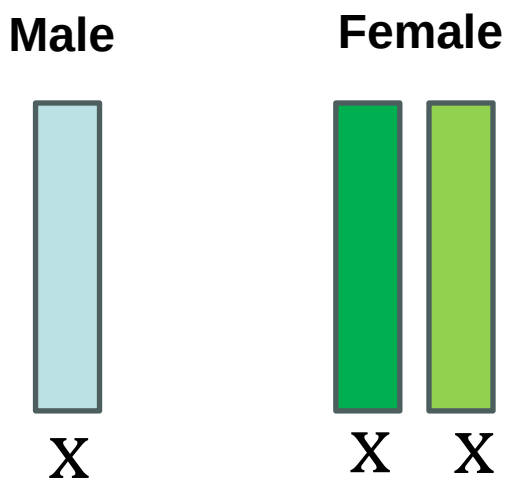


Question: Are child 1 and child 2 paternal half-sibs. or unrelated

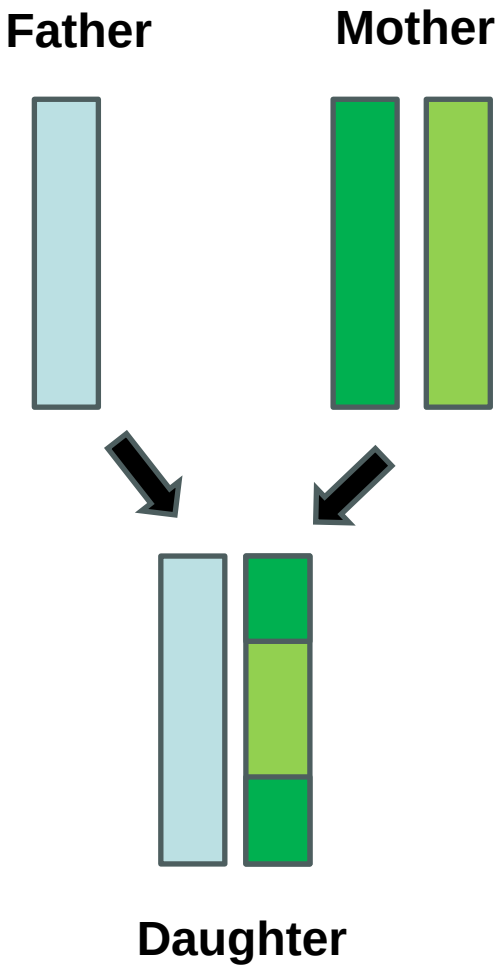
Genetic data: 4-12 X-STRs

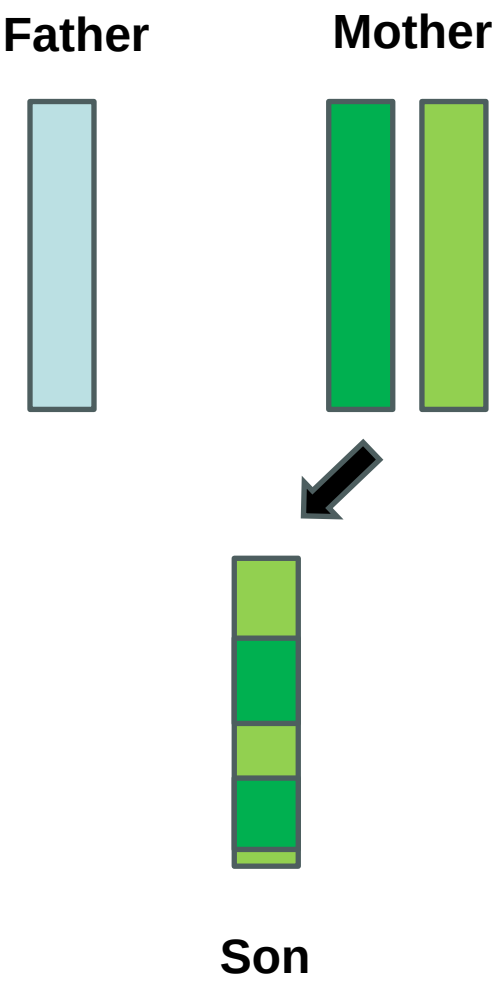
X chromosome in humans

- 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

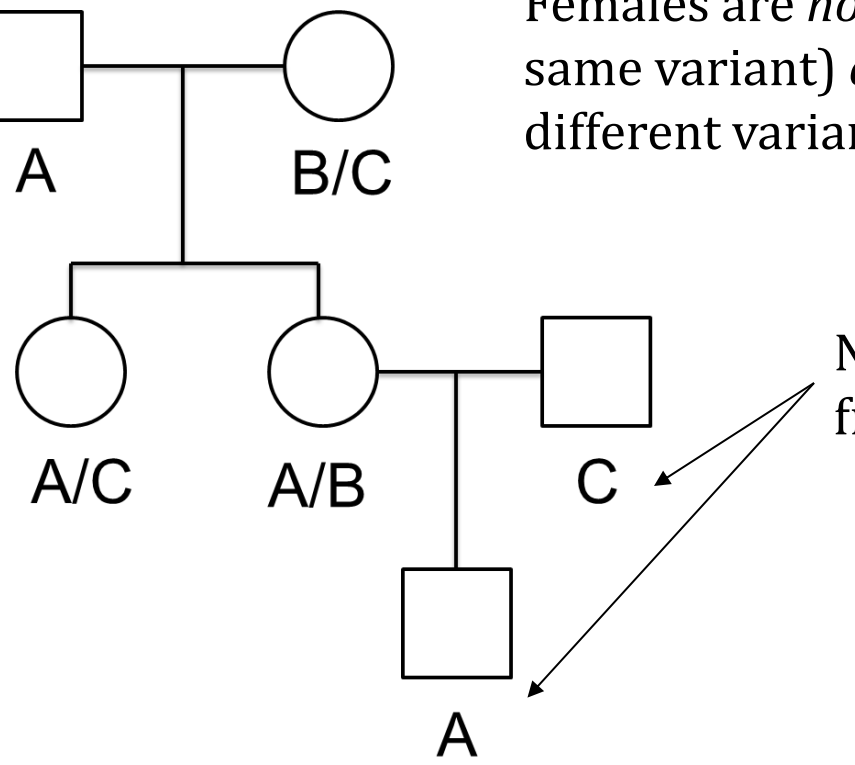




Inheritance pattern (one X locus)

Males are *hemizygous*

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



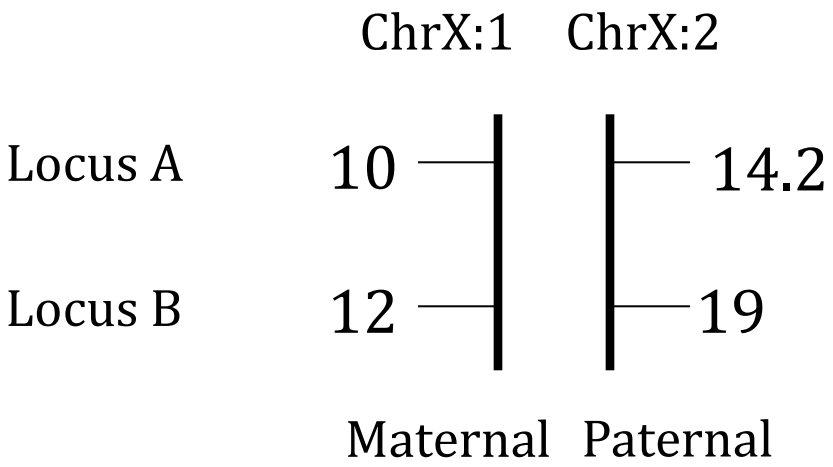
Forced transmission from father to daughter

No transmission from father to son

Basic notations:

Allele. haplotype. genotype. diplotype

Female

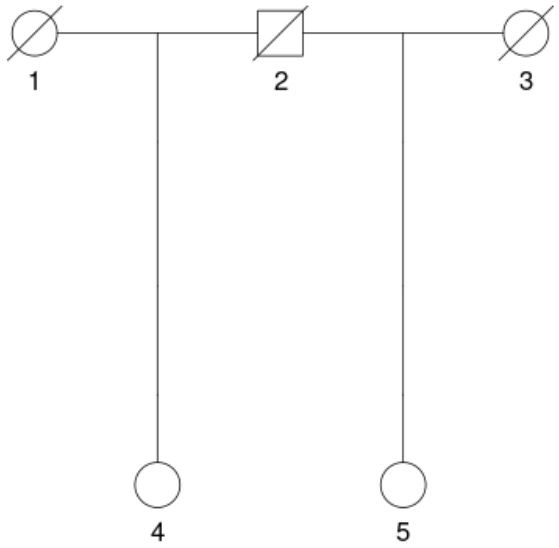


- 10 is an *allele*
- 10/14.2 (or 10.14.2) is a *genotype*
- 10_12 is a *haplotype*
- 10_12/14.2_19 is a *diplotype*
  (or 10_12|14.2_19)
  (or 10|14.2
  12|19)

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

Generally more informative

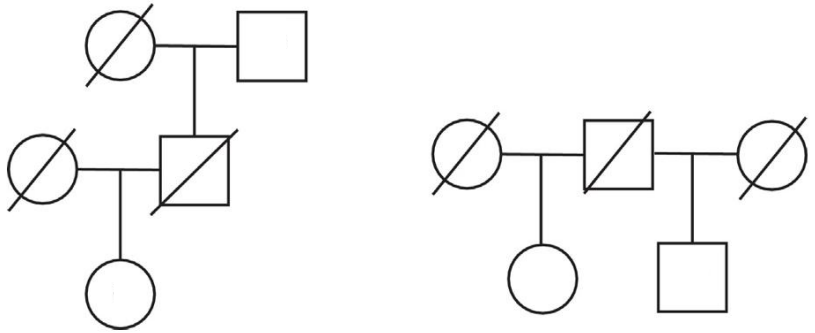
- Paternal half-sisters vs unrelated
- Paternal grandmother/granddaughter vs unrelated
- For many pedigress. the exclusion probability is not null



See Pinto et al.. 2011

Generally less informative

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



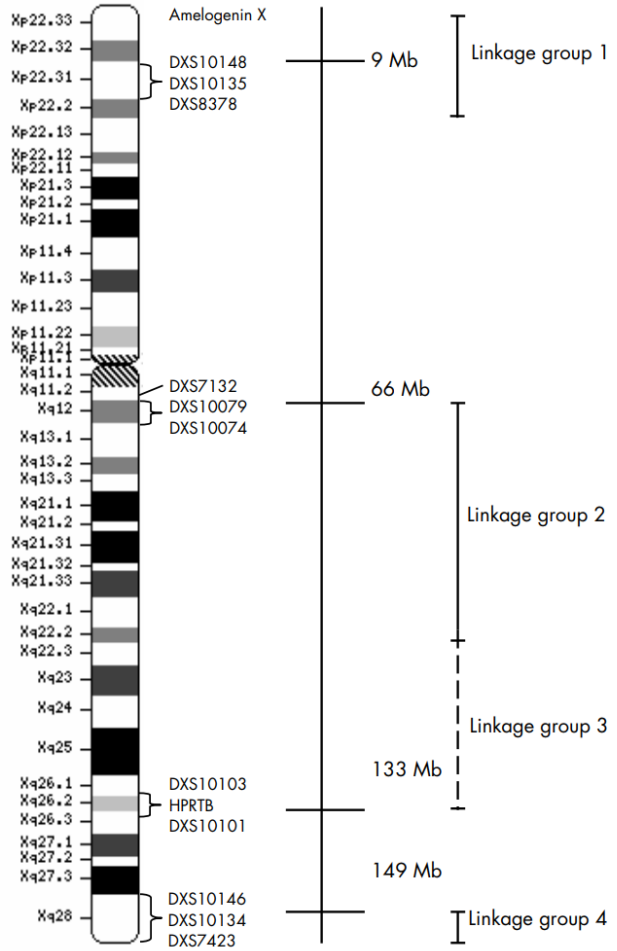
Tillmar et al.. 2017

Two common X-chromosomal marker panels

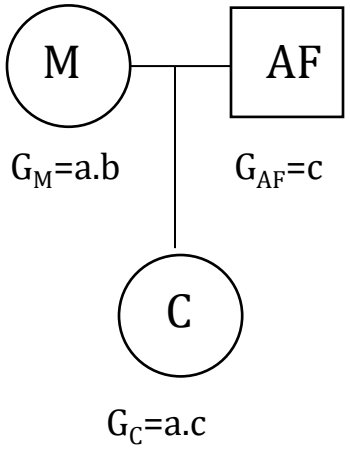
- STRs (short tandem repeats)
- **"X-Decaplex"**
  - 10 X STRs. in genetic linkage but mostly not in linkage disequilibrium (LD. allelic association).
  - Developed by GEP-ISFG (Gusmao et al.. 2009)

Focus in the first lecture

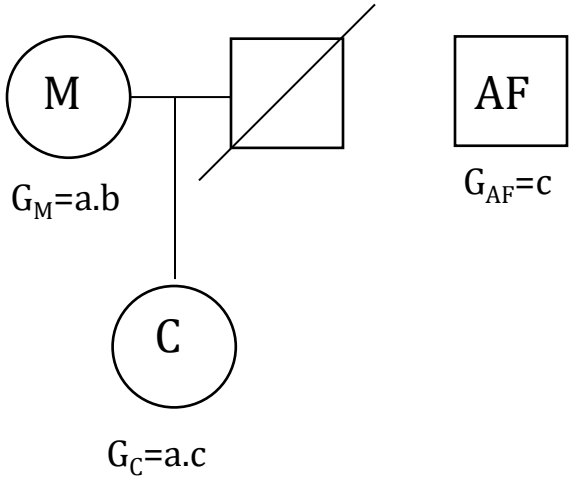
- **Argus X-12QS**
  - 12 X STRs. in four "linkage groups". in genetic linkage but mostly not in linkage disequilibrium (LD. allelic association).
  - Investigator Argus X-12 QS (Qiagen)



Which hypothesis is best supported by observed DNA profiles?



$\Pr(DNA \mid H_1)$

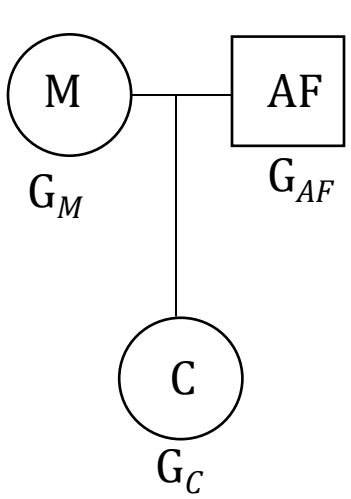


$\Pr(DNA \mid H_2)$

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

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

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



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*

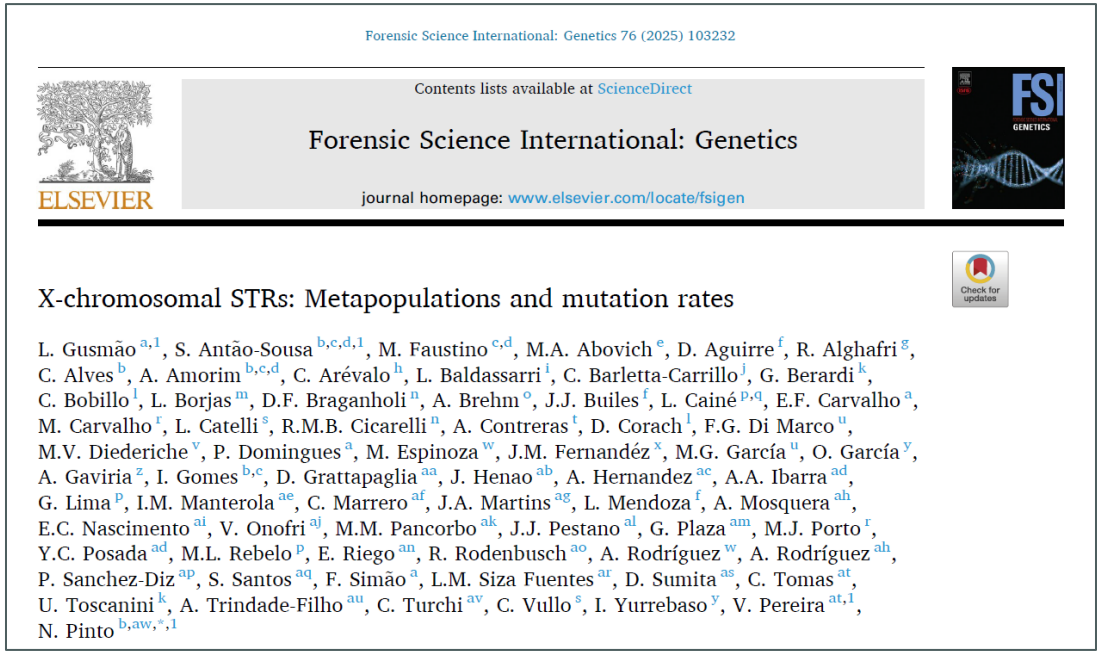
N Total number of observed haplotypes in the population database

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

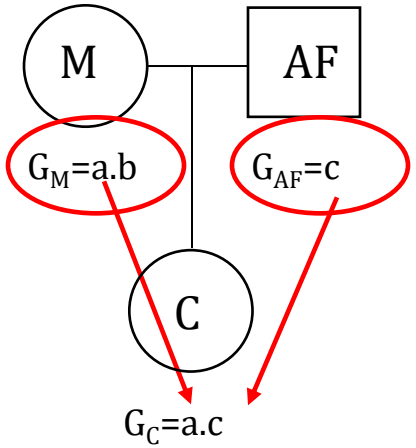
X-chromosomal STRs: Metapopulations and mutation rates

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

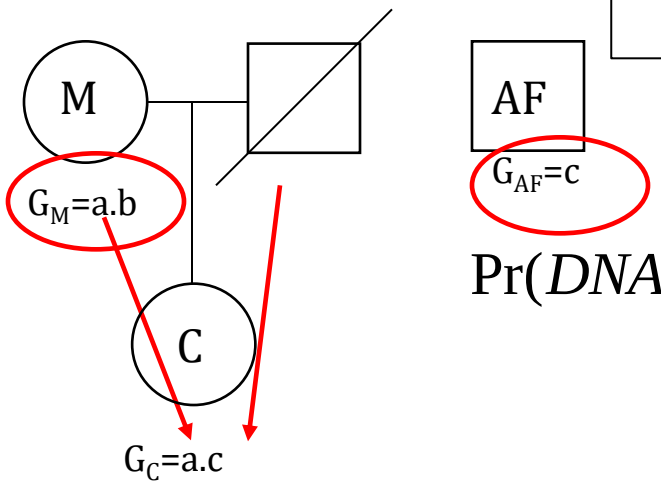
Paternity trio – A simple example with one X-chromosomal marker

$$\Pr(DNA \mid H_1) = \Pr(G_C, G_{AF}, G_M \mid H_1) = \Pr(G_{AF}, G_M \mid H_1) \cdot \Pr(G_C \mid G_{AF}, G_M, H_1)$$



$$\Pr(DNA \mid H_1) = 2 \cdot p_a \cdot p_b \cdot p_c \cdot 0.5 \cdot 1$$

$$\text{LR} = \frac{\Pr(DNA \mid H_1) = \quad 2 \cdot p_a \cdot p_b \cdot p_c \cdot 0.5 \cdot 1}{\Pr(DNA \mid H_2) = \quad 2 \cdot p_a \cdot p_b \cdot p_c \cdot 0.5 \cdot p_c} = \frac{1}{p_c}$$



$$\Pr(DNA \mid H_2) = 2 \cdot p_a \cdot p_b \cdot p_c \cdot 0.5 \cdot p_c$$

Two common X-chromosomal marker panels

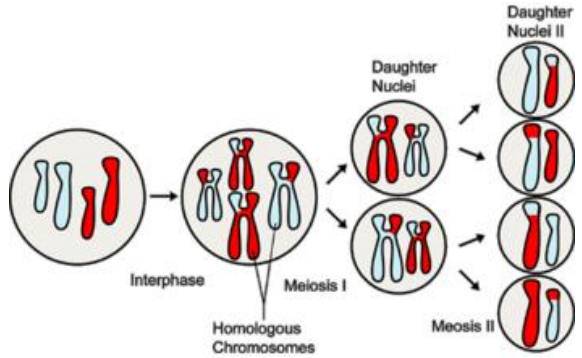
- STRs (short tandem repeats)
- **"X-Decaplex"** In focus first lecture
  - 10 X STRs. in genetic linkage but mostly not in linkage disequilibrium (LD. allelic association).
  - Developed by GEP-ISFG (Gusmao et al.. 2009)
- **Argus X-12QS**
  - 12 X STRs. in four "linkage groups". in genetic linkage but mostly not in linkage disequilibrium (LD. allelic association).
  - Investigator Argus X-12 QS (Qiagen)

DXS8378
DXS9898
DXS7133
GATA31E08
GATA172D05
DXS7423
DXS6809
DXS7132
DXS9902
DXS6789

- All markers located on the same chromosome.
- Markers are **genetically linked**

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!**

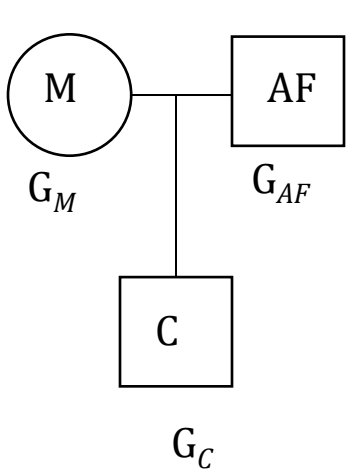


- Linkage disequilibrium (LD)
  - Allelic association.
  - Two alleles (at two different markers) which is observed more often/less often than can be expected.
  - **Effects the founder genotype probabilities. not the transmission probabilities!**
  - Haplotype frequencies rather than allele frequencies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

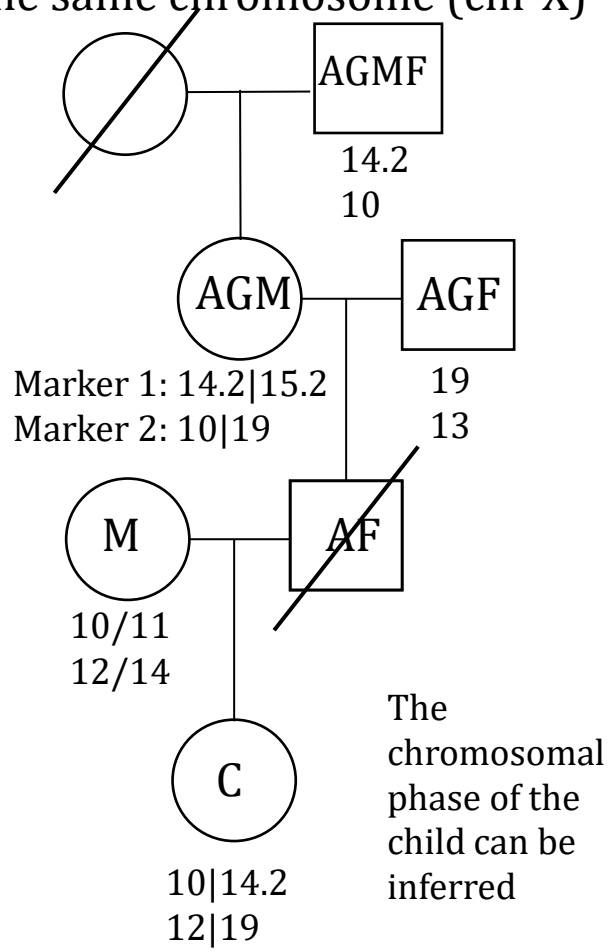
Mendelian factor

Recombination frequency/rate

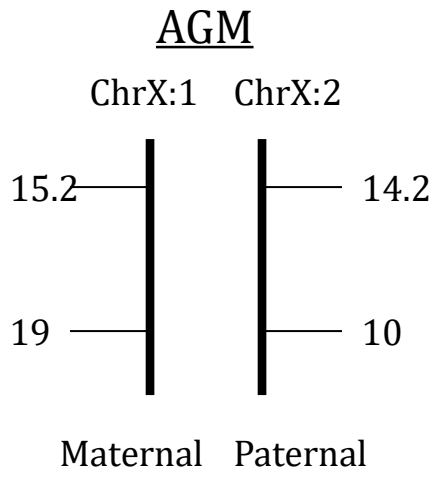
Linkage

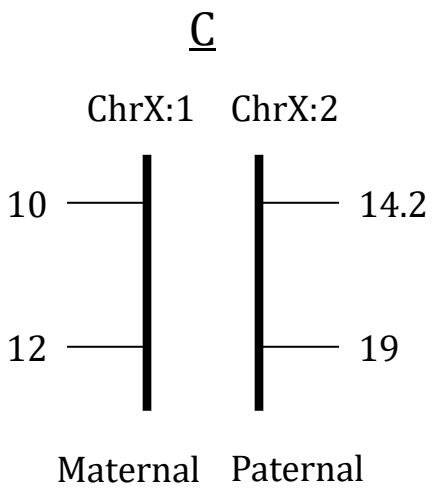
Linkage and how it impacts the LR

Marker 1 and 2 are located on the same chromosome (chr X)



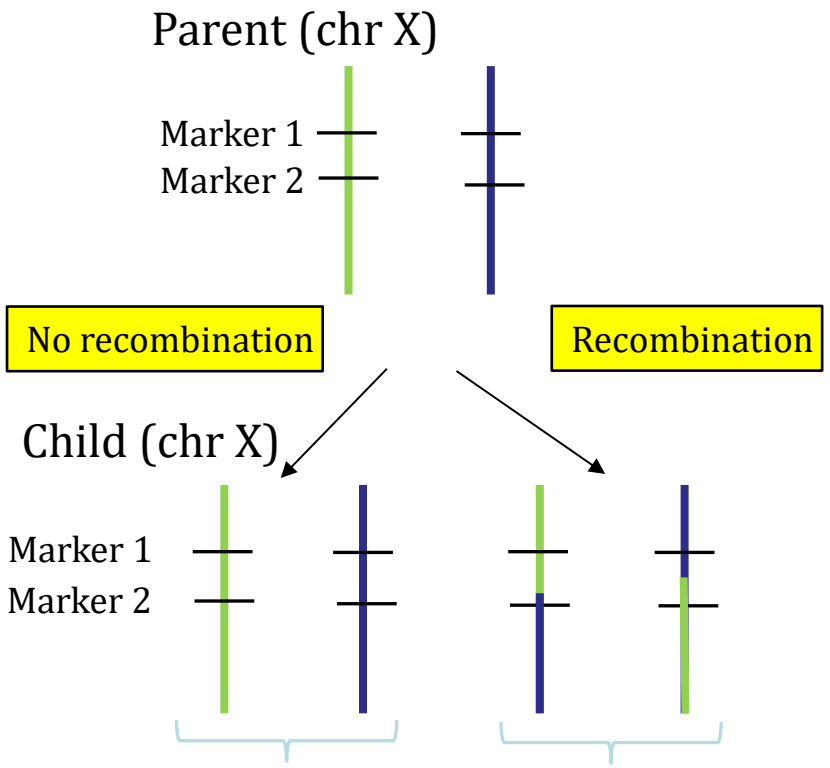
The chromosomal phase of the child can be inferred





1. A recombination must have occurred at AGM to explain the data (given this pedigree).
2. The probability of the observed DNA data (given this pedigree) depends on the recombination rate between marker 1 and marker 2!
3. E.g. if this recombination rate is very low. the probability is very low. Also. if the recombination rate is 0 (very very close markers). the observed data is not possible (given this pedigree)
4. **Ignoring genetic linkage may result in false LRs**

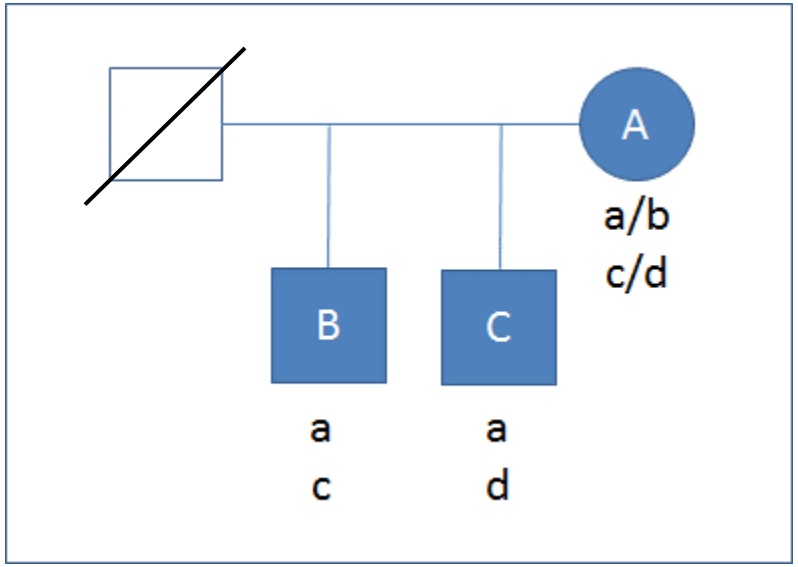
Genetic linkage

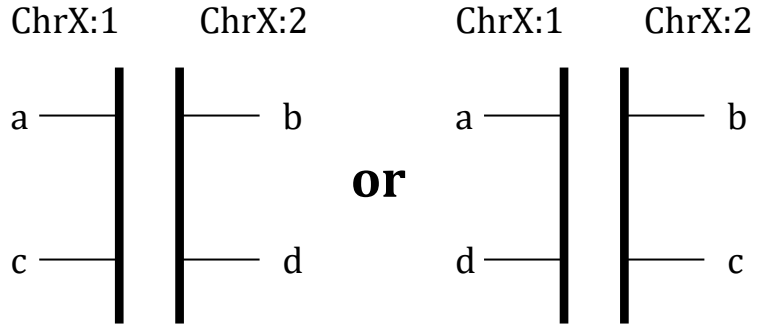


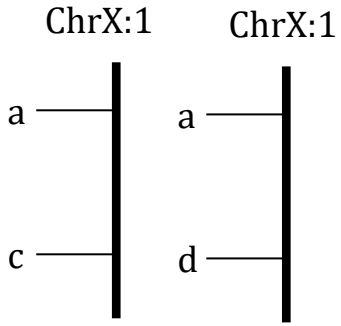
No Linkage	50%	50%
Linkage	>50%	<50%

Example of the effect of LR

2 STRs



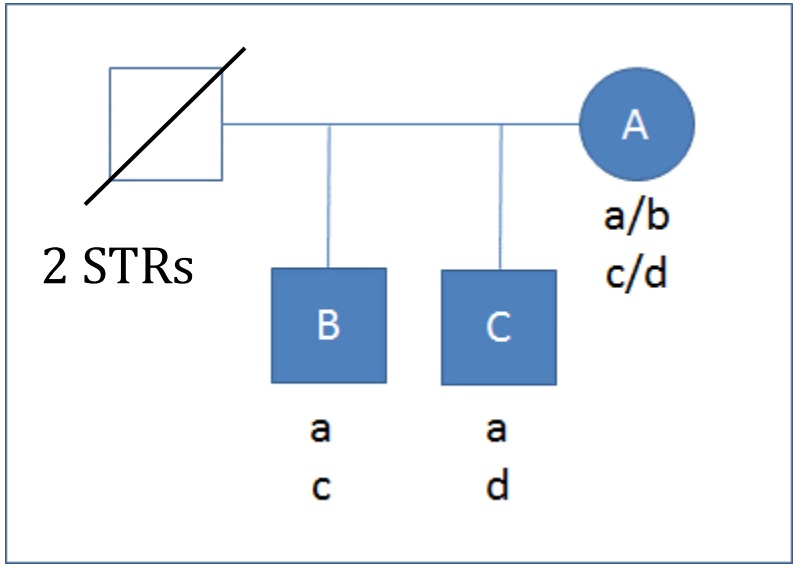


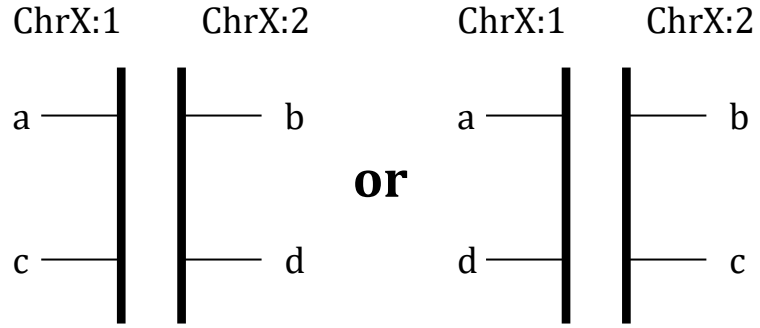


A recombination must have occurred at A to explain the data (given this pedigree).

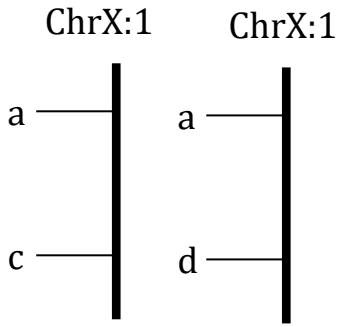
Example of the effect of LR

$$P(data|H1) = 2 \cdot p_{a-c} \cdot p_{b-d} \cdot 0.5 \cdot (1-r) \cdot 0.5 \cdot (r) + 2 \cdot p_{b-c} \cdot p_{a-d} \cdot 0.5 \cdot (r) \cdot 0.5 \cdot (1-r) = 0.5 \cdot (r - r^2) \cdot (p_{a-c} \cdot p_{b-d} + p_{b-c} \cdot p_{a-d})$$



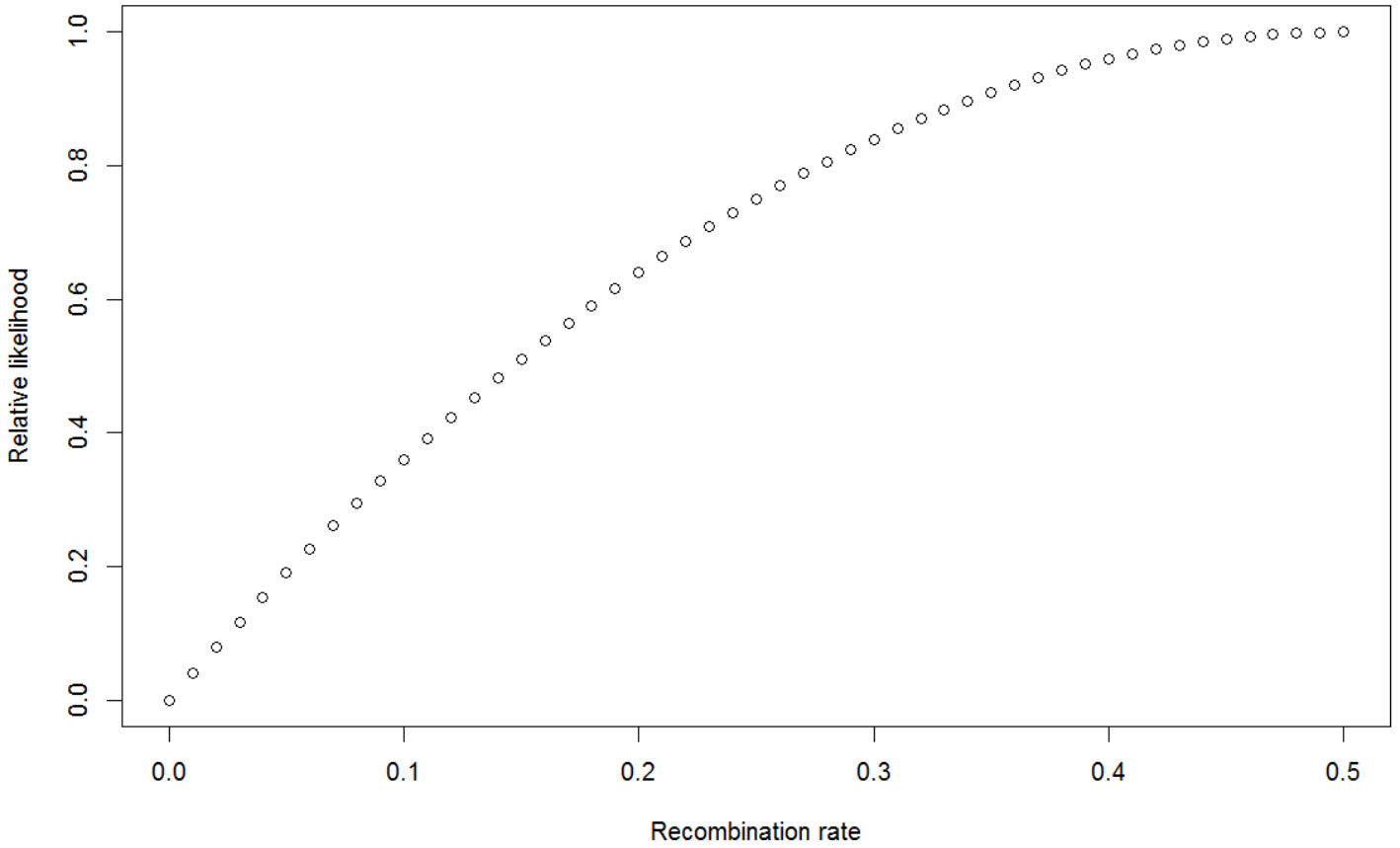


$$\frac{Pr(data|H1, r = 0 \ldots 0.5)}{Pr(data|H1, r = 0.5)} = 4r(1-r)$$



- A recombination must have occurred at A to explain the data (given this pedigree).
- The probability is correlated to the recombination frequency

Example of the effect of LR



Recombination rate/frequency

- When gametes are formed in meiosis. the two copies of each chromosome may be mixed together via **crossovers**.
- Closer chromosomal segments have a higher probability of staying together.
- If one. or an odd number. of crossovers occurs. a recombination has occurred.
- The probability of a recombination event to occur is the recombination rate/frequency.
- The recombination rate/frequency is used during likelihood calculation as the transmission probability.

- The recombination rate is normally correlated to the physical positions. but recombination hot spots exist!

We need information about the **genetic distance** between loci

- Centimorgan (cM) is a unit of genetic distance
- 1 cM ≈ 1% recombination frequency
- More precise estimators are Haldane's and Kosmabi's mapping functions

Haldane's Mapping Function

- Assumes no crossover interference
- Uses Poisson distribution
- The relationship between recombination rate ("r") and genetic distance ("d") can be estimate via Haldane's mapping function as:

Formula [edit]

$$r = \frac{1}{2}(1 - e^{-2d})$$

Inverse [edit]

$$d = -\frac{1}{2}\ln(1 - 2r)$$

Kosambi's Mapping Function

- Accounts for crossover interference
- More accurate for larger distances
- The relationship between recombination rate ("r") and genetic distance ("d") can be estimate via Kosambi's mapping function as:

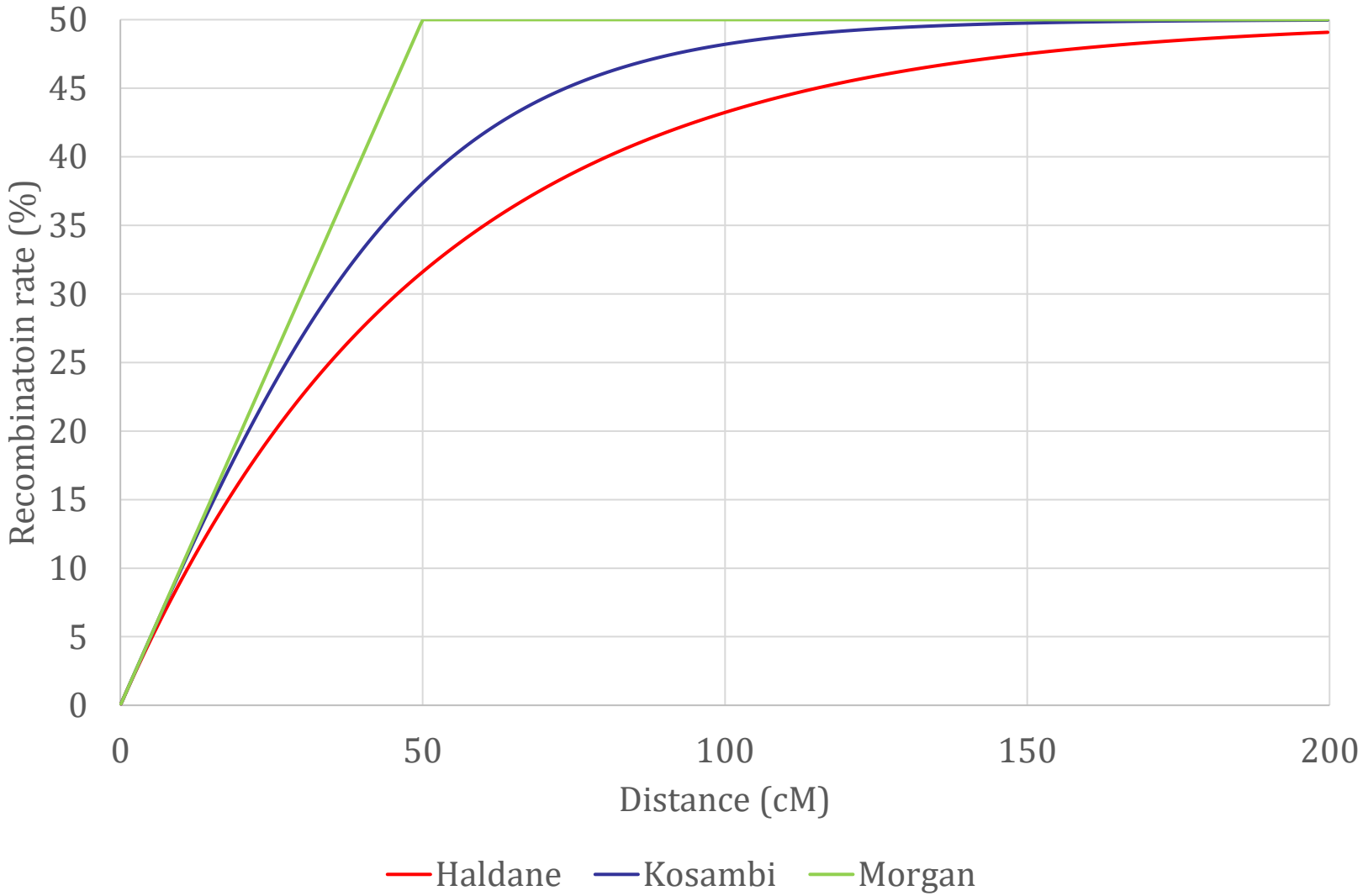
Formula [edit]

$$r = \frac{1}{2}\tanh(2d) = \frac{1}{2}\frac{e^{4d} - 1}{e^{4d} + 1}$$

Inverse [edit]

$$d = \frac{1}{2}\tanh^{-1}(2r) = \frac{1}{4}\ln\left(\frac{1 + 2r}{1 - 2r}\right)$$

Correlation between genetic distance (cM) and recombination rate



Genetic Maps: Rutgers. deCODE. and HapMap

Rutgers Genetic Map

•The Rutgers map integrates data from various genetic studies. including linkage disequilibrium and pedigree-based analyses. The primary method for estimating genetic distances involves interpolation of recombination rates derived from multi-generational family data. which is further refined using computational models to improve accuracy.
•http://compgen.rutgers.edu/rutgers_maps.shtml

deCODE Genetic Map

• The deCODE map estimates genetic distances by analyzing recombination events in a large Icelandic pedigree database. Recombination fractions between markers are directly observed from meiotic events within families. allowing for precise distance calculations.
•https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg38&g=recombRate2
•https://www.science.org/doi/10.1126/science.aau1043

HapMap Genetic Map

•Unlike traditional family-based genetic maps. the HapMap project estimates recombination rates using population-based LD data. Genetic distances are inferred by analyzing correlations between genetic variants and identifying historical recombination events within populations.
•https://www.genome.gov/10001688/international-hapmap-project

X-Decaplex STR genetic positions

Marker	Physical (cM)	Rutgers v2 (cM)
DXS8378	9.33	20.21
DXS9902	15.23	32.32
DXS7132	64.57	90.75
DXS9898	87.68	101.29
DXS6809	94.83	108.12
DXS6789	95.34	108.47
DXS7133	108.93	118.18
GATA172D05	113.06	124.36
GATA31E08	140.06	160.54
DXS7423	149.46	184.19

https://chrx-str.org/xdb/marker.jsf?marker=DXS7133

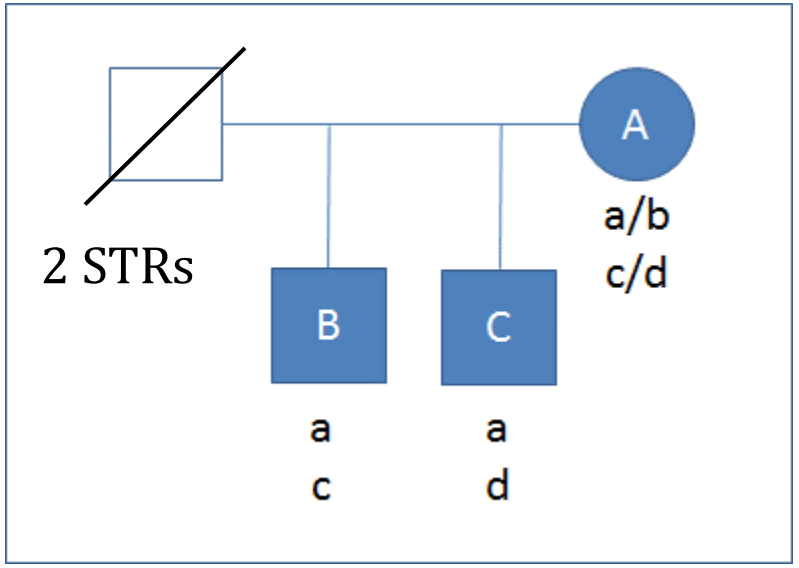
X-Decaplex STR genetic positions

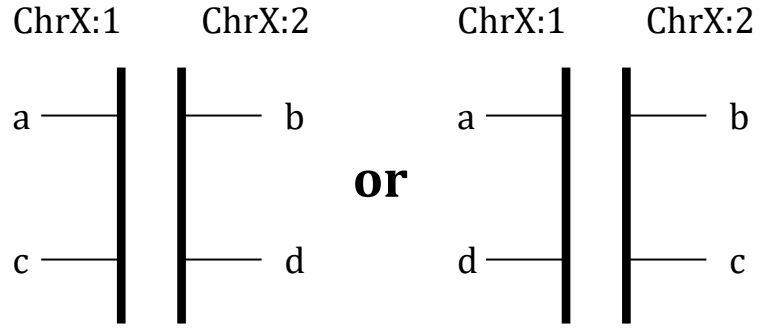
(distance between neighboring STRs)

Marker	Physical	Rutgers v2		Physical recomb (%)	Rutgers recomb (%)
DXS8378	9.33	20.21		0	0
DXS9902	15.23	32.32		5.57	10.76
DXS7132	64.57	90.75		31.36	34.46
DXS9898	87.68	101.29		18.51	9.50
DXS6809	94.83	108.12		6.66	6.38
DXS6789	95.34	108.47		0.51	0.35
DXS7133	108.93	118.18		11.90	8.83
GATA172D05	113.06	124.36		3.97	5.81
GATA31E08	140.06	160.54		20.86	25.75
DXS7423	149.46	184.19		8.57	18.84

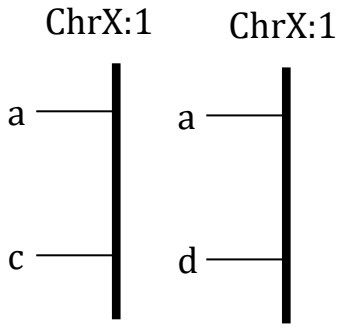
Example of the effect of LR

$$P(data|H1) = 2 \cdot p_{a-c} \cdot p_{b-d} \cdot 0.5 \cdot (1-r) \cdot 0.5 \cdot (r) + 2 \cdot p_{b-c} \cdot p_{a-d} \cdot 0.5 \cdot (r) \cdot 0.5 \cdot (1-r) = 0.5 \cdot (r - r^2) \cdot (p_{a-c} \cdot p_{b-d} + p_{b-c} \cdot p_{a-d})$$

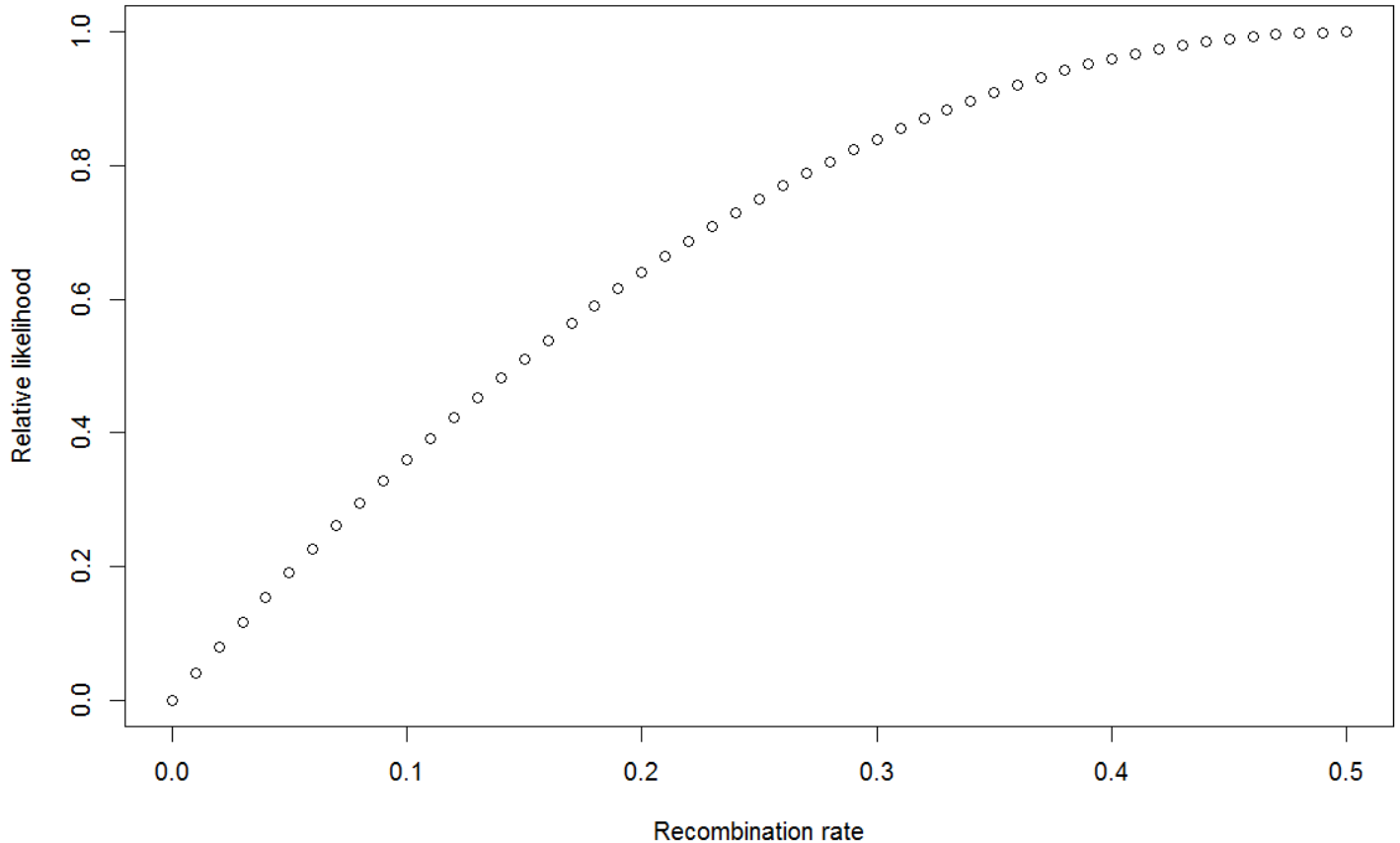




$$\frac{Pr(data|H1, r = 0 \ldots 0.5)}{Pr(data|H1, r = 0.5)} = 4r(1-r)$$



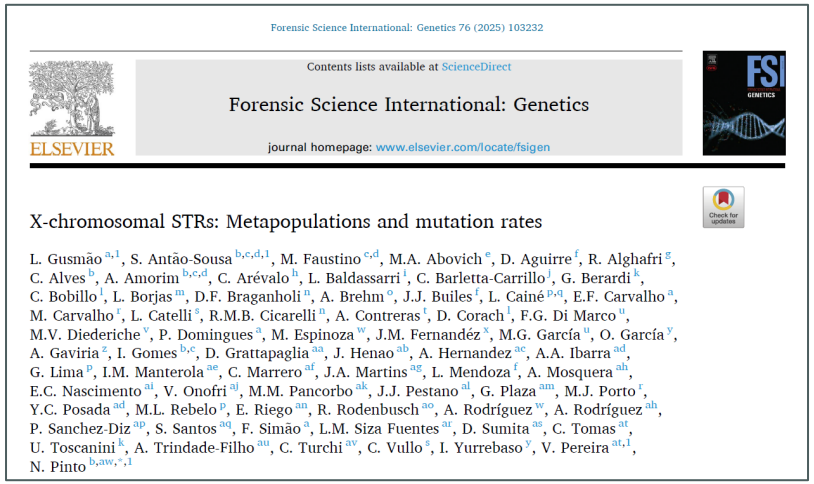
- A recombination must have occurred at A to explain the data (given this pedigree).
- The probability is correlated to the recombination frequency


Relative likelihood
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Recombination rate

STR mutations

"The possibility of mutation shall be taken into account whenever a genetic inconsistency is observed" (Gjertson et al.. 2007)

- Brinkmann et al (1998) found 23 mutations in 10.844 parent/child offsprings. Out of these 22 were single step and 1 were two-step mutations.
- Gusmao et al (2025)



Forensic Science International: Genetics 76 (2025) 103232

Contents lists available at ScienceDirect

Forensic Science International: Genetics

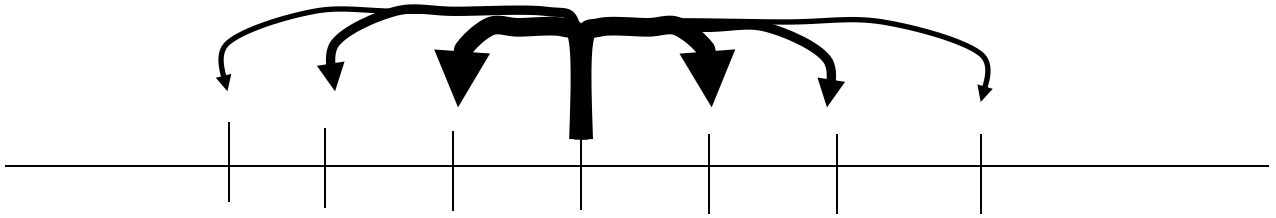
journal homepage: www.elsevier.com/locate/fsigen

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 [i], 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]

- Mutation rate may depend on marker. sex (female/male). age of individual. allele size

Models for STR mutations



Allele:	9	10	11	12	13	14	15
	-3	-2	-1	0	+1	+2	+3
				$1-\mu_{tot}$			
	μ_{-3}	μ_{-2}	μ_{-1}		μ_{+1}	μ_{+2}	μ_{+3}

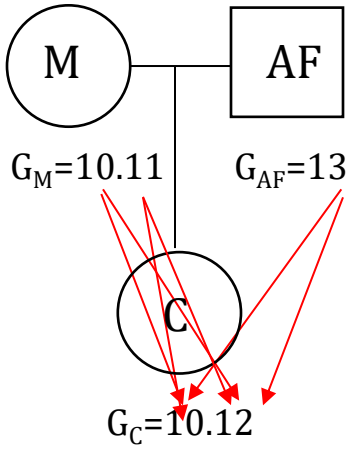
$$\mu_{tot}=\mu_{+1}+\mu_{-1}+\mu_{+2}+\mu_{-2}+\ldots$$

Different approaches to calculate LR accounting for mutations exist.

The most used one *"Stepwise mutation model"*

LR ~ (μ_{tot}*adj_steps)/p(paternal allele)

Paternity trio - Mutation



$$\Pr(mut_{13\rightarrow 12})$$

Mutation model decreasing with range

A software like FamLinkX will consider all mutation possibilities (if themutation rate is set to >0)

Total mutation rate for the locus

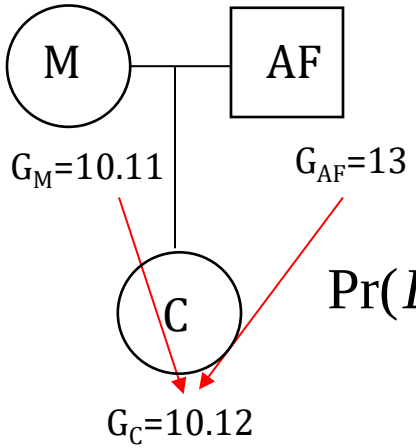
50% are loss of fragment size

$$\Pr(mut_{13->12}) = \{ \text{“Stepwise decreasing with range”} \} = 1 \cdot \mu_{Tot} \cdot 0.9 \cdot 0.5$$

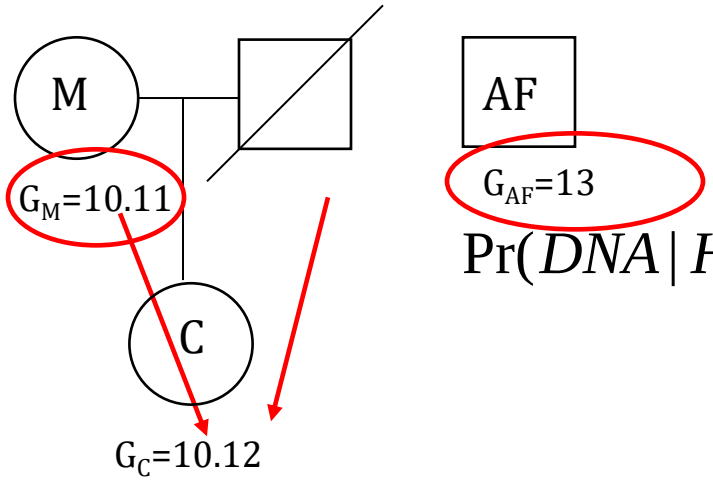
100% for 13 to be the parental original allele

90% of mutations are 1-step

Paternity trio - Mutation



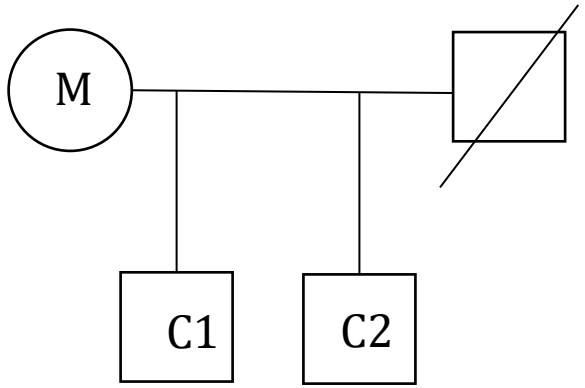
$$\Pr(DNA \mid H_1) = 2 \cdot p_{10} \cdot p_{11} \cdot p_{13} \cdot 0.5 \cdot \underbrace{(mut_{13->12})}_{1 \cdot \mu_{Tot} \cdot 0.9 \cdot 0.5}$$



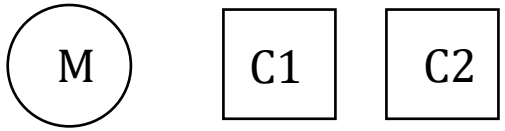
$$\Pr(DNA \mid H_2) = 2 \cdot p_{10} \cdot p_{11} \cdot p_{13} \cdot 0.5 \cdot p_{12}$$

$$LR = \frac{1 \cdot \mu_{Tot} \cdot 0.9 \cdot 0.5}{p_{12}}$$

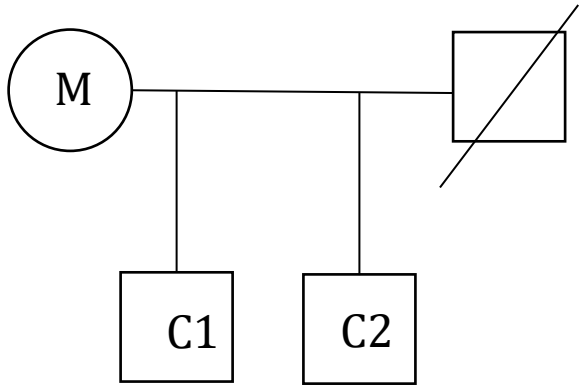
FamLinkX



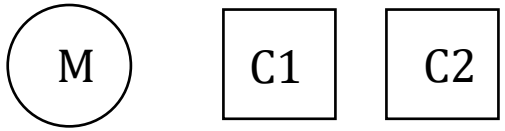
H1: C1 and C2 are full siblings. M being the mother of both C1 and C2



H2: M. C1 and C2 are all unrelated



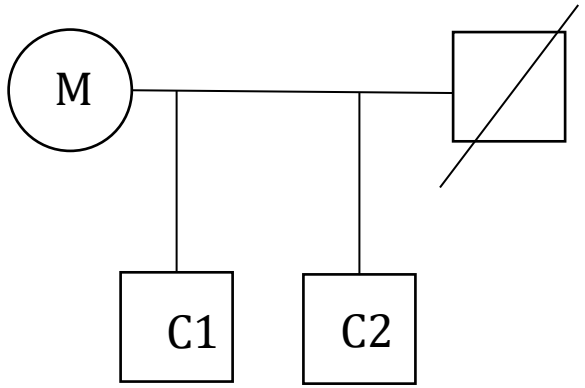
H1: C1 and C2 are full siblings. M being the mother of both C1 and C2



H2: M. C1 and C2 are all unrelated

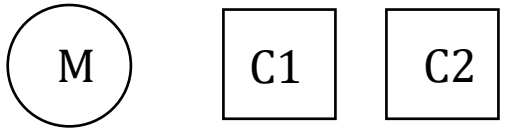
Name	Genetic position (Rutgers)	M	C1	C2
Amel	-	X/X	X/Y	X/Y
DXS8378	20.21	10/11	10	10
DXS9902	32.32	11/12	11	11
DXS7132	90.75	13/14	13	13
DXS9898	101.29	11/12	11	11
DXS6809	108.12	31/32	31	31
DXS6789	108.47	20/21	20	21
DXS7133	118.18	9/10	9	9
GATA172D05	124.36	10/11	10	10
GATA31E08	160.54	11/12	11	11
DXS7423	184.19	14/15	14	14

No genetic inconsistencies between M and C1. and M and C2



H1: C1 and C2 are full siblings. M being the mother of both C1 and C2

Name	Genetic position (Rutgers)	M	C1	C2	Marginal LR
Amel	-	X/X	X/Y	X/Y	
DXS8378	20.21	10/11	10	10	1.8
DXS9902	32.32	11/12	11	11	3.6
DXS7132	90.75	13/14	13	13	4.6
DXS9898	101.29	11/12	11	11	26.7
DXS6809	108.12	31/32	31	31	29.1
DXS6789	108.47	20/21	20	21	0.03
DXS7133	118.18	9/10	9	9	0.3
GATA172D05	124.36	10/11	10	10	4.4
GATA31E08	160.54	11/12	11	11	7.0
DXS7423	184.19	14/15	14	14	4.2

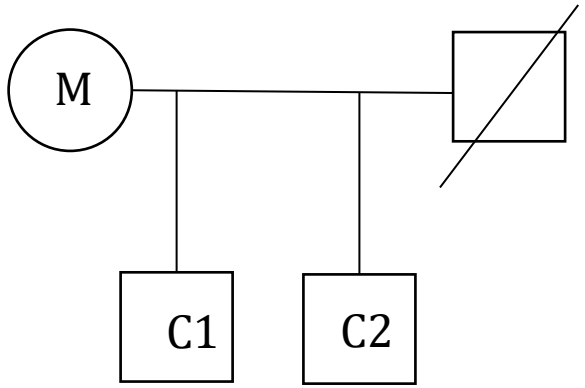


H2: M. C1 and C2 are all unrelated

Marginal LR for DXS6789 is very low. even tough no apparent inconsistency between M and C1. C2!

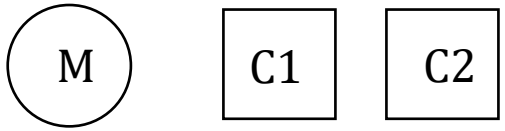
⇒ A recombination event must have occurred to explain the observed data.

⇒ Only 0.5 cM between DXS6809 and DXS6789



H1: C1 and C2 are full siblings. M being the mother of both C1 and C2

Name	Genetic position (Rutgers)	M	C1	C2	Marginal LR
Amel	-	X/X	X/Y	X/Y	
DXS8378	20.21	10/11	10	10	1.8
DXS9902	32.32	11/12	11	11	3.6
DXS7132	90.75	13/14	13	13	4.6
DXS9898	101.29	11/12	11	11	26.7
DXS6809	108.12	31/32	31	31	29.1
DXS6789	108.47	20/21	20	21	0.03
DXS7133	118.18	9/10	9	9	0.3
GATA172D05	124.36	10/11	10	10	4.4
GATA31E08	160.54	11/12	11	11	7.0
DXS7423	184.19	14/15	14	14	4.2

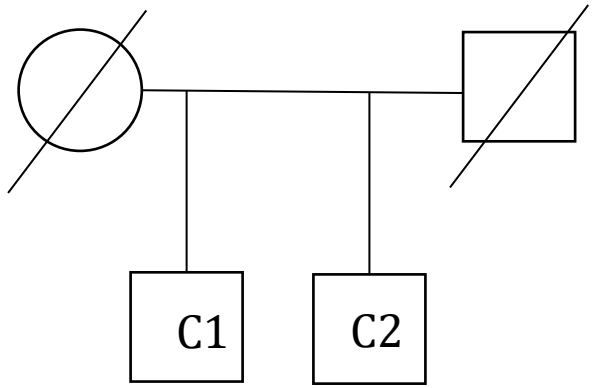


H2: M. C1 and C2 are all unrelated

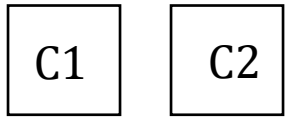
Marginal LR for DXS6789 is very low. even tough no apparent inconsistency between M and C1. C2!

⇒ A recombination event must have occurred to explain the observed data.

⇒ Only 0.5 cM between DXS6809 and DXS6789



H1: C1 and C2 are full siblings



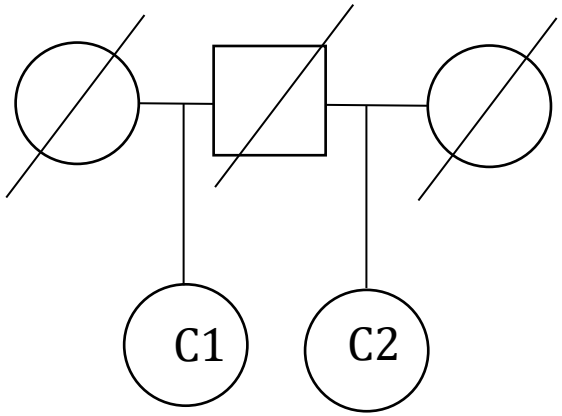
H2: C1 and C2 are all unrelated

Name	Genetic position (Rutgers)	C1	C2	Marginal LR
Amel	-	X/Y	X/Y	
DXS8378	20.21	10	10	1.8
DXS9902	32.32	11	11	2.3
DXS7132	90.75	13	13	2.7
DXS9898	101.29	11	11	6.0
DXS6809	108.12	31	31	7.0
DXS6789	108.47	20	21	0.03
DXS7133	118.18	9	9	1.1
GATA172D05	124.36	10	10	1.7
GATA31E08	160.54	11	11	3.0
DXS7423	184.19	14	14	2.6

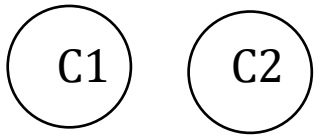
Marginal LR for DXS6789 is very low. even tough no data from. M!

⇒ A recombination event is probably to have occurred to explain the observed data (given population frequencies).

⇒ Only 0.5 cM between DXS6809 and DXS6789



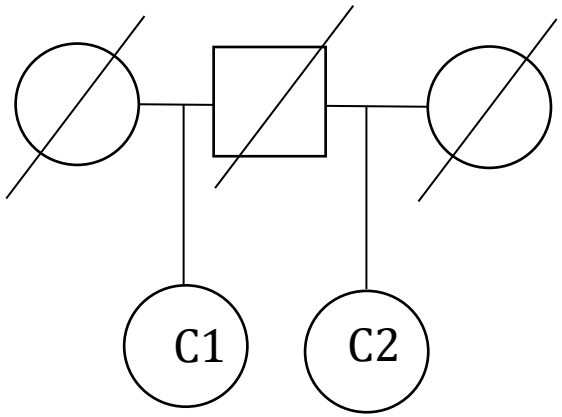
H1: C1 and C2 are paternal half siblings



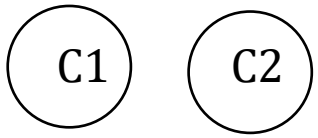
H2: C1 and C2 are unrelated

Name	Genetic position (Rutgers)	C1	C2	
Amel	-	X/X	X/X	
DXS8378	20.21	10/12	11/12	
DXS9902	32.32	11/13	11/13	
DXS7132	90.75	13/15	13/15	
DXS9898	101.29	11/13	12/13	
DXS6809	108.12	31/33	31/33	
DXS6789	108.47	20/22	21/22	
DXS7133	118.18	9/11	9/11	
GATA172D05	124.36	10/12	11/12	
GATA31E08	160.54	11/13	11/13	
DXS7423	184.19	14/16	15/16	

No genetic inconsistencies between C1 and C2
$\Rightarrow$ LR=322. **Is this expected? Let's simulate**



H1: C1 and C2 are paternal half siblings



H2: C1 and C2 are unrelated

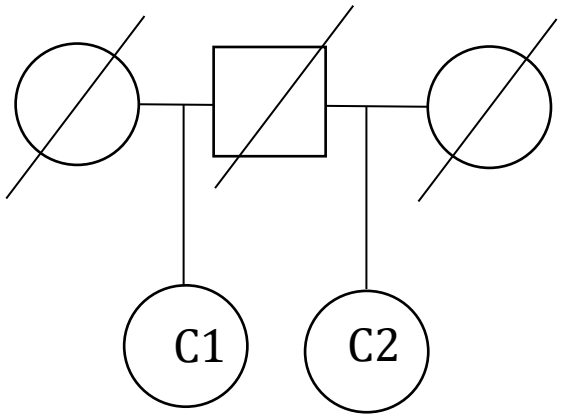
Name	Genetic position (Rutgers)	C1	C2	
Amel	-	X/X	X/X	
DXS8378	20.21	10/12	11/12	
DXS9902	32.32	11/13	11/13	
DXS7132	90.75	13/15	13/15	
DXS9898	101.29	11/13	12/13	
DXS6809	108.12	31/33	31/33	
DXS6789	108.47	20/22	21/22	
DXS7133	118.18	9/11	9/11	
GATA172D05	124.36	10/12	11/12	
GATA31E08	160.54	11/13	11/13	
DXS7423	184.19	14/16	15/16	

No genetic inconsistencies between C1 and C2

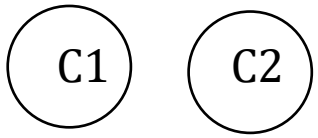
$\Rightarrow$ LR=322. **Is this expected? Let's simulate:**

$\Rightarrow$ **Median around LR=300**

What about mutations?



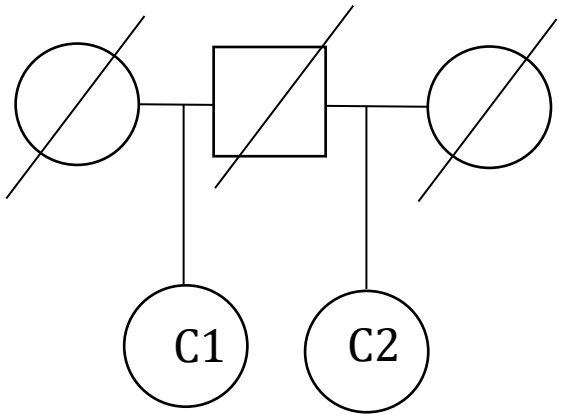
H1: C1 and C2 are paternal half siblings



H2: C1 and C2 are unrelated

Name	Genetic position (Rutgers)	C1	C2	
Amel	-	X/X	X/X	
DXS8378	20.21	10/12	11/11	#
DXS9902	32.32	11/13	11/13	
DXS7132	90.75	13/15	13/15	
DXS9898	101.29	11/13	12/13	
DXS6809	108.12	31/33	31/33	
DXS6789	108.47	20/22	21/22	
DXS7133	118.18	9/11	9/11	
GATA172D05	124.36	10/12	11/12	
GATA31E08	160.54	11/13	11/13	
DXS7423	184.19	14/16	15/16	

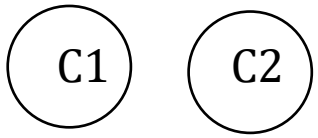
LR=1.2 (LR decreases with a factor 300. Is this to expect?)



H1: C1 and C2 are paternal half siblings

Name	Genetic position (Rutgers)	C1	C2	
Amel	-	X/X	X/X	
DXS8378	20.21	10/12	11/11	#
DXS9902	32.32	11/13	11/13	
DXS7132	90.75	13/15	13/15	
DXS9898	101.29	11/13	12/13	
DXS6809	108.12	31/33	31/33	
DXS6789	108.47	20/22	21/22	
DXS7133	118.18	9/11	9/11	
GATA172D05	124.36	10/12	11/12	
GATA31E08	160.54	11/13	11/13	
DXS7423	184.19	14/16	15/16	

LR=1.2 (LR decreases with a factor 300. Is this to expect?)



H2: C1 and C2 are unrelated

$$\Pr(mut_{10->11} or\ mut_{12->11}\ or\ mut_{11->10}\ or\ mut_{11->12}) \approx 4 \cdot (1 \cdot \mu_{Tot} \cdot 0.9 \cdot 0.5)$$

X-chromosomal markers in Forensic Genetics

GHEP 2025 Virtual workshop series.

March ~~10~~.17 and 24th

Daniel Kling and Andreas Tillmar

Linkage disequilibrium, LD

- Allelic association at a population level
  - ”12” at STR 1 is observed with ”16” at STR 2 much more often than expected
- Causes:
  - **Genetic linkage**: When loci are physically close on the same chromosome, recombination is less likely to separate them, leading to LD.
  - **Mutation**: A new mutation at one locus can create LD if it arises on a specific haplotype and recombination has not yet had time to break the association.
  - **Population genetic effects**: Drift, Founder effects, Bottlenecks; If a population undergoes a sharp reduction in size or is founded by a small number of individuals, certain allele combinations can become more common, creating LD.
- “Break down” of LD
  - **Recombinations** and **random mating.**