gembloux agro bio tech

fnrs LA LIBERTÉ DE CHERCHER

Université de Liège

Forensic science technique applied for calculation of kinship index

E. Bömcke [1,2], and N. Gengler [1,3]

[1] Animal Science Unit, Gembloux Agro-Bio Tech, University of Liege (GxABT, ULg) – Gembloux, Belgium
[2] FRIA - Brussels, Belgium
[3] National Fund for Scientific Research (FRS-FNRS) – Brussels, Belgium

EAAP 2010
August 23 – August 27, Heraklion, Greece

Context

- Relationship coefficients = useful tool to improve genetic management of small and/or endangered populations
- Relationship coefficients are traditionally calculated on pedigree data
- But in pedigrees
  - Often presence of errors
  - Some parents are missing

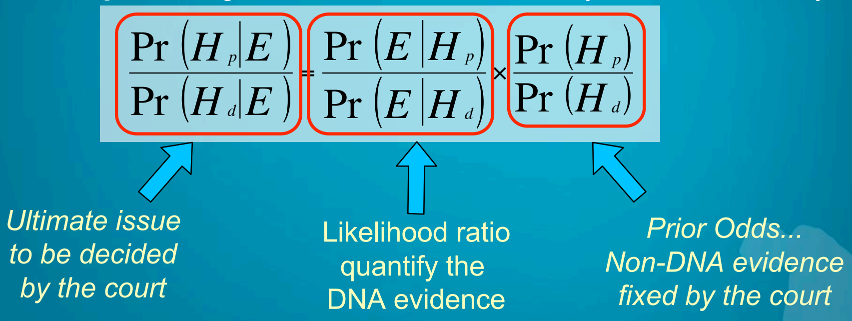
Objective

- Find tools to help the breeders for the management of endangered population or population with incomplete pedigrees
- Increase the knowledge of kinship trough detection in the pedigrees of
  - False parents
  - Non-recorded parents

➔ Calculation of kinship index

Methods

- Based on calculation of Likelihood ratio (LR)
- Principle - Bayes' Theorem in court (DNA evidence)



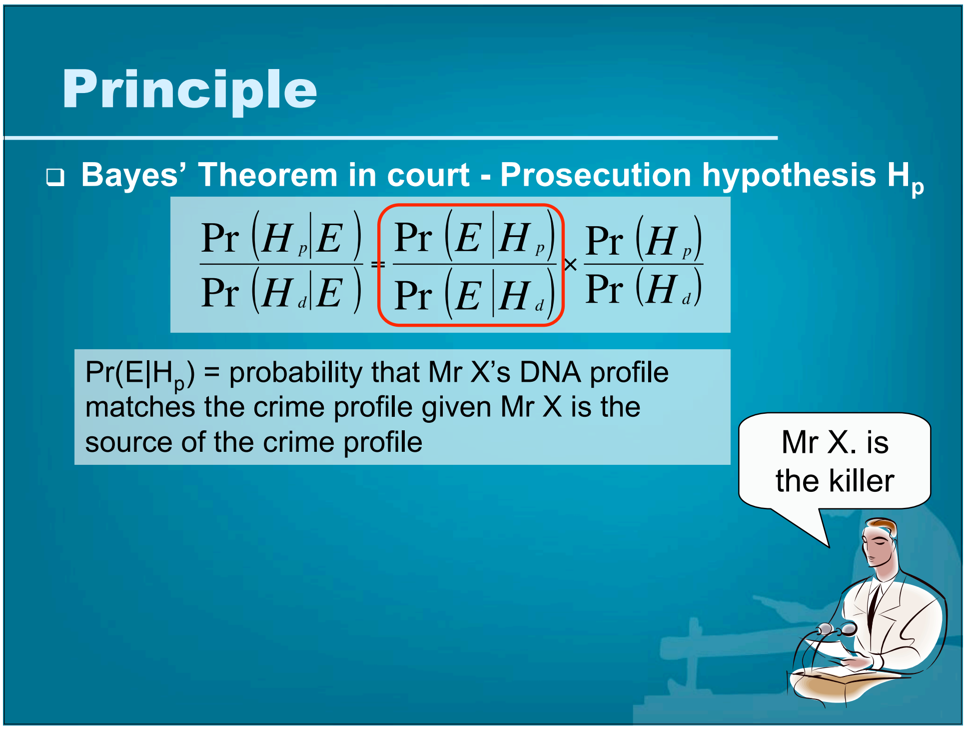
$$\boxed{\frac{\Pr(H_p|E)}{\Pr(H_d|E)}} = \boxed{\frac{\Pr(E|H_p)}{\Pr(E|H_d)}} \times \boxed{\frac{\Pr(H_p)}{\Pr(H_d)}}$$

Ultimate issue to be decided by the court

Likelihood ratio quantify the DNA evidence

Prior Odds... Non-DNA evidence fixed by the court

- H_p = prosecution hypothesis
- H_d = defense hypothesis
- Pr(.|.) = conditional probability



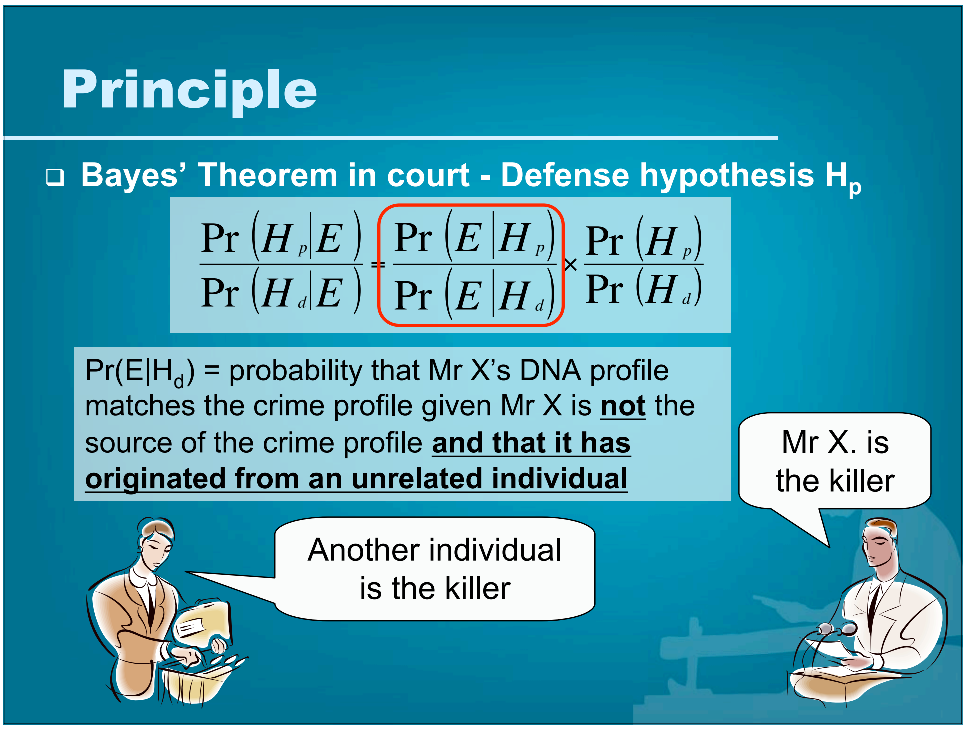
Principle

- Bayes' Theorem in court - Prosecution hypothesis H_p

$$\frac{\Pr\left(H_p|E\right)}{\Pr\left(H_d|E\right)}=\boxed{\frac{\Pr\left(E|H_p\right)}{\Pr\left(E|H_d\right)}}\times\frac{\Pr\left(H_p\right)}{\Pr\left(H_d\right)}$$

$\Pr(E|H_p)$ = probability that Mr X's DNA profile matches the crime profile given Mr X is the source of the crime profile

Mr X. is the killer



Principle

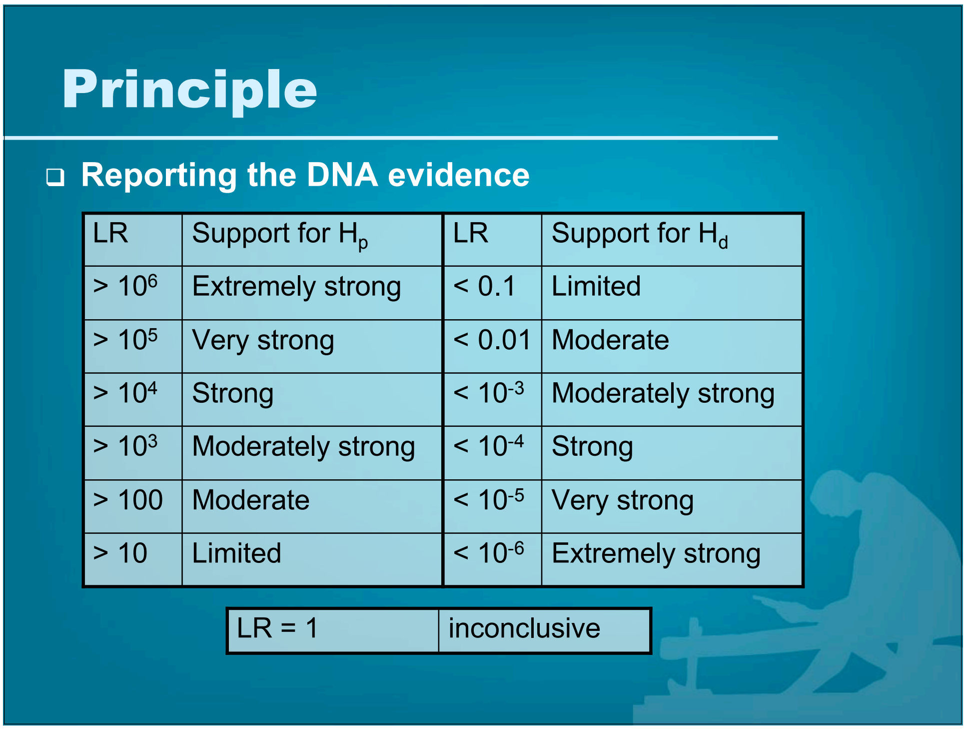
- Bayes' Theorem in court - Defense hypothesis H_p

$$\frac{\Pr\left(H_p|E\right)}{\Pr\left(H_d|E\right)}=\boxed{\frac{\Pr\left(E|H_p\right)}{\Pr\left(E|H_d\right)}}\times\frac{\Pr\left(H_p\right)}{\Pr\left(H_d\right)}$$

$\Pr(E|H_d)$ = probability that Mr X's DNA profile matches the crime profile given Mr X is **not** the source of the crime profile **and that it has originated from an unrelated individual**

Mr X. is the killer

Another individual is the killer



Principle

- Reporting the DNA evidence

LR	Support for H_p	LR	Support for H_d
> 10^6	Extremely strong	< 0.1	Limited
> 10^5	Very strong	< 0.01	Moderate
> 10^4	Strong	< 10^{-3}	Moderately strong
> 10^3	Moderately strong	< 10^{-4}	Strong
> 100	Moderate	< 10^{-5}	Very strong
> 10	Limited	< 10^{-6}	Extremely strong

LR = 1	inconclusive

Methods

- In court, DNA evidence match directly to criminal profiles
- But, LR can also be used to do Familial searching

Methods

- In court, DNA evidence match directly to criminal profiles
- But, LR can also be used to do Familial searching
- Familial searching
  - Method used in forensic science
  - Search for people in a database who are related to DNA evidence
  - New hypothesis for parentage calculations
    - H_p ➔ H1: the alleged father (or mother) is the true parent
    - H_d ➔ H2: the alleged father (or mother) is not the parent

---

Methods - parent/child LR

- 2 individuals (x and y), 1 locus, 4 allele positions

➔

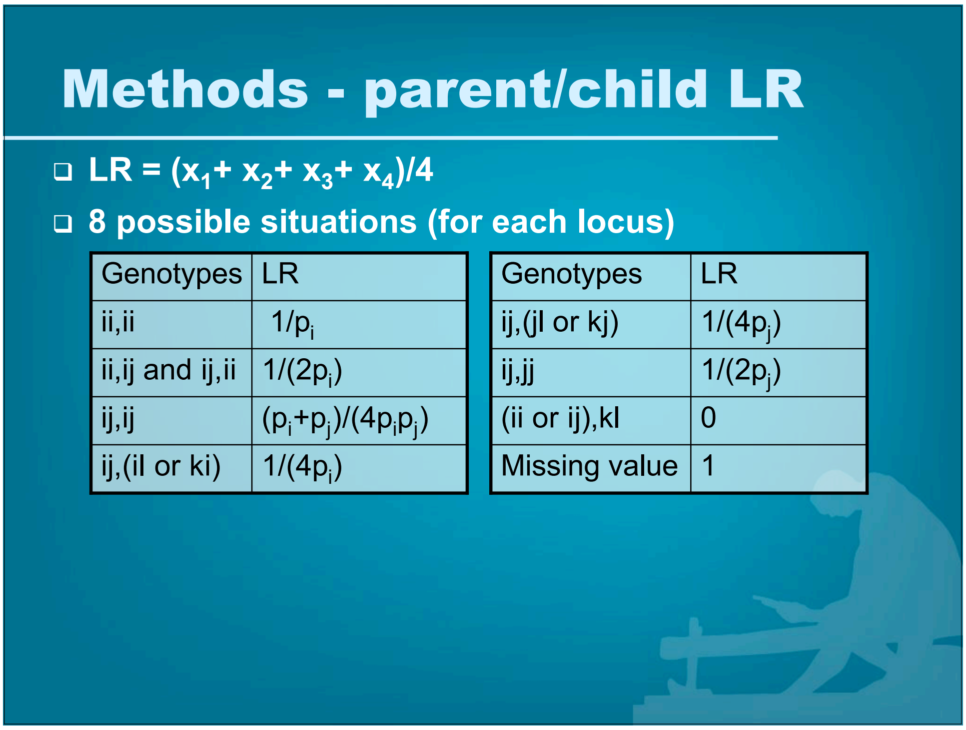
x	a	b
y	c	d

$$LR = \frac{\Pr(ab,cd \mid H_1)}{\Pr(ab,cd \mid H_2)} = U$$

- U depends on the frequencies of shared alleles

$$U=(x_1+x_2+x_3+x_4)/4$$

- Define x_1, x_2, x_3, x_4 for the allele pairs ac, ad, bc, bd
- $x_i=1/p_i$ if the two alleles are the same type
- $x_i=0$ if the 2 alleles are not the same type

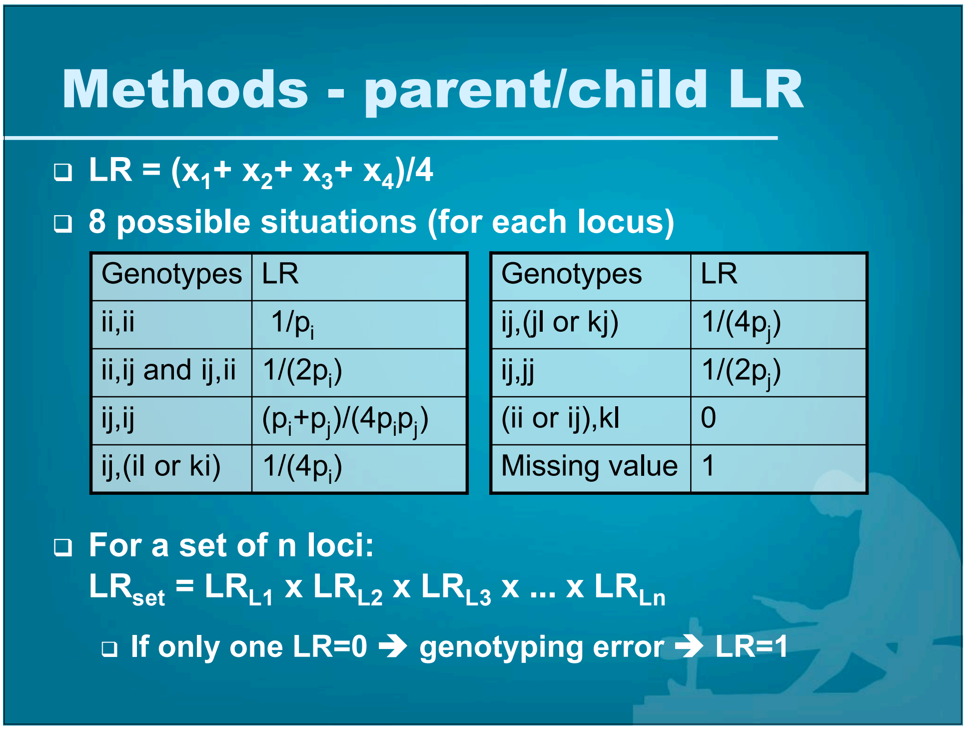


Methods - parent/child LR

- $LR = (x_1 + x_2 + x_3 + x_4)/4$
- 8 possible situations (for each locus)

Genotypes	LR
ii,ii	$1/p_i$
ii,ij and ij,ii	$1/(2p_i)$
ij,ij	$(p_i+p_j)/(4p_ip_j)$
ij,(il or ki)	$1/(4p_i)$

Genotypes	LR
ij,(jl or kj)	$1/(4p_j)$
ij,jj	$1/(2p_j)$
(ii or ij),kl	0
Missing value	1



Methods - parent/child LR

- $LR = (x_1 + x_2 + x_3 + x_4)/4$
- 8 possible situations (for each locus)

Genotypes	LR
ii,ii	$1/p_i$
ii,ij and ij,ii	$1/(2p_i)$
ij,ij	$(p_i+p_j)/(4p_ip_j)$
ij,(il or ki)	$1/(4p_i)$

Genotypes	LR
ij,(jl or kj)	$1/(4p_j)$
ij,jj	$1/(2p_j)$
(ii or ij),kl	0
Missing value	1

- For a set of n loci:
  $LR_{set} = LR_{L1} \times LR_{L2} \times LR_{L3} \times \ldots \times LR_{Ln}$
  - If only one LR=0 ➔ genotyping error ➔ LR=1

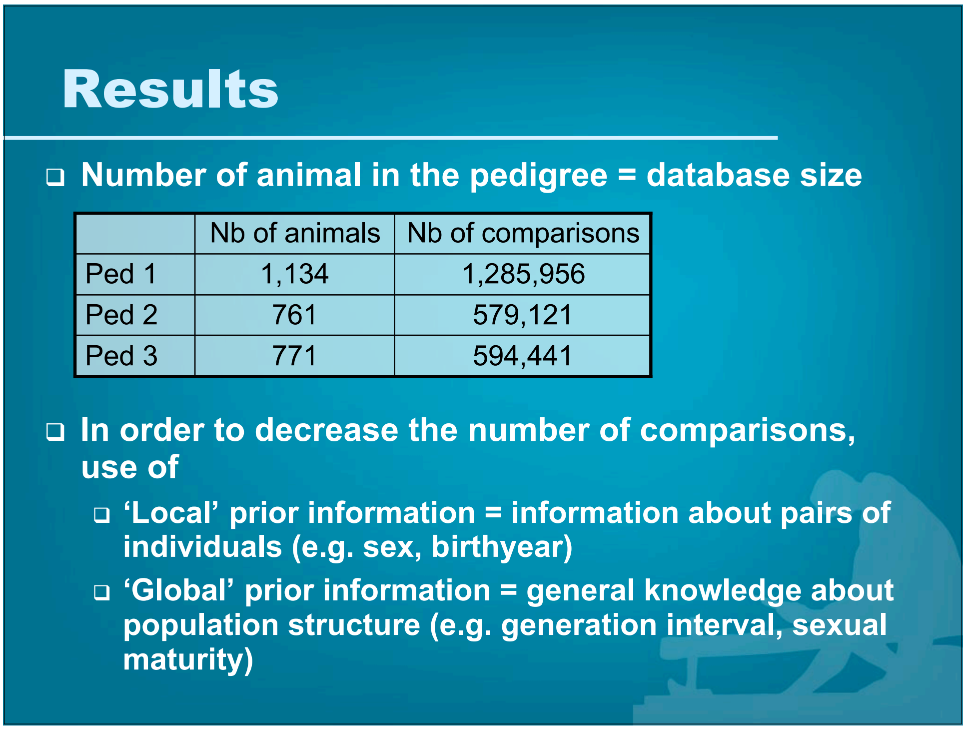
Data simulation

- Pedigree:
  - 100 years of simulation (1907-2007)
  - 3 repetitions with 3 levels of inbreeding
    - Ped1: ~17%
    - Ped2: ~27%
    - Ped3: ~37%
- Genotypes:
  - 25 microsatellites with 3 to 15 alleles
  - Equal allele frequencies in founder population
  - 5 repetitions/complete pedigree (data1 to data5)

Results

- Number of animal in the pedigree = database size

	Nb of animals	Nb of comparisons
Ped 1	1,134	1,285,956
Ped 2	761	579,121
Ped 3	771	594,441

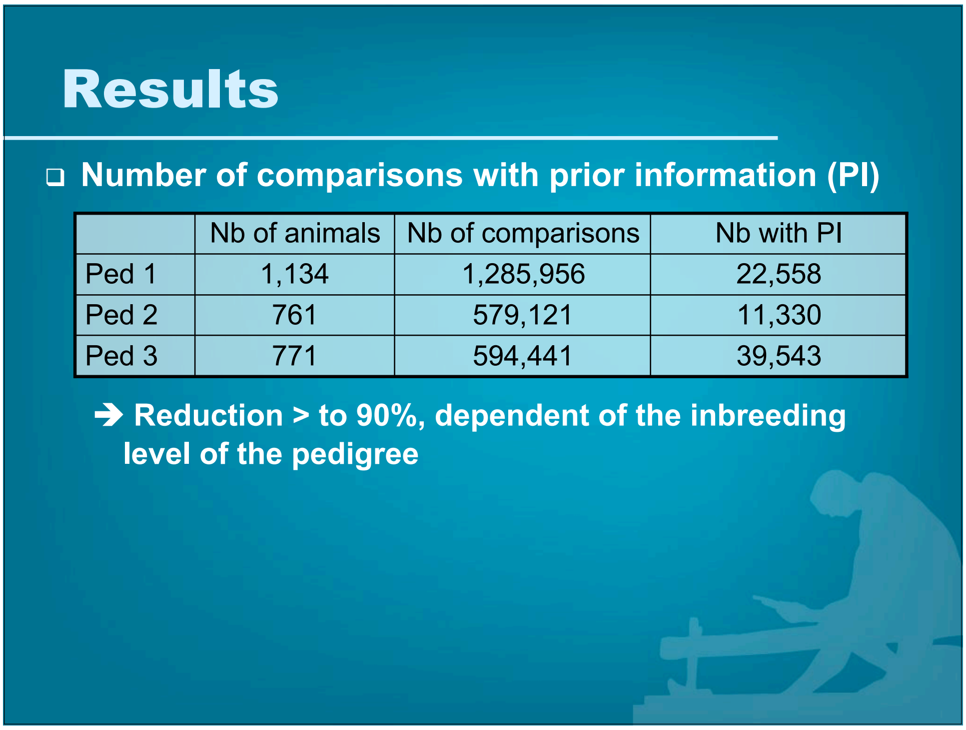


Results

- Number of animal in the pedigree = database size

	Nb of animals	Nb of comparisons
Ped 1	1,134	1,285,956
Ped 2	761	579,121
Ped 3	771	594,441

- In order to decrease the number of comparisons, use of
  - ‘Local’ prior information = information about pairs of individuals (e.g. sex, birthyear)
  - ‘Global’ prior information = general knowledge about population structure (e.g. generation interval, sexual maturity)



Results

- Number of comparisons with prior information (PI)

	Nb of animals	Nb of comparisons	Nb with PI
Ped 1	1,134	1,285,956	22,558
Ped 2	761	579,121	11,330
Ped 3	771	594,441	39,543

➔ Reduction > to 90%, dependent of the inbreeding level of the pedigree

Results

- ❑ **Number of comparisons with LR > 0 = possible parents**

	Nb of animals	Nb with PI	Nb with LR > 0
Ped 1	1,134	22,558	307
Ped 2	761	11,330	334
Ped 3	771	39,543	2,815

➔ Up to 98 % of the calculated parent-child combination are rejected (LR = 0), dependent from inbreeding

---

Results

- ❑ **Ranking of true parents (%)**

	Ped1 (F ≈ 17%)		Ped2 (F ≈ 27%)		Ped3 (F ≈ 37%)	
	Case A[1]	Case B[2]	Case A[1]	Case B[2]	Case A[1]	Case B[2]
In first position	49.26	49.68	48.44	49.53	42.75	44.94
In 1st and 2nd position	97.16	96.95	90.94	92.03	73.71	73.31
In first 5 positions	100.0	100.0	99.69	99.84	93.93	93.76
In first 10 positions			100.0	100.0	98.26	98.88
In first 40 positions					100.0	100.0

[1] **Case A: allele frequencies = frequencies from base population**

[2] **Case B: allele frequencies = frequencies from genotyped population**

Results

- Ranking of true parents (%)

	Ped1 (F ≈ 17%)		Ped2 (F ≈ 27%)		Ped3 (F ≈ 37%)	
	Case A[1]	Case B[2]	Case A[1]	Case B[2]	Case A[1]	Case B[2]
In first position	49.26	49.68	48.44	49.53	42.75	44.94
In 1st and 2nd position	97.16	96.95	90.94	92.03	73.71	73.31
In first 5 positions	100.0	100.0	99.69	99.84	93.93	93.76
In first 10 positions			100.0	100.0	98.26	98.88
In first 40 positions					100.0	100.0

➔ **High impact of inbreeding**

Highly inbred parents are more often miss-ranked: full- (half-) sibs of the (grand-)parents are in higher position

---

Results

- Mean values of LR

	Ped1 (F ≈ 17%)		Ped2 (F ≈ 27%)		Ped3 (F ≈ 37%)	
	Case A	Case B	Case A	Case B	Case A	Case B
In first position	4.68E15	2.43E8	1.08E17	7.19E6	4.53E18	3.07E6
In 1st and 2nd position	2.67E15	1.24E8	6.50E16	3.66E6	2.68E18	1.54E6
In first 5 positions	1.41E15	6.23E7	2.84E16	1.47E6	1.24E18	6.18E5
In first 10 positions			2.04E16	1.22E6	7.58E17	3.09E5
In first 40 positions					3.38E17	1.54E5
Max and min values	1.33E17 2.54E11	5.27E10 72.22	2.13E18 3.25E13	5.97E8 17.92	5.46E20 1.04E15	2.10E9 0.272

Results

- Mean values of LR

	Ped1 (F ≈ 17%)		Ped2 (F ≈ 27%)		Ped3 (F ≈ 37%)	
	Case A	Case B	Case A	Case B	Case A	Case B
In first position	4.68E15	2.43E8	1.08E17	7.19E6	4.53E18	3.07E6
In 1st and 2nd position	2.67E15	1.24E8	6.50E16	3.66E6	2.68E18	1.54E6
In first 5 positions	1					
In first 10 positions						
In first 40 positions						
Max and min values	1.33E17 2.54E11	5.27E10 72.22	2.13E18 3.25E13	5.97E8 17.92	5.46E20 1.04E15	2.10E9 0.272

➔ **High impact of allele frequencies**

Disequilibrium in allele frequencies (rare vs very common alleles) decreases the mean value of LR and increases the range of values

Inbred parents, carrying common alleles have very low LR values = risk to reject true parentage

---

Application to real data

- **Skyros pony: an endangered Greek horse breed**
  - **Population size: about 200 individuals**
- **Available data for the breed**
  - **Partial pedigree**
    - **Total of 395 individuals**
    - **Pedigree deepness: 1.5 generation-equivalents**
  - **Genotypes of half of the living population (99 ind.)**

Application to real data

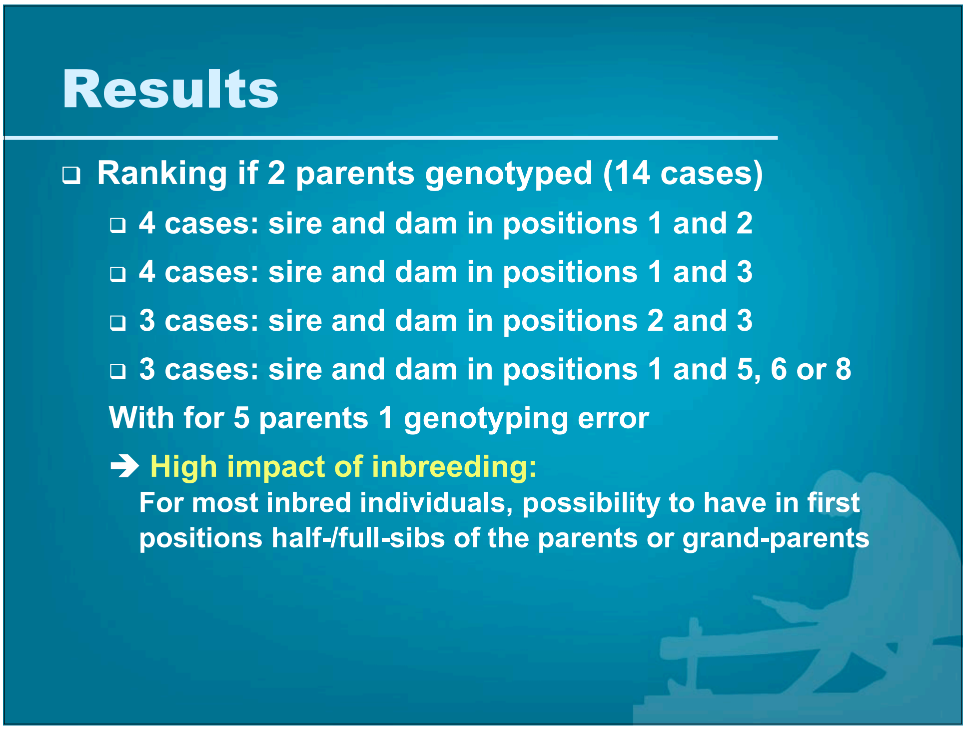
- Skyros pony - 99 individuals genotyped with
  - 2 parents genotyped (14 ind.)
  - Dam genotyped (28 ind.) → 62 parents
  - Sire genotyped (6 ind.)

- Application of Familial Searching
  - 1 genotyping error max

Results

- Ranking if 2 parents genotyped (14 cases)
  - 4 cases: sire and dam in positions 1 and 2
  - 4 cases: sire and dam in positions 1 and 3
  - 3 cases: sire and dam in positions 2 and 3
  - 3 cases: sire and dam in positions 1 and 5, 6 or 8

  With for 5 parents 1 genotyping error

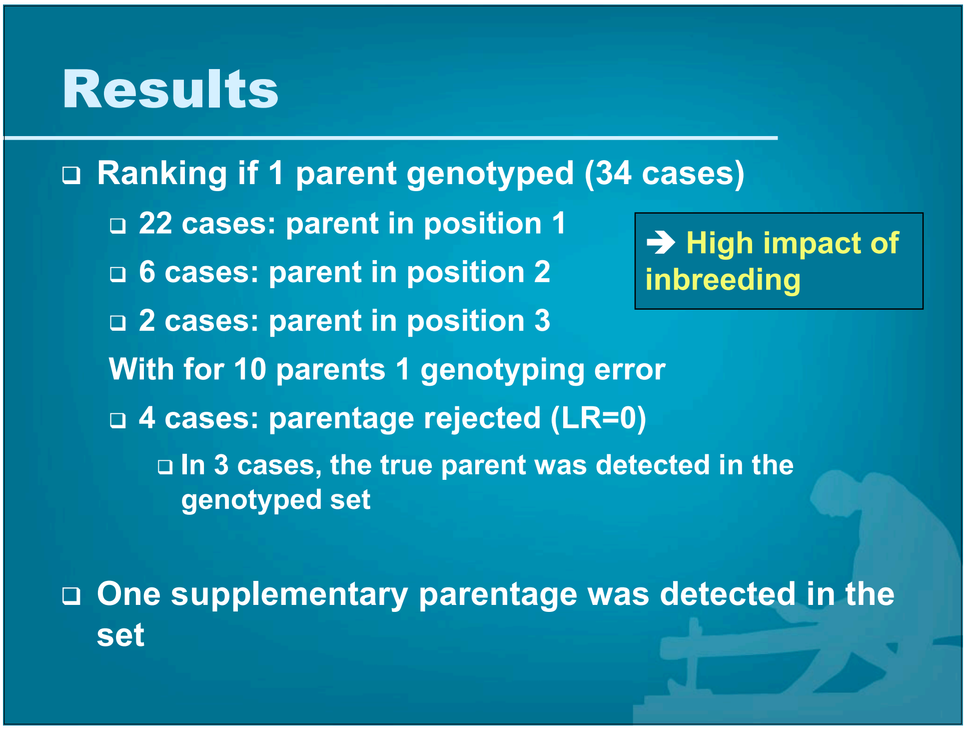


Results

- Ranking if 2 parents genotyped (14 cases)
  - 4 cases: sire and dam in positions 1 and 2
  - 4 cases: sire and dam in positions 1 and 3
  - 3 cases: sire and dam in positions 2 and 3
  - 3 cases: sire and dam in positions 1 and 5, 6 or 8

  With for 5 parents 1 genotyping error

  ➔ High impact of inbreeding:
  For most inbred individuals, possibility to have in first positions half-/full-sibs of the parents or grand-parents



Results

- Ranking if 1 parent genotyped (34 cases)
  - 22 cases: parent in position 1
  - 6 cases: parent in position 2
  - 2 cases: parent in position 3

  With for 10 parents 1 genotyping error
  - 4 cases: parentage rejected (LR=0)
    - In 3 cases, the true parent was detected in the genotyped set

➔ High impact of inbreeding

- One supplementary parentage was detected in the set

Conclusions

- Familial searching offers promising results for improvement of pedigrees
  ➔ Detection of non-recorded or false parentages
- Efficiency of programs can be increased using prior information
- Ranking / LR values of true parents influenced by
  - Inbreeding
  - Allele frequencies used to calculate LR
  - Presence of genotyping errors / missing values

Corresponding author's email:
elisabeth.bomcke@ulg.ac.be

Thank you for your attention!

Study supported by:

- National Fund for Scientific Research (FRS – FNRS)

gembloux agro bio tech

fnrs LA LIBERTÉ DE CHERCHER

Université de Liège ulg