

La ressemblance entre apparentés

La ressemblance entre apparentés

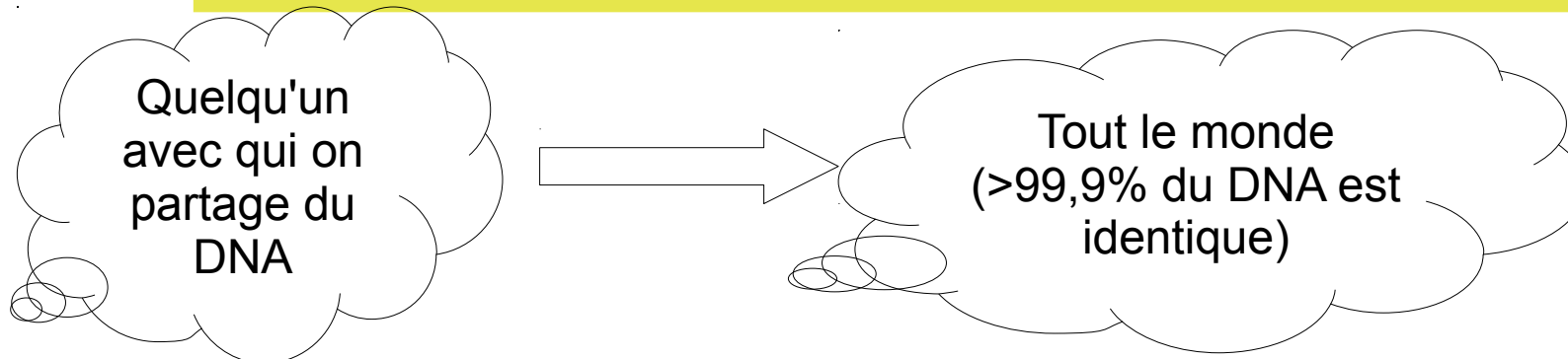
La ressemblance entre apparentés est un de principaux sujets de la génétique

Qu'est ce qu'un apparenté?

La ressemblance entre apparentés

La ressemblance entre apparentés est un de principaux sujets de la génétique

Qu'est ce qu'un apparenté?



De plus, ce n'est pas très pratique car nous voulons expliquer en quoi la génétique explique *la différence*

Tant pour comprendre les mécanismes évolutifs

Comme pour sélectionner ceux qui sont meilleurs

et pour prédire l'occurrence d'une maladie

La ressemblance entre apparentés

Qu'est ce qu'un apparenté?

Quelqu'un avec qui on partage du DNA

Tout le monde (>99,9% du DNA est identique)

Quelqu'un avec qui on partage des allèles aux gènes qui expliquent la ressemblance phénotypique

➤ Soit parce que l'on partage de vrais allèles aux gènes



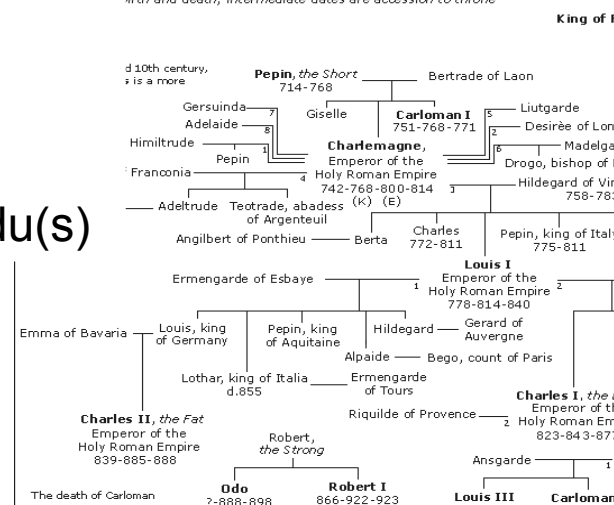
Partagent le type sanguin (O)

➤ Soit parce que *possiblement* on partage des allèles car on a une généalogie qui nous relie

- Mais les généalogies sont incomplètes
- Et de toutes manières nous descendons tous du même individu(s)

Kings of France family tree: 1st Dynasty: Carolingians (800-987)

birth and death; intermediate dates are accession to throne



La ressemblance entre apparentés

Clarification des notions intuitives

La notion d'identité par état (identity by state= IBS)

s'applique à des vrais gènes
mais aussi à les marqueurs (qui ne le sont pas)

La notion d'identité par descendance (identity by descent= IBD)

Le coefficient de parenté

La relation génétique additive

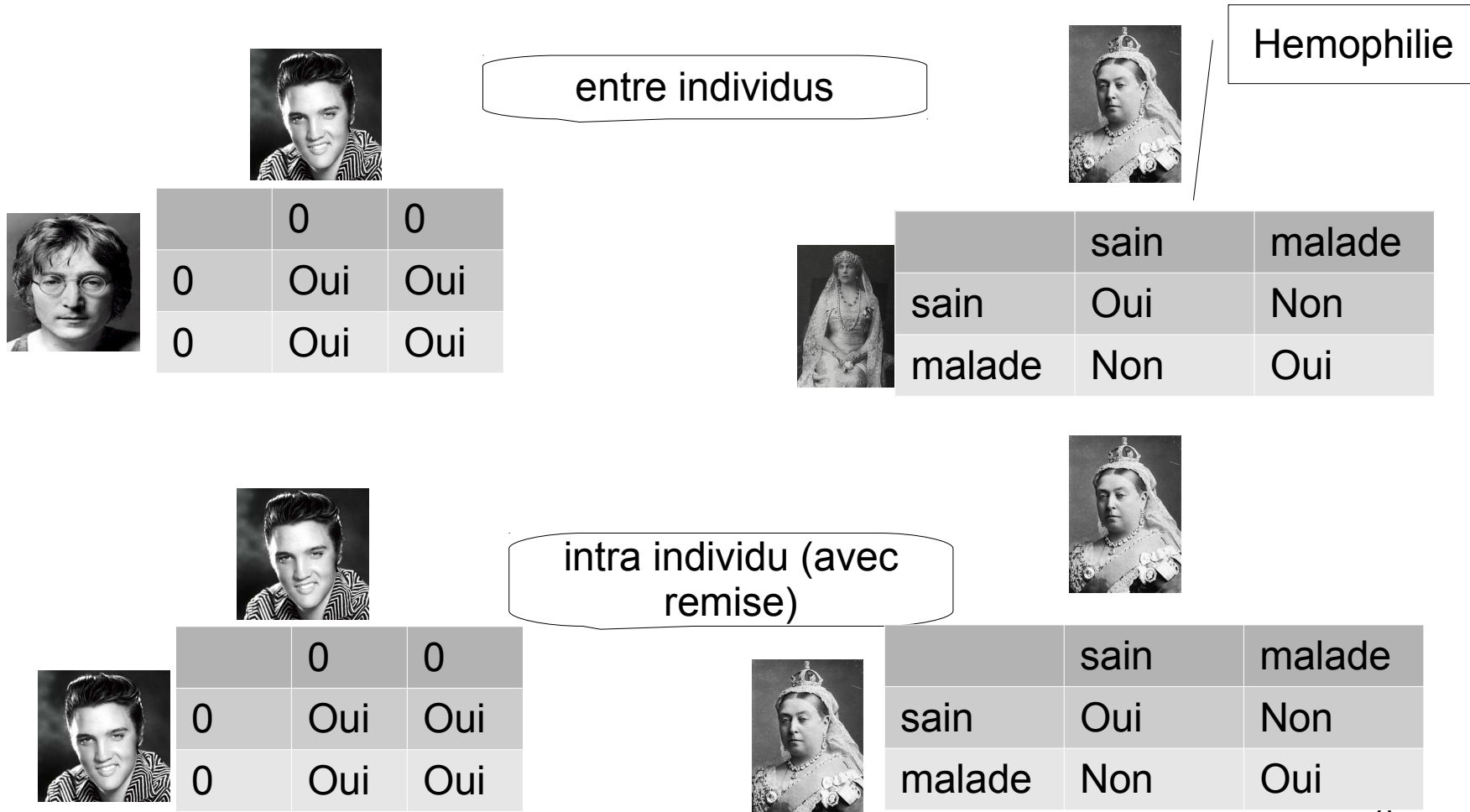
s'applique à des généalogies
(car on ne « voit » pas les gènes)

généralisation à tout un
ensemble d'individus

et finalement ça sert à expliquer la
ressemblance « phénotypique »

Identité par état

L'identité par état (*identity by state; IBS*) est le phénomène de que deux allèles au gène soient identiques

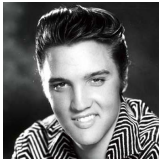


Identité par état

Cela se résume de deux manières:

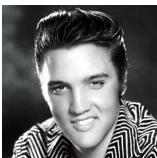
1- Probabilité d'identité par état


Pour deux individus i, j prenons deux allèles au hasard (avec remise si $i=j$). Quelle est la probabilité de qu'ils soient identiques?



1


	0	0
0	Oui	Oui
0	Oui	Oui





0.5

	sain	malade
sain	Oui	Non
malade	Non	Oui



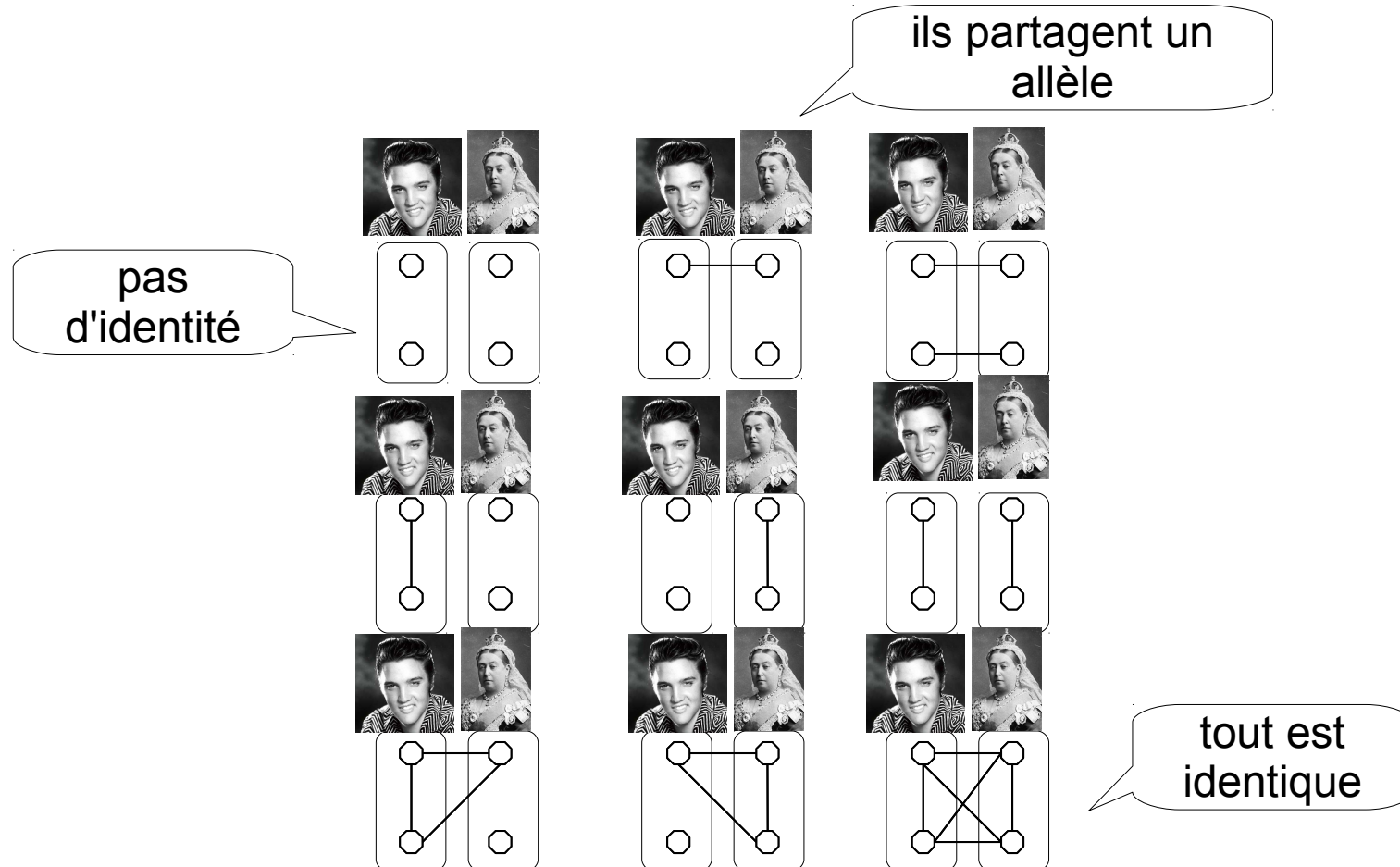
	A	B
C	Non	Non
B	Non	Oui

0.25

Identité par état

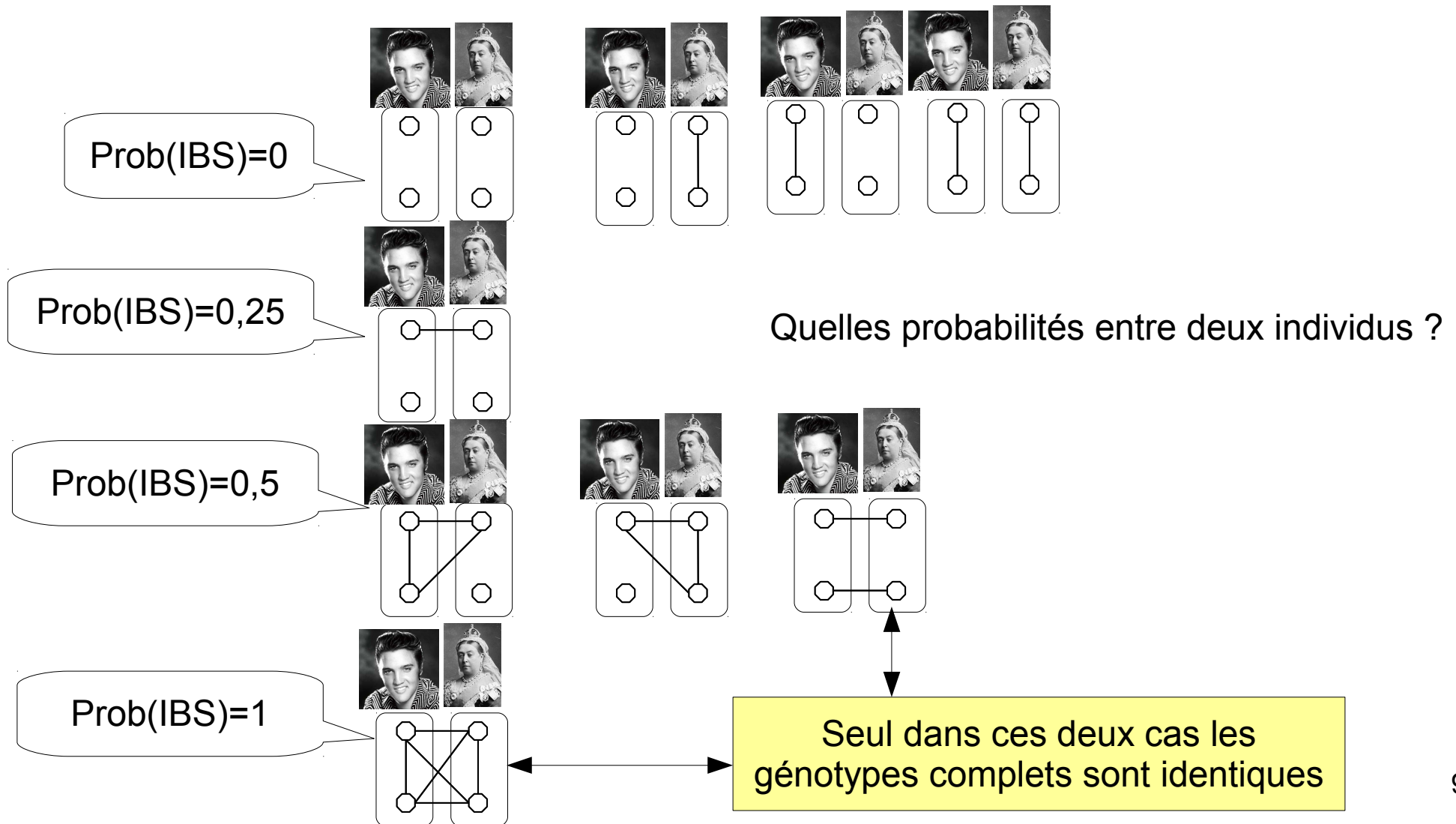
2- Etats d'identité

Quelles configurations d'identité peuvent prendre deux individus?



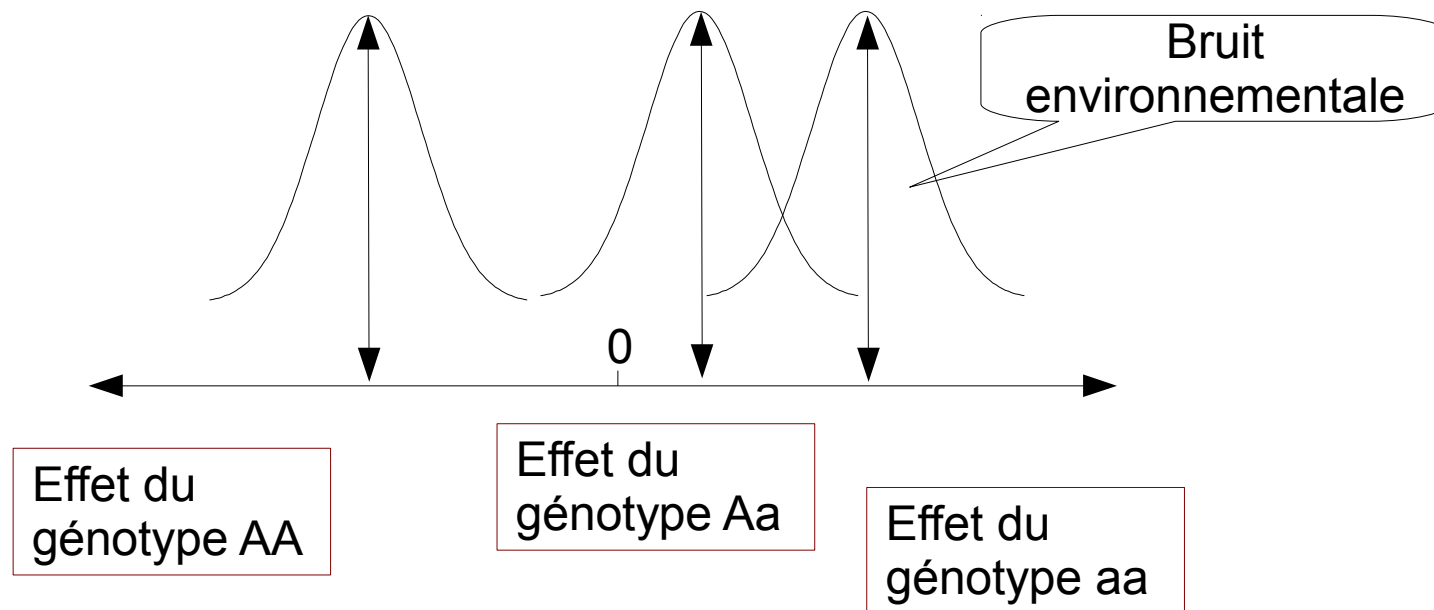
Identité par état

2- Etats d'identité

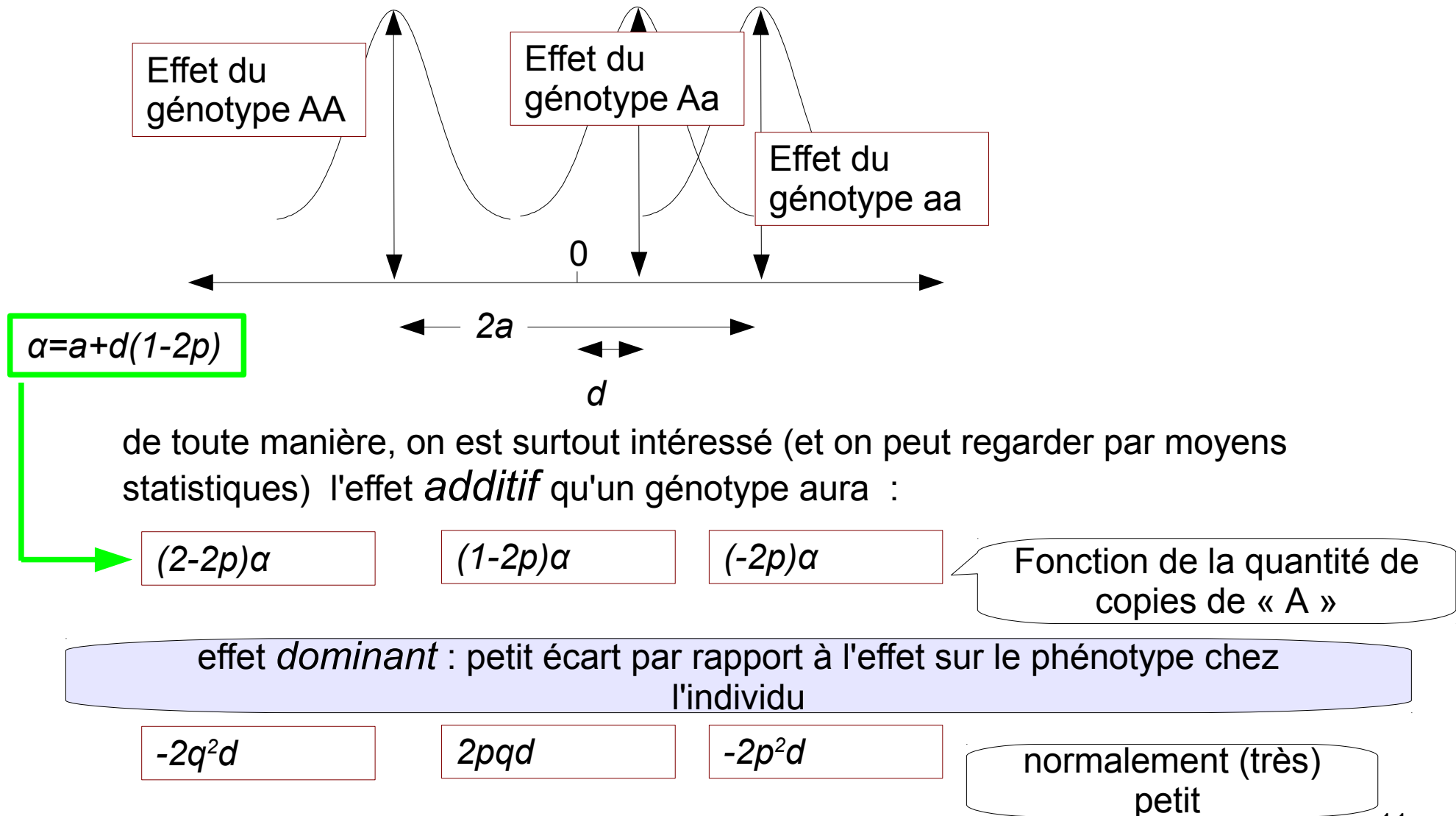


La ressemblance entre les phénotypes des apparentés

Si l'on voyait le gène, on pourrait facilement décrire la population



La ressemblance entre les phénotypes des apparentés

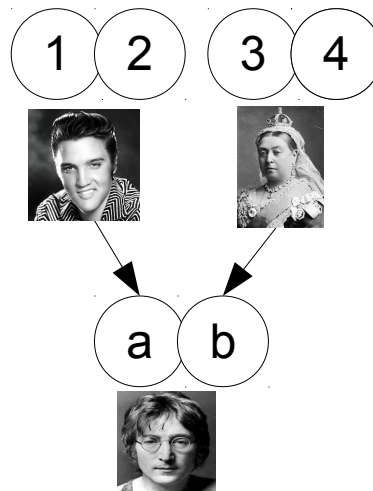


La ressemblance entre les phénotypes des apparentés

Donc on sait bien décrire les choses mais on ne voit pas ces gènes

Comment faire?

Hypothèse radical : chez les fondateurs (ceux qui n'ont pas de parents connus), *tous les allèles au gène sont différents*



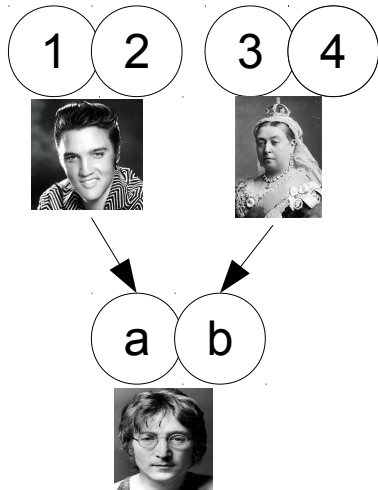
Supposons (c'est le plus fréquent) que les gènes sont transmis intacts (sans mutation)

Donc un allèle transmis est un allèle identique (la chaîne de DNA n'est pas modifié)

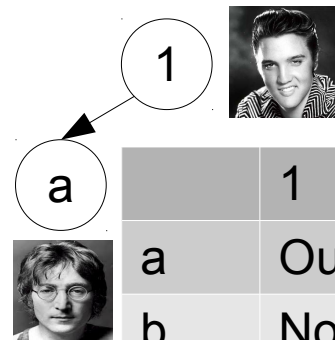
Probabilité d'identité par descendance (IBD de Malécot)

Pour deux individus i, j prenons deux allèles au hasard (avec remise si $i=j$). Quelle est la probabilité – Φ_{ij} – de qu'ils soient identiques par descendance? C'est à dire, qu'ils soient issus du même individu (ancêtre)

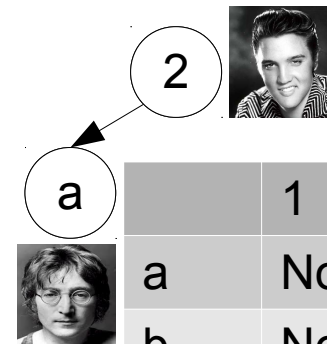
Probabilité d'identité par descendance



Événements possibles



	1	2
a	Oui	Non
b	Non	Non



	1	2
a	Non	Oui
b	Non	Non

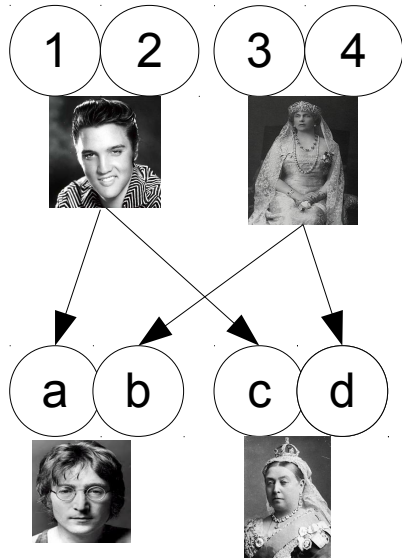
$$\Phi_{ij} = 2/8 = 0.25$$

Pour deux individus i, j prenons deux allèles au hasard (avec remise si $i=j$). Quelle est la probabilité de qu'ils soient identiques par descendance? C'est à dire, qu'ils aient le même origine ?

	1	2
1	Oui	Non
2	Non	Oui

$$\Phi_{ii} = 2/4 = 0.5$$

Probabilité d'identité par descendance



Événements possibles

	1	2
1	Oui	Non
2	Non	Oui

	3	4
1	Non	Non
2	Non	Non

$$\Phi_{ij} = 4/16 = 0.25$$

	1	2
3	Non	Non
4	Non	Non

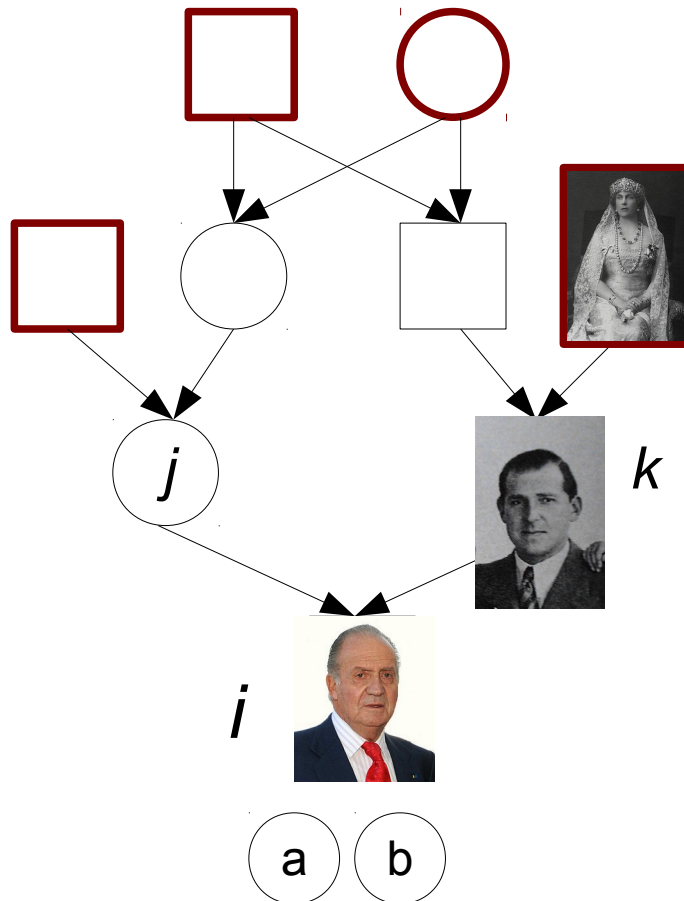
	3	4
3	Oui	Non
4	Non	Oui

Pour deux individus i, j prenons deux allèles au hasard (avec remise si $i=j$). Quelle est la probabilité de qu'ils soient identiques par descendance? C'est à dire, qu'ils aient le même origine ?

Probabilité d'identité par descendance

La consanguinité

Fondateurs



La coefficient de consanguinité (f) est la probabilité de que $a=b$ par descendance

La différence avec le coefficient de parenté est que le tirage soit fait avec remise (Φ_{ij}) ou pas (f_i) mais en fait

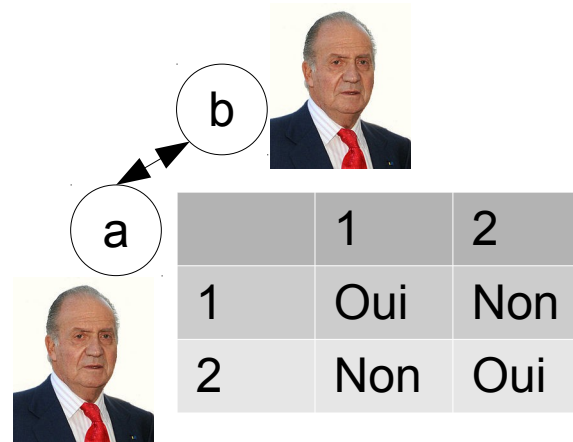
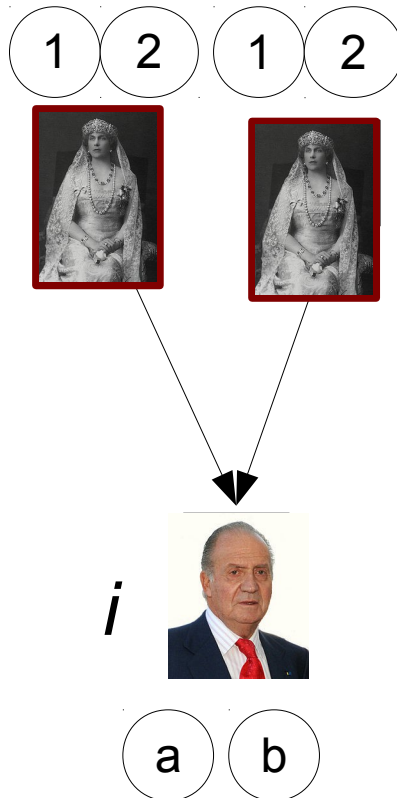
$$1+f_i=2\Phi_{ij}$$

et aussi, $f_i=\Phi_{jk}$, le coefficient de parenté de ses parents

Probabilité d'identité par descendance

La consanguinité

Fondateurs

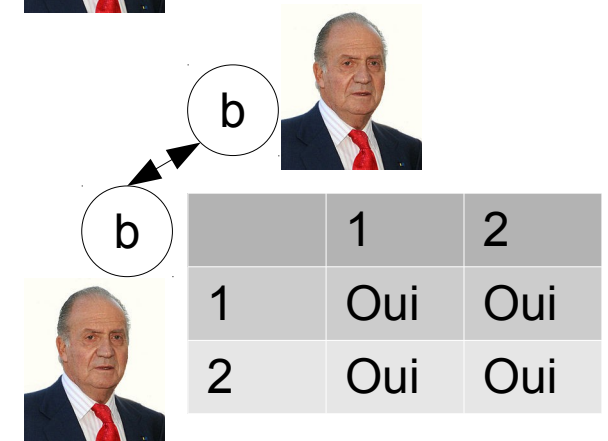
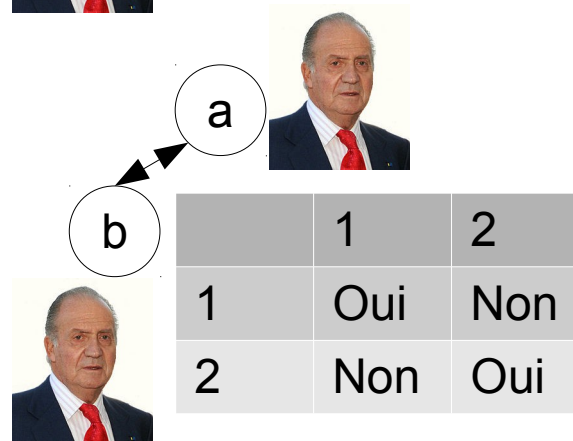
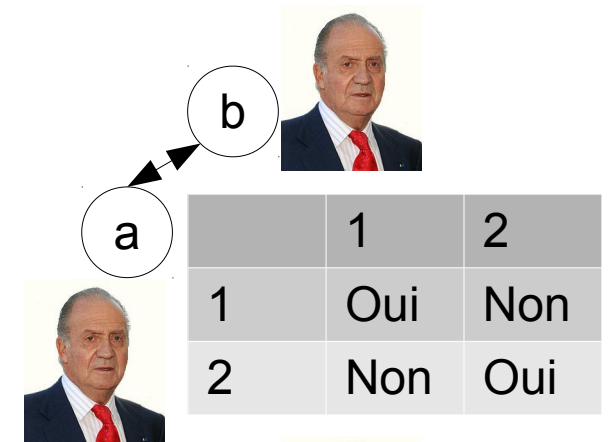
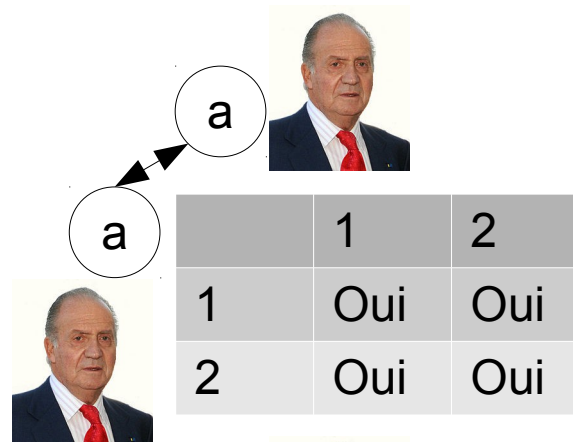
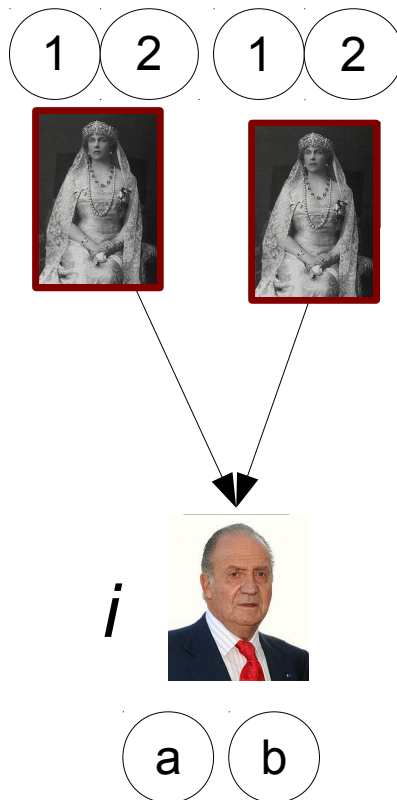


Pour cet individu, $f_i = 1/2$

Probabilité d'identité par descendance

La consanguinité

Fondateurs

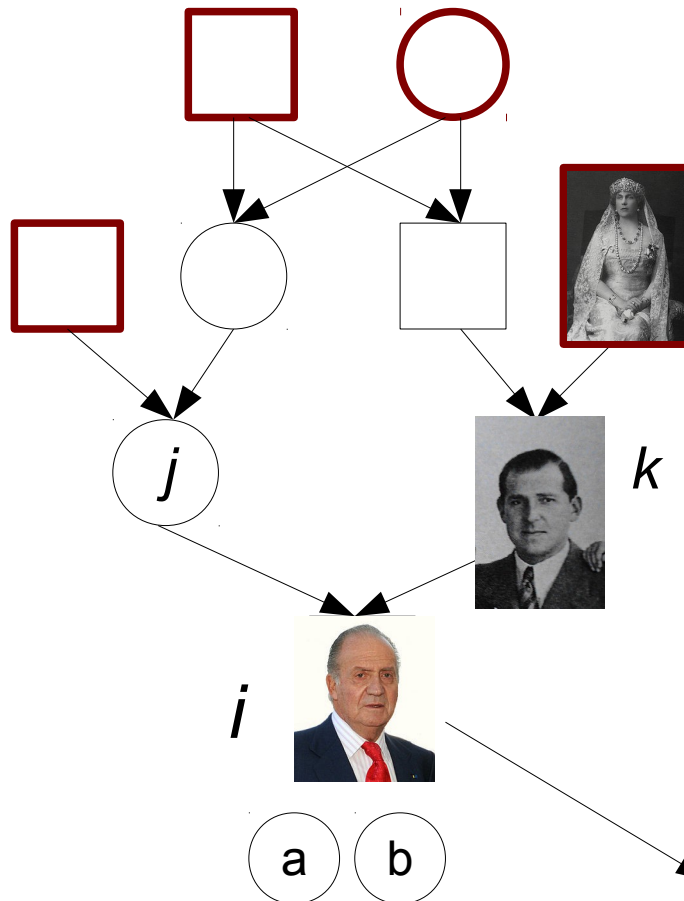


Pour cet individu, $\Phi_{ii} = 12/16 = 0,75$

Probabilité d'identité par descendance

La consanguinité

Fondateurs



La coefficient de consanguinité (f) est la probabilité de que $a=b$ par descendance

La différence avec le coefficient de parenté est que le tirage soit fait avec remise (Φ_{ij}) ou pas (f_i) mais en fait

$$1+f_i=2\Phi_{ij}$$

et aussi, $f_i=\Phi_{jk}$, le coefficient de parenté de ses parents

Pour cet individu, $f_i = 1/16$

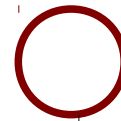
Probabilité d'identité par descendance

Le calcul

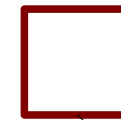
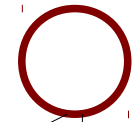
f et ϕ peuvent se calculer à la main dans les cas simples... ➡

i

Fondateurs



j

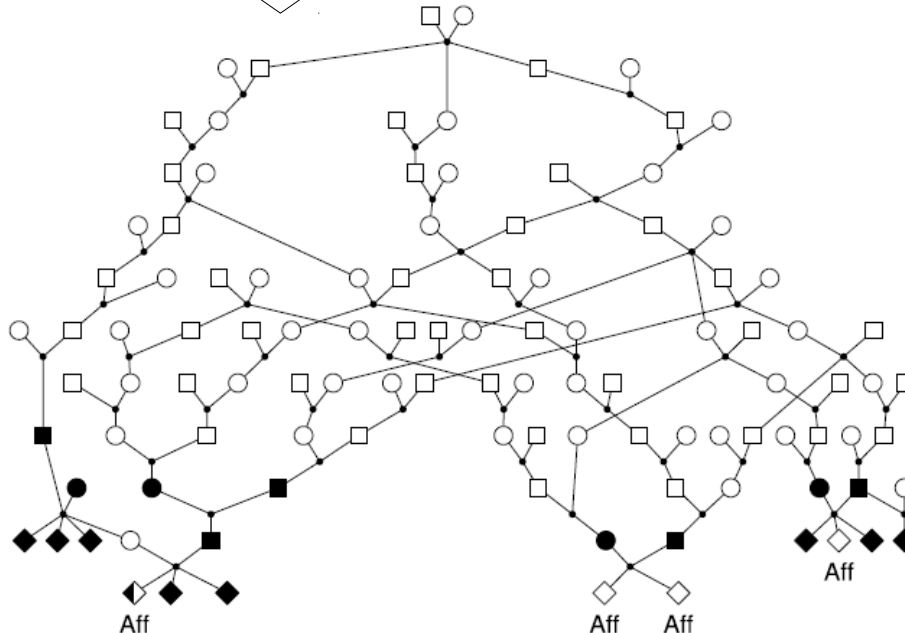


i



j

Il y a des cas compliqués... ⬇



Heureusement il y a des méthodes systématiques qui permettent de calculer des f et ϕ à l'intérieur des généalogies très complexes

- Voir TP

Probabilité d'identité par descendance

Utilisation

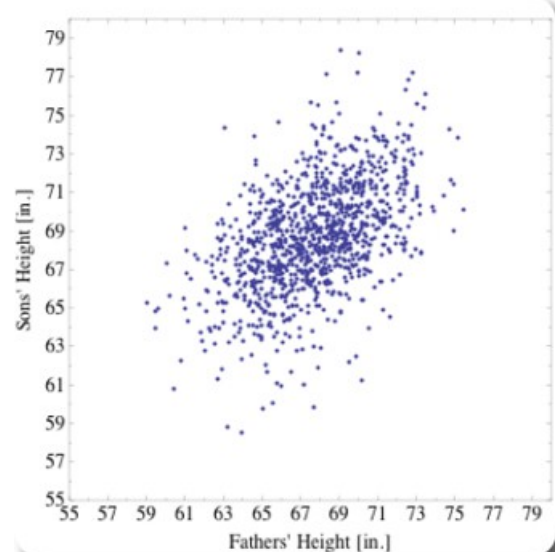
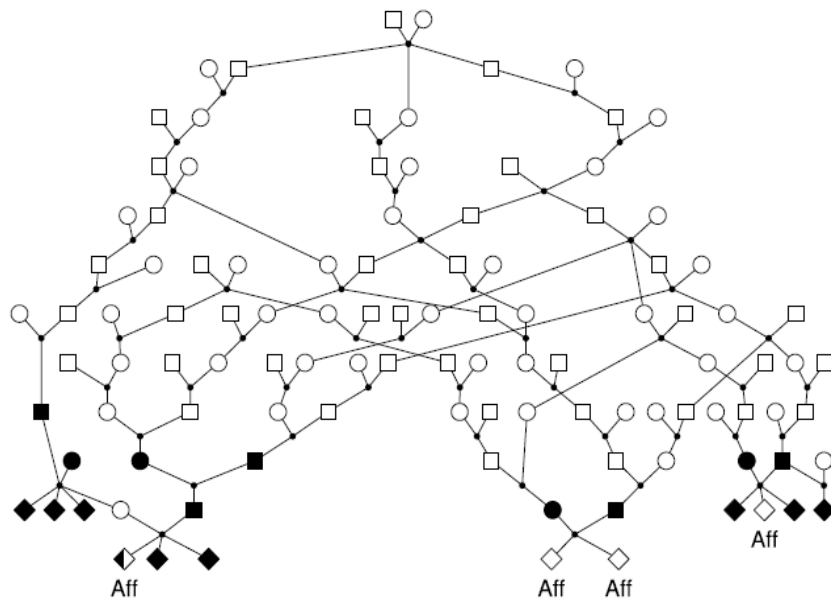
En quoi tous ces coefficients servent à quelque chose ?

Le coefficient de consanguinité f_i sert

- à établir de risques d'être atteint des maladies récessives
- à gérer des populations rares (espèces protégées, Zoos)
- À gérer des populations pour qu'elles n'aient pas de la dépression consanguine

Le coefficient de parenté Φ_{ij} sert

- à quantifier le degré de ressemblance (corrélation) phénotypique due au parenté, pour un caractère quantitatif



Probabilité d'identité par descendance

La relation génétique additive

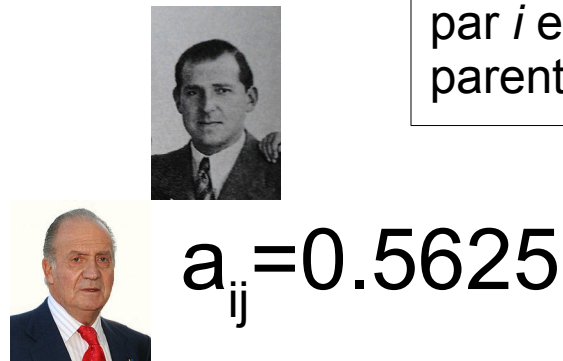
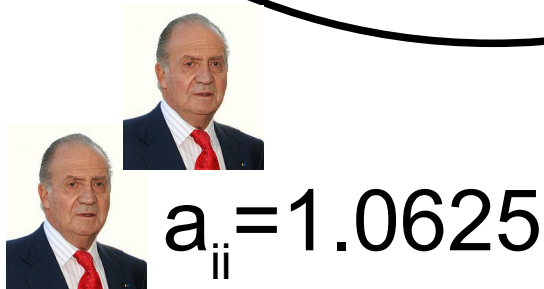
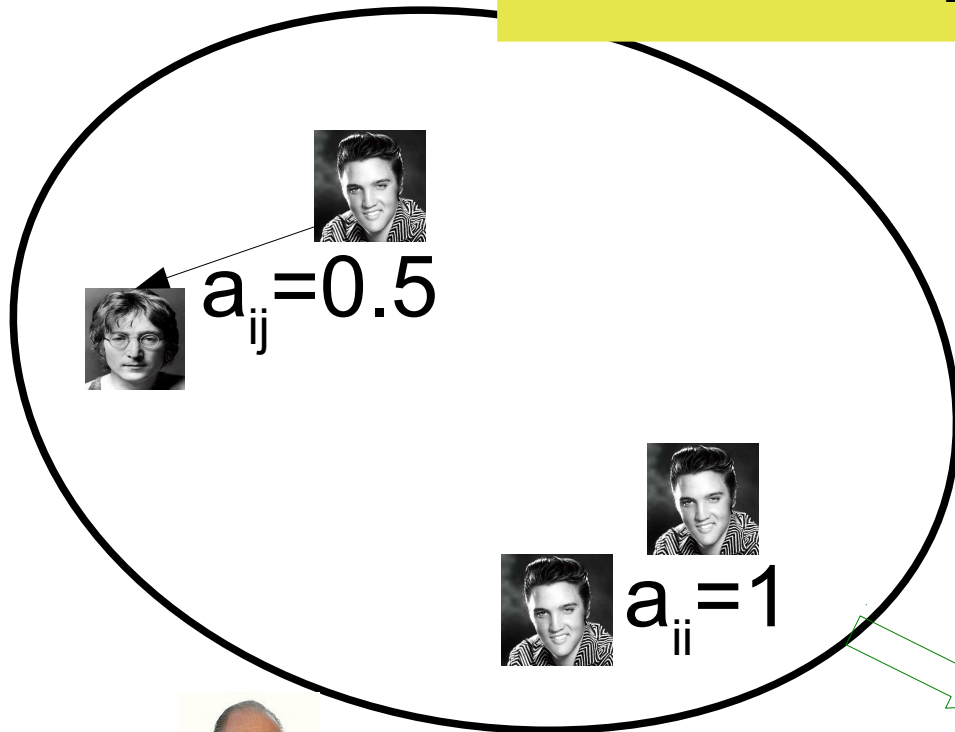
La relation additive est tout simplement

$$a_{ij} = 2\Phi_{ij}$$

du coup

$$a_{ii} = 2\Phi_{ii} = 1 + f_i$$

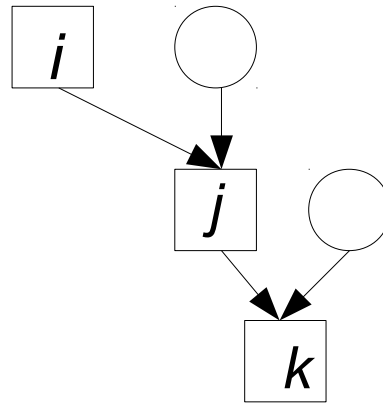
En absence de consanguinité, a_{ij} est la fraction du génome théoriquement partagée par i et j due au fait qu'ils ont de liens de parenté



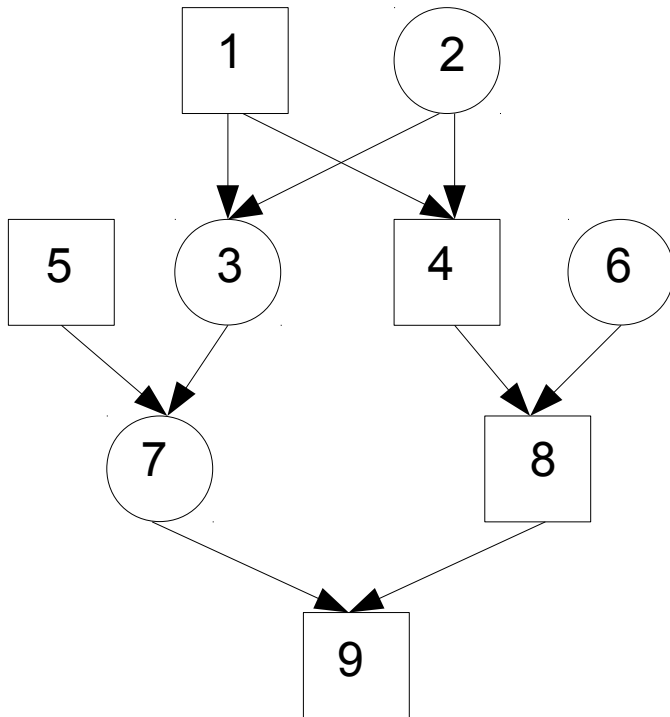
Probabilité d'identité par descendance

La matrice de parenté ou matrice de relations additives

On peut (et on doit) organiser les a_{ij}
de manière simple :
matrice de relations de parenté ou
matrice de parenté (**A**)



	i	j	k
i	1	0.5	0.25
j	0.5	1	0.5
k	0.25	0.5	1

[illegible]

Probabilité d'identité par descendance

Méthode tabulaire (Emik and Terrill 1949)

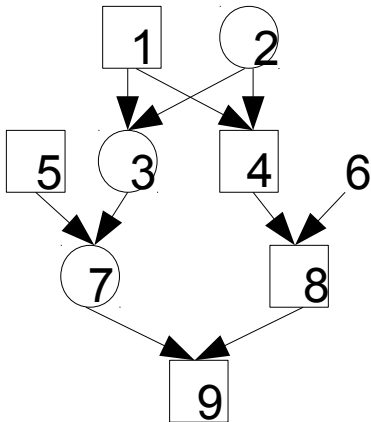
- On organise les individus par ordre de naissance
- On remplit les « carrés » 1, 2, ... par ordre avec deux règles :

$$a_{ii} = 1 + 0.5a_{pm}$$

$$a_{ij} = 0.5(a_{pj} + a_{dj})$$

$$a_{ji} = a_{ij}$$

Où p et m sont les parents de i et i est plus jeune que j

[illegible]

Probabilité d'identité par descendance

Méthode tabulaire (Emik and Terrill 1949)

Ped :
1 0 0
2 0 0
3 1 2
4 1 2
5 0 0
...etc

Probabilité d'identité par descendance

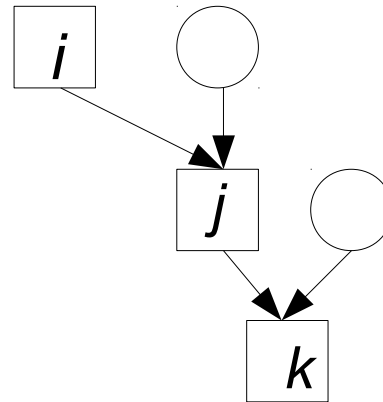
Méthode tabulaire (Emik and Terrill 1949)

```
getA <- function(pedigree){  
  # A, tabular method  
  A=matrix(0,nrow(pedigree),nrow(pedigree))  
  f=rep(0,nrow(pedigree))  
  #read pedigree  
  for(i in 1:nrow(pedigree)){  
    A[i,i]=1  
    if(pedigree[i,2]!=0 && pedigree[i,3]!=0){  
      A[i,i]=A[i,i]+0.5*A[pedigree[i,2],pedigree[i,3]]  
    }  
    f[i]=A[i,i]  
    for (j in 1:(i-1)){  
      if(pedigree[i,2]!=0) A[i,j]=A[i,j]+0.5*A[j,pedigree[i,2]]  
      if(pedigree[i,3]!=0) A[i,j]=A[i,j]+0.5*A[j,pedigree[i,3]]  
      A[j,i]=A[i,j]  
    }  
  }  
  list(A=A, f=f)  
}
```

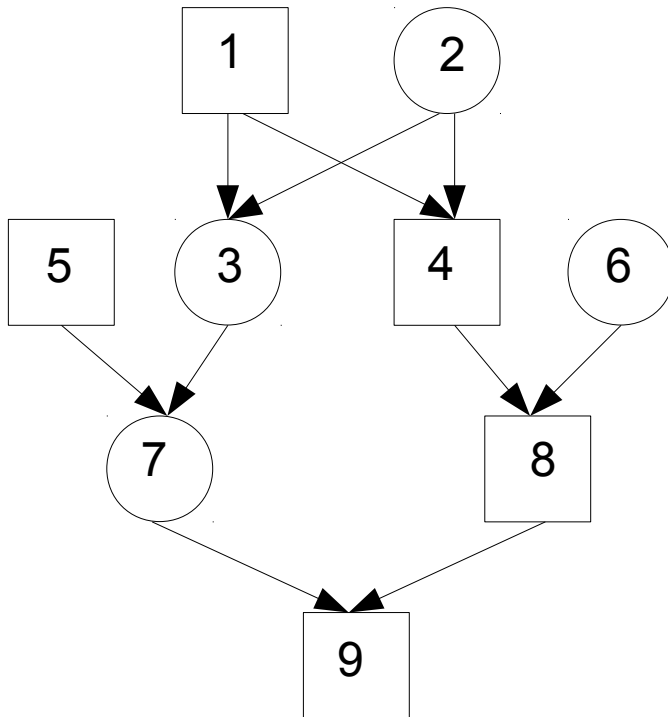
Probabilité d'identité par descendance

La matrice de parenté ou matrice de relations additives

On peut (et on doit) organiser les a_{ij} de manière simple :
matrice de relations de parenté ou *matrice de parenté (A)*



	i	j	k
i	1	0.5	0.25
j	0.5	1	0.5
k	0.25	0.5	1



	1	0	0,5	0,5	0	0	0,25	0,25	0,25
0	1	0,5	0,5	0	0	0,25	0,25	0,25	0,25
0,5	0,5	1	0,5	0	0	0,5	0,25	0,375	0,375
0,5	0,5	0,5	1	0	0	0,25	0,5	0,375	0,375
0	0	0	0	1	0	0,5	0	0,25	0,25
0	0	0	0	0	1	0	0,5	0,25	0,25
0,25	0,25	0,5	0,25	0,5	0	1	0,125	0,5625	0,5625
0,25	0,25	0,25	0,5	0	0,5	0,125	1	0,5625	0,5625
0,25	0,25	0,375	0,375	0,25	0,25	0,5625	0,5625	1,0625	1,0625

Probabilité d'identité par descendance et ressemblance entre les phénotypes des individus

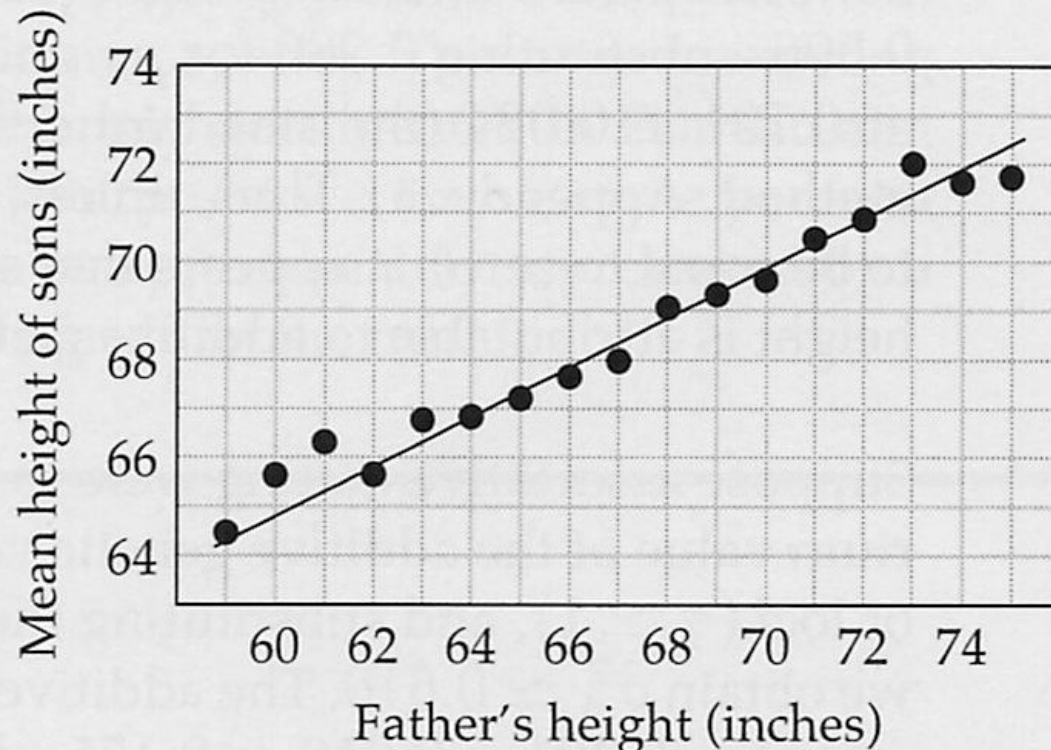
La covariance entre les phénotypes mesurés sur différent individus

Example 5. As an example of the application of the preceding theory, we consider an early data set on height in British families (Pearson and Lee 1903). Pearson recruited college students to obtain data from approximately 1300 families, recording whenever possible the stature of father, mother, and eldest son and daughter (ignoring offspring less than 18 years of age). This was a very large data set for the precomputer era, and it took two years to calculate the statistics by hand. The data are remarkable for their essentially normal distribution and for the linearity of the regressions between relatives.

Probabilité d'identité par descendance et ressemblance entre les phénotypes des individus

La covariance entre les phénotypes mesurés sur différent individus

Example 5. As an example of the application of the preceding theory, we consider an early data set on height in British families (Pearson and Lee 1903). Pearson recruited college students to obtain data from approximately 1300 families, recording whenever possible the stature of father, mother, and eldest son and daughter (ignoring offspring set for the precomputer era, hand. The data are remarkable for the linearity of the regression



Probabilité d'identité par descendance et ressemblance entre les phénotypes des individus

La covariance entre les phénotypes mesurés sur différent individus

Répétition expérience de Pearson (Gillespie)

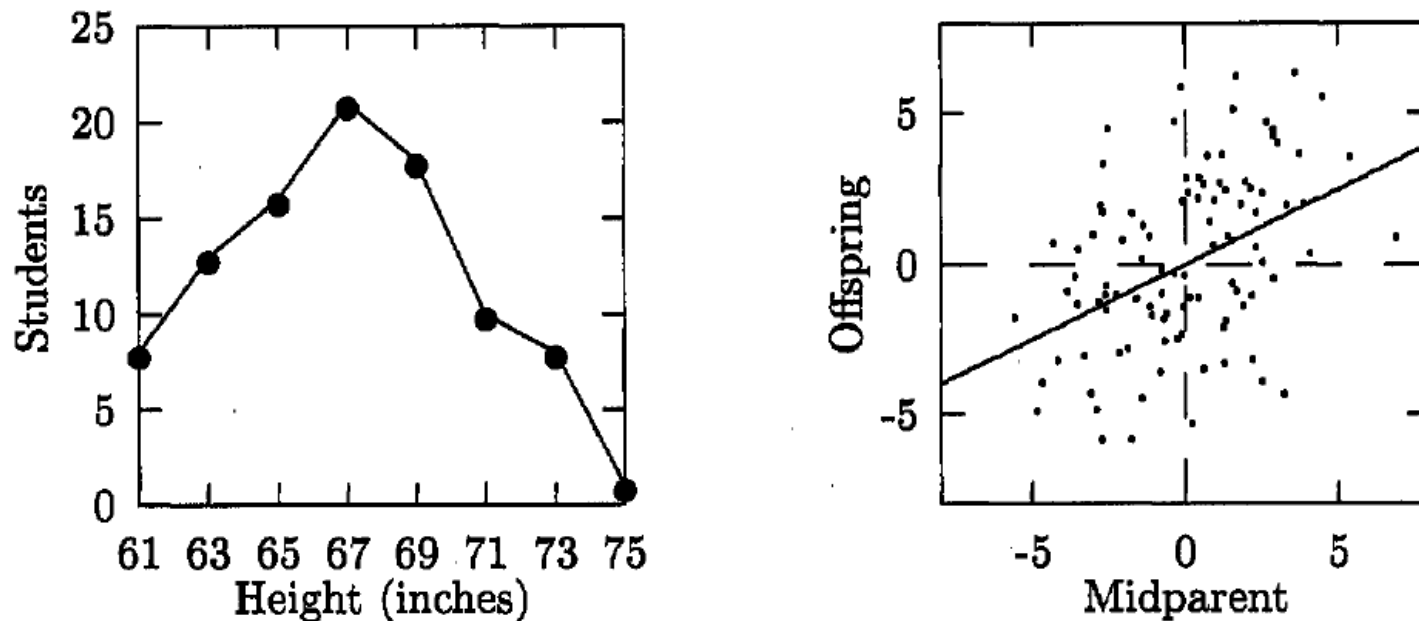


Figure 5.1: The left-hand figure is a histogram of the number of students of a particular height in an evolution class at UC Davis. The right-hand figure graphs the deviation of a student's height from the population mean against the deviation of the student's parents' average height from the population mean.

Probabilité d'identité par descendance et ressemblance entre les phénotypes des individus

La covariance entre les phénotypes mesurés sur différent individus

...encore une avec des espèces sauvages...

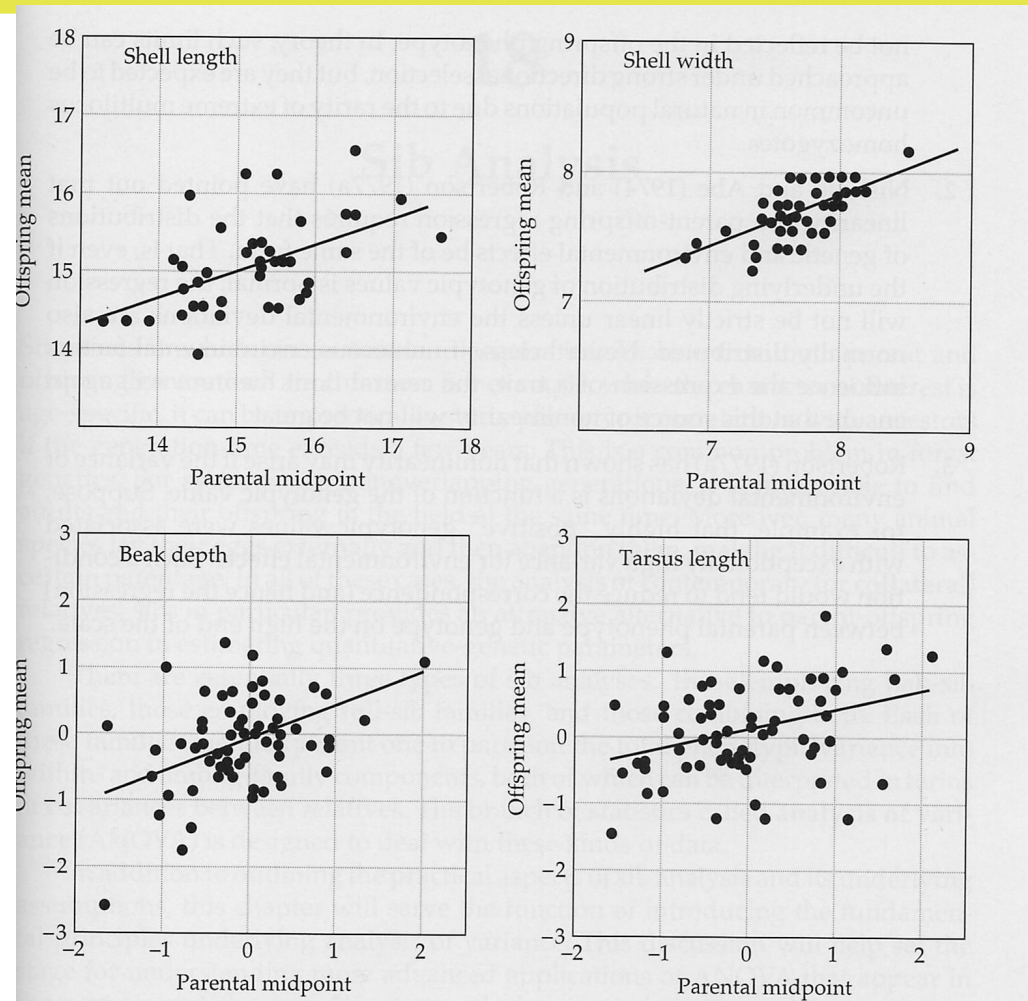
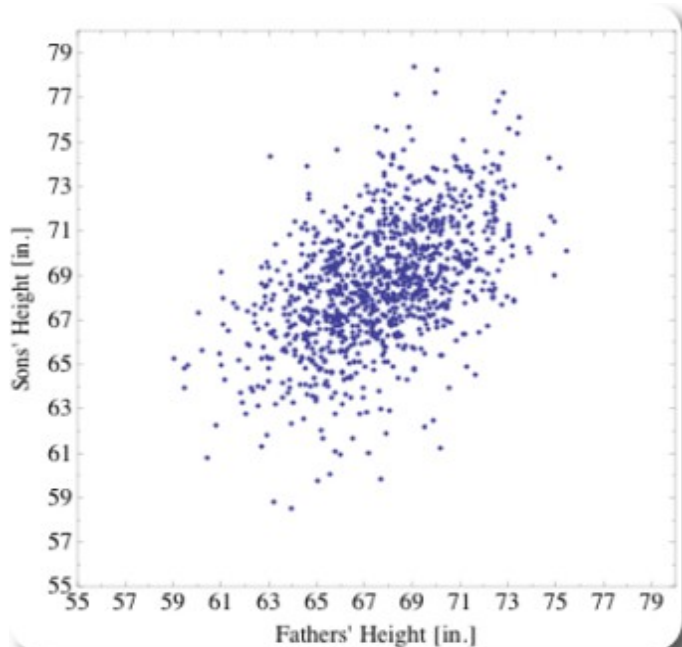


Figure 17.4 Examples of midparent-offspring regressions. **Top panels:** Land snail *Partula taeniata* (Murray and Clarke 1968). **Bottom panels:** Song sparrow *Melospiza melodia* (Smith and Zach 1979).

Probabilité d'identité par descendance et ressemblance entre les phénotypes des individus

La covariance entre les phénotypes mesurés sur différent individus



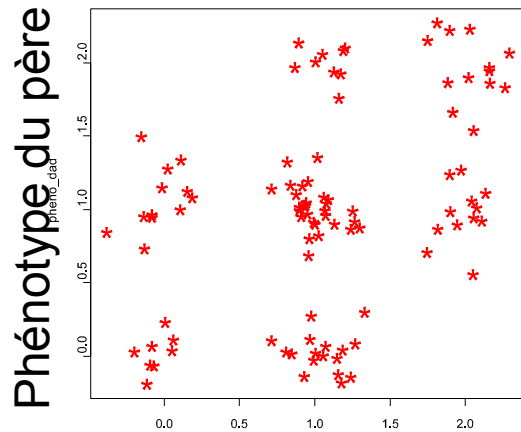
Pearson observa une corrélation de 0,5 père-fils pour la taille

Les « mendéliennistes » trouvaient louche cet notion de corrélation car elle était en conflit avec la ségrégation mendélienne

Mais en fait rien n'est en conflit, car on ne voit pas la ségrégation à cause de la multitude de gènes causaux

Probabilité d'identité par descendance et ressemblance entre les phénotypes

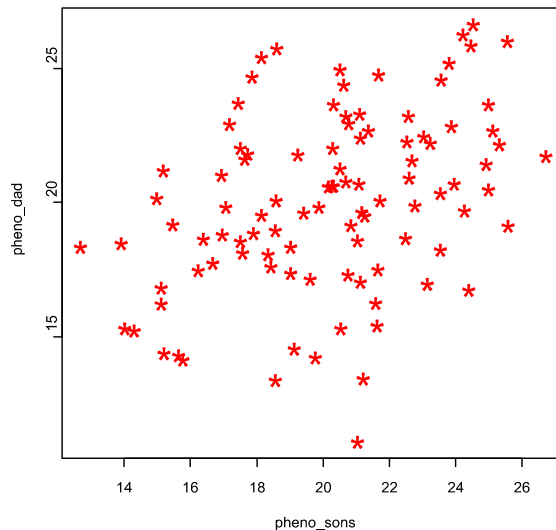
Distribution des phénotypes, n loci causaux bialléliques inconnues, $h^2=0,95$



1 locus

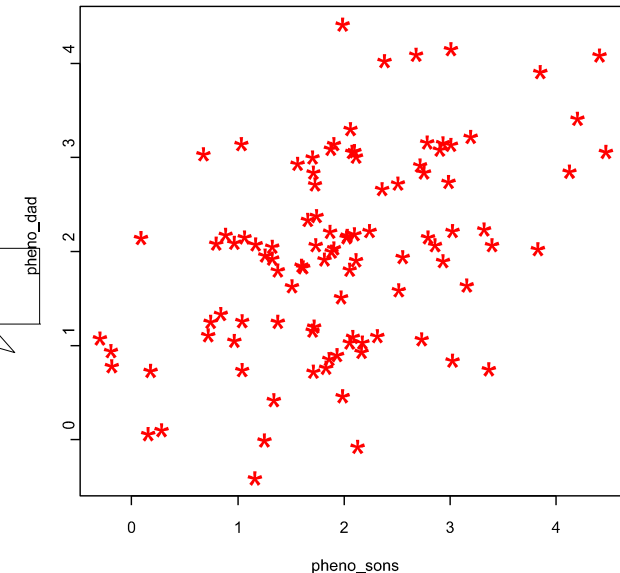
On voit bien tous les génotypes possibles du père et fils

Phénotype du fils



20 loci

On voit plus rien...
néanmoins la corrélation est
toujours de 0,5



2 loci

Probabilité d'identité par descendance et ressemblance entre les phénotypes

La covariance entre individus

$$P = G + E$$

Soit la décomposition de la valeur phénotypique :

$$\text{Var}(P) = \text{Var}(G) + \text{Var}(E)$$

$$\text{Cov}(G, E) = 0$$

$$G = \alpha + \alpha' + \delta_{ij}$$

La **valeur génétique additive (A)** d'un individu est la somme des effets moyens des allèles (α et α') qu'il possède.

L'effet d'interaction est appelé **valeur ou résidu de dominance (D)**

Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

Soit 2 individus i et j appartenant à la même population

$$P_i = G_i + E_i \quad P_j = G_j + E_j$$

• Lorsque l'on peut admettre l'absence d'effet d'environnement commun (E_i, E_j non corrélés), on peut démontrer que :

✓ l'expression de la covariance entre valeurs génétiques et phénotypiques est

$$\text{cov}(P_i, P_j) = \text{cov}(G_i, G_j)$$

$$\text{cov}(G_i, G_j) = \text{cov}(A_i, A_j) + \text{cov}(D_i, D_j) = a_{ij} V_A + d_{ij} V_D$$

Coefficient de relation
génétique additive



$$a_{ij} = 2 \Phi_{ij} \quad (\Phi_{ij} = \text{coefficient de parenté entre } i \text{ et } j)$$



Variance des valeurs génétiques
(additives) à la génération des fondateurs

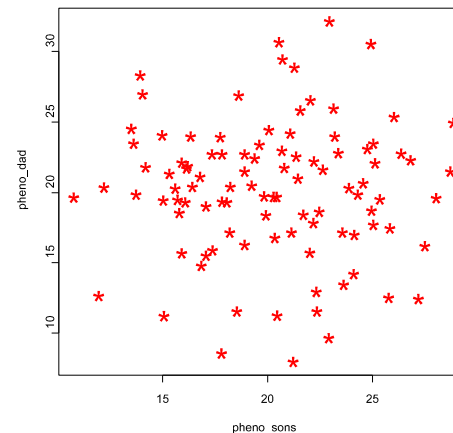
Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

$$\text{cov}(P_i, P_j) = \text{cov}(G_i, G_j) = \text{cov}(A_i, A_j) + \text{cov}(D_i, D_j) = a_{ij} V_A + d_{ij} V_D$$

Donc on peut, par exemple, estimer la variance additive génétique à partir de la covariance phénotypique entre individus

- 1- estime la covariance phénotypique entre deux séries d'individus ayant le même coefficient de parenté (père fils par exemple)
- 2- calcule $V_a = \text{Cov} / a_{ij}$
- 3- calcule la variance brute phénotypique (V_p)
- 4- calcule l'heritabilité du caractère comme V_a / V_p



Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

$$\text{cov}(P_i, P_j) = \text{cov}(G_i, G_j) = \text{cov}(A_i, A_j) + \text{cov}(D_i, D_j) = a_{ij} V_A + d_{ij} V_D$$

On peut aussi estimer la valeur génétique d'un individu par régression à partir de son phénotype P_i

$$\hat{G}_i = \frac{\text{cov}(P_i, G_i)}{\text{Var}(P_i)} P_i = \frac{a_{i,i} V_A}{a_{i,i} V_A + V_E} P_i$$

et d'un individu i par rapport à un autre j

$$\hat{G}_i = \frac{\text{cov}(G_i, G_j)}{\text{Var}(G_j)} \hat{G}_j = \frac{a_{i,j} V_A}{a_{j,j} V_A} \hat{G}_j = \frac{a_{i,j}}{a_{j,j}} \hat{G}_j$$

Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

La covariance entre apparentés généralisé

On peut généraliser cette équation à des nombreux individus simultanément *via* la matrice de parenté **A**

$$\text{Var} \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_n \end{pmatrix} = \begin{pmatrix} \text{Var}(G_1, G_1) & \text{Cov}(G_1, G_2) & \dots & \text{Cov}(G_1, G_n) \\ \text{Cov}(G_2, G_1) & \text{Var}(G_2, G_2) & \dots & \text{Cov}(G_2, G_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(G_n, G_1) & \text{Cov}(G_n, G_2) & \dots & \text{Var}(G_n, G_n) \end{pmatrix}$$
$$\text{Var} \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_n \end{pmatrix} = \mathbf{A} \mathbf{V}_A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix} \mathbf{V}_A$$

Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

La covariance entre apparentés généralisé

$$\text{Var} \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_n \end{pmatrix} = \begin{pmatrix} \text{Var}(G_1, G_1) & \text{Cov}(G_1, G_2) & \dots & \text{Cov}(G_1, G_n) \\ \text{Cov}(G_2, G_1) & \text{Var}(G_2, G_2) & \dots & \text{Cov}(G_2, G_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(G_n, G_1) & \text{Cov}(G_n, G_2) & \dots & \text{Var}(G_n, G_n) \end{pmatrix}$$

$$\text{Var} \begin{pmatrix} G_1 \\ G_2 \\ G_3 \end{pmatrix} = A V_A = \begin{pmatrix} 1 & 0,5 & 0,25 \\ 0,5 & 1 & 0,5 \\ 0,25 & 0,5 & 1 \end{pmatrix} V_A$$

On peut généraliser cette équation à des nombreux individus simultanément *via* la matrice de parenté **A**

Cela permet d'introduire toute l'information en même temps

- 1) Estimation de valeurs génétiques : BLUP
- 2) Estimation de paramètres génétiques
- 3) Correction du « fond » génétique pour la localisation de gènes causaux (QTL)

Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

La covariance entre apparentés généralisé

$$Var \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_n \end{pmatrix} = \begin{pmatrix} Var(G_1, G_1) & Cov(G_1, G_2) & \dots & Cov(G_1, G_n) \\ Cov(G_2, G_1) & Var(G_2, G_2) & \dots & Cov(G_2, G_n) \\ \dots & \dots & \dots & \dots \\ Cov(G_n, G_1) & Cov(G_n, G_2) & \dots & Var(G_n, G_n) \end{pmatrix}$$

On peut généraliser cette équation à des nombreux individus simultanément *via* la matrice de parenté **A**

Pour les « stateux » : la manière d'introduire cette matrice est avec la distribution normal multivariée :

$$P(G_1, G_2, \dots, G_n | A, V_A) = N(0, A V_A)$$

Probabilité d'identité par descendance et ressemblance entre les phénotypes

La dépression consanguine

Phénomène observé chez les espèces *non* autofécondantes (mammifères, maïs...)

Les individus consanguin sont moins viables (productifs, grands, etc.) que ce qui ne le sont pas

Explication biologique possible : accumulation d'effets négatifs de gènes à action récessive

A éviter (par gestion de la consanguinité) :

- Éviter les couples d'individus apparentés proches

Probabilité d'identité par descendance et ressemblance entre les phénotypes

La dépression consanguine

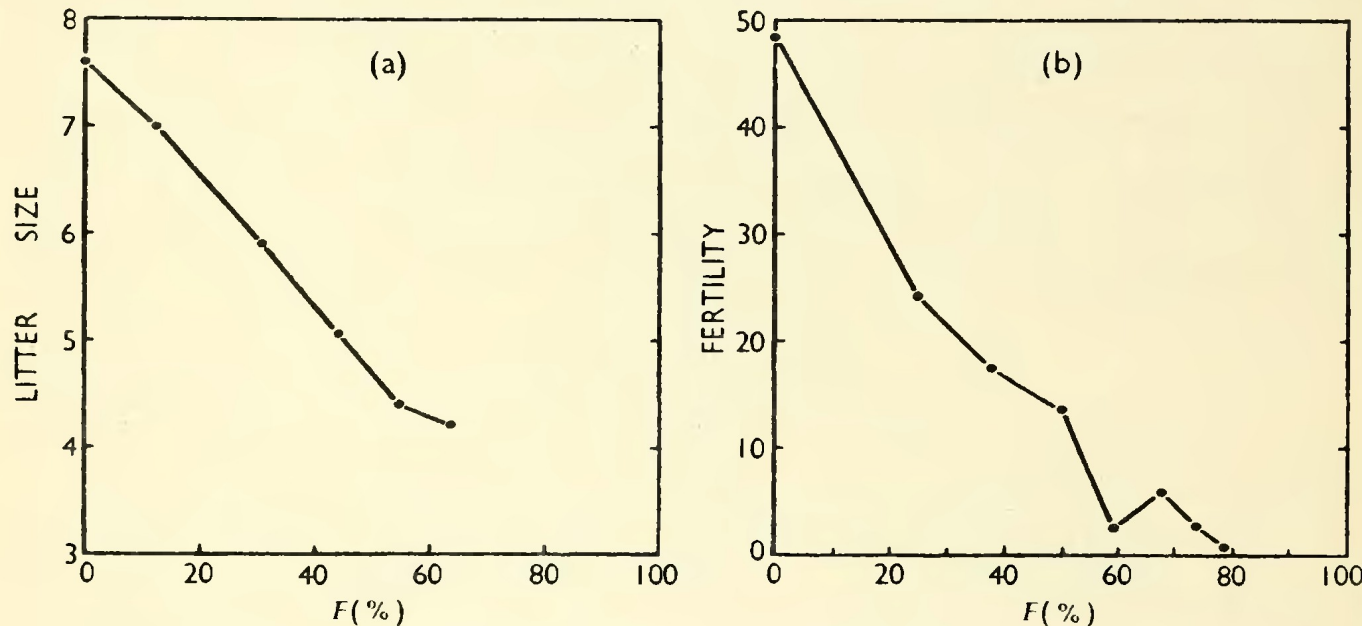


FIG. 14.1. Examples of inbreeding depression affecting fertility. (a) Litter-size in mice (original data). Mean number born alive in 1st litters, plotted against the coefficient of inbreeding of the litters. The first generation was by double-first-cousin mating; thereafter by full-sib mating. No selection was practised. (b) Fertility in *Drosophila subobscura*. Mean number of adult progeny per pair per day, plotted against the inbreeding coefficient of the parents. Consecutive full-sib matings. (Redrawn from Hollingsworth & Smith, 1955.)

Utilisation de marqueurs ADN

Mais en fait, la plupart des populations n'ont pas une généalogie :

- Ni les populations naturelles (sauvages)
- Ni les populations en captivité (certains animaux domestiques mais aussi en conservation)

Deux possibilités :

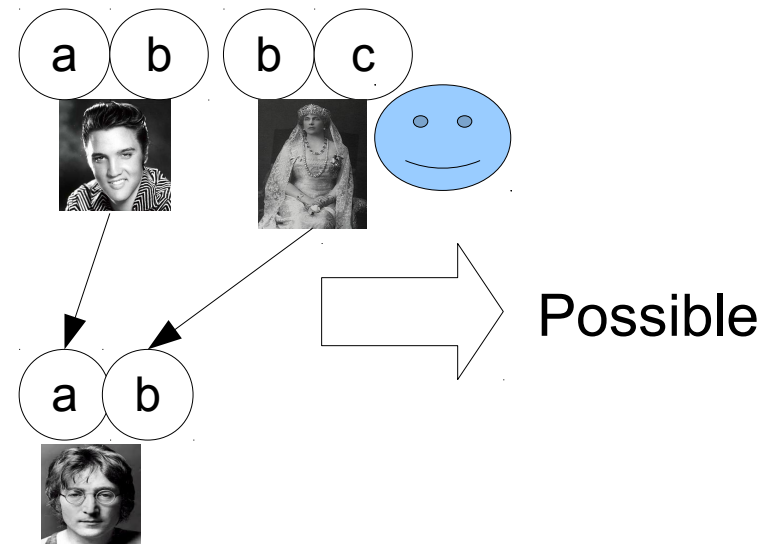
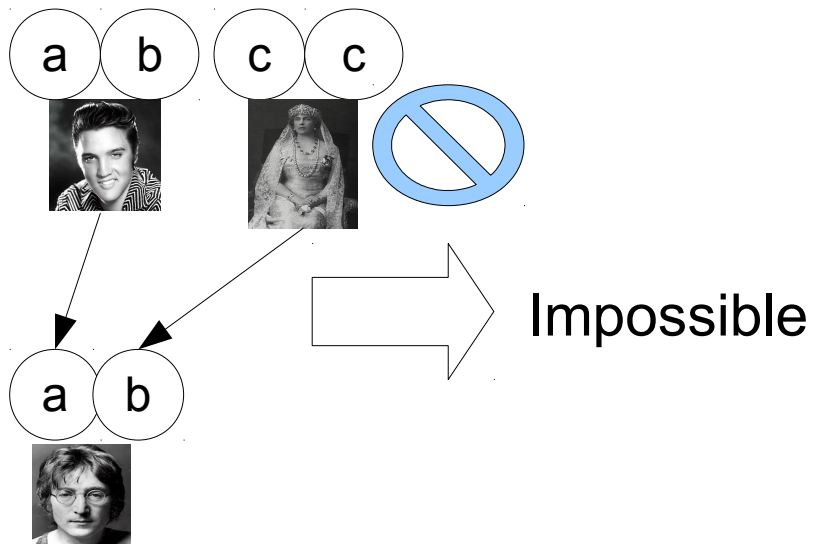
- On connaît une généalogie partielle ou incomplète → vérification de paternités
- On ne connaît rien :
 - estimation de la généalogie la plus probable ou
 - utilisation du parenté moléculaire

Solution : utilisation de marqueurs ADN

Utilisation de marqueurs ADN

Verification de paternités

On génotype l'individu et un (ou deux) de ses parents pour un marqueur (biallélique ou autre)



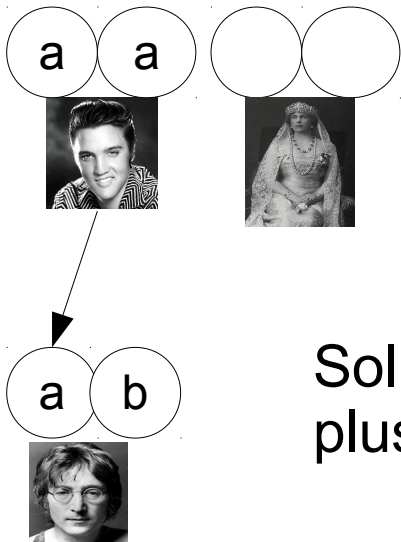
Mais il peut arriver que l'on se trompe par hasard !!

Utilisation de marqueurs ADN

Vérification de paternités

Cas simplifié : « b » vient de la mère
Il suffit que la « fausse » mère soit porteuse de l'allèle b pour que on ne l'exclue pas, avec probabilité :

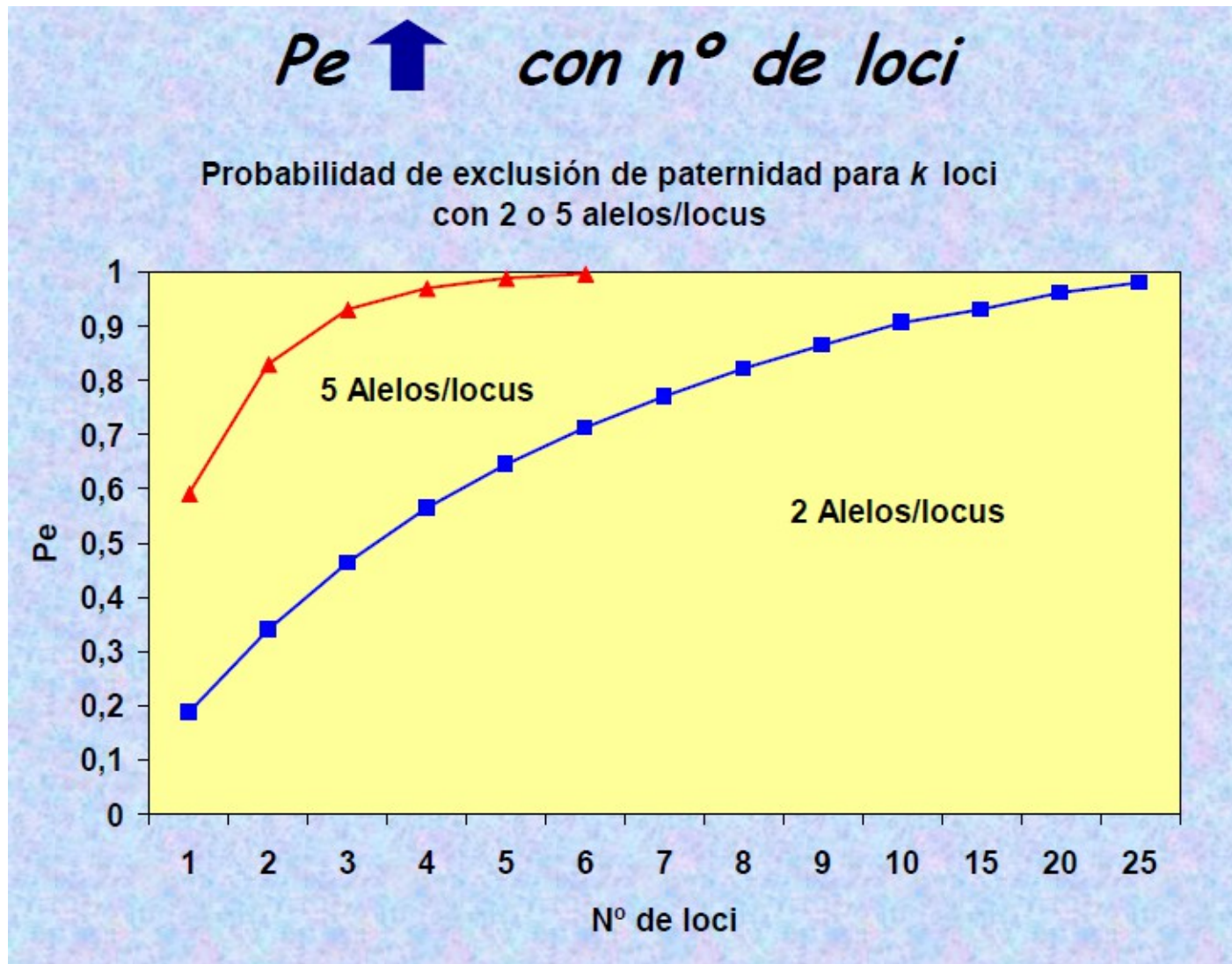
$$\Pr(\text{no exclusion} \mid \text{n'est pas mère}) = 2 * \text{freq}(b)$$



Solution : utiliser plusieurs marqueurs, avec (si possible) plusieurs allèles chacun

Utilisation de marqueurs ADN

D'après J Fernández & M Toro :



Utilisation de marqueurs ADN

Verification de paternités

En pratique, on utilise, soit un *kit* de microsatellites (~20) avec 4-10 allèles par microsatellites, soit un ensemble de >100 marqueurs SNP

Ces tests sont faits de manière systématique dans certains zoos, espèces de rente, et aussi quelques populations sauvages

Utilisation de marqueurs ADN

Assignation de parentés ou reconstruction de la généalogie

Beaucoup plus difficile

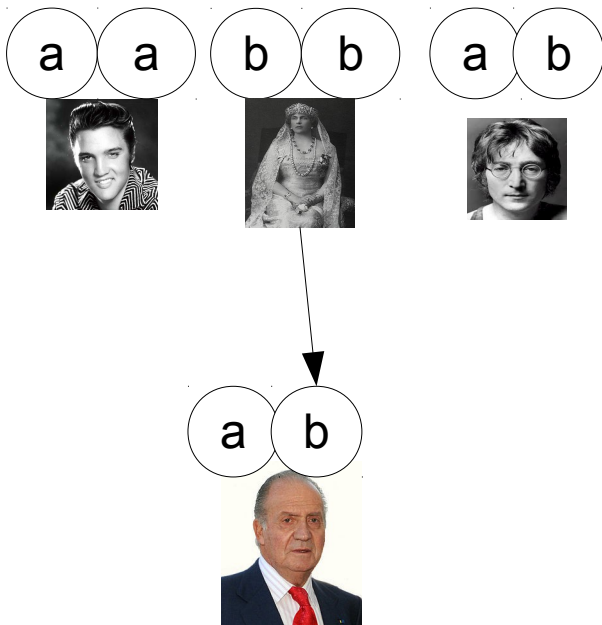
Sachant un ensemble d'individus et ses génotypes, il s'agit de reconstruire, soit les parentés « proches » (pere-file, plein germains, demi germains)

Cas assez difficile

Utilisation de marqueurs ADN

Assignation de parentés ou reconstruction de la généalogie

Qui est le papa le plus probable?



Elvis aurait certainement passé l'allèle « a »

Lennon aurait passé l'allèle « a » la moitié des fois

$\text{Pr}(\text{Elvis papa}) = 2/3$

$\text{Pr}(\text{Lennon papa}) = 1/3$

Utilisation de marqueurs ADN

Assignation de parentés ou reconstruction de la généalogie

Field Genetics - Cervus: Overview - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?


Field Genetics - Cervus: Overview

www.fieldgenetics.com/pages/aboutCervus_Overview.jsp

start [CTIGWiki] RENATER - service See... LAPACK: Main Page biblio Stack Overflow Les plus visités Débuter avec Firefox À la une Google

FIELD GENETICS
www.fieldgenetics.com

HOME FIELD GENETICS CERVUS DOWNLOAD



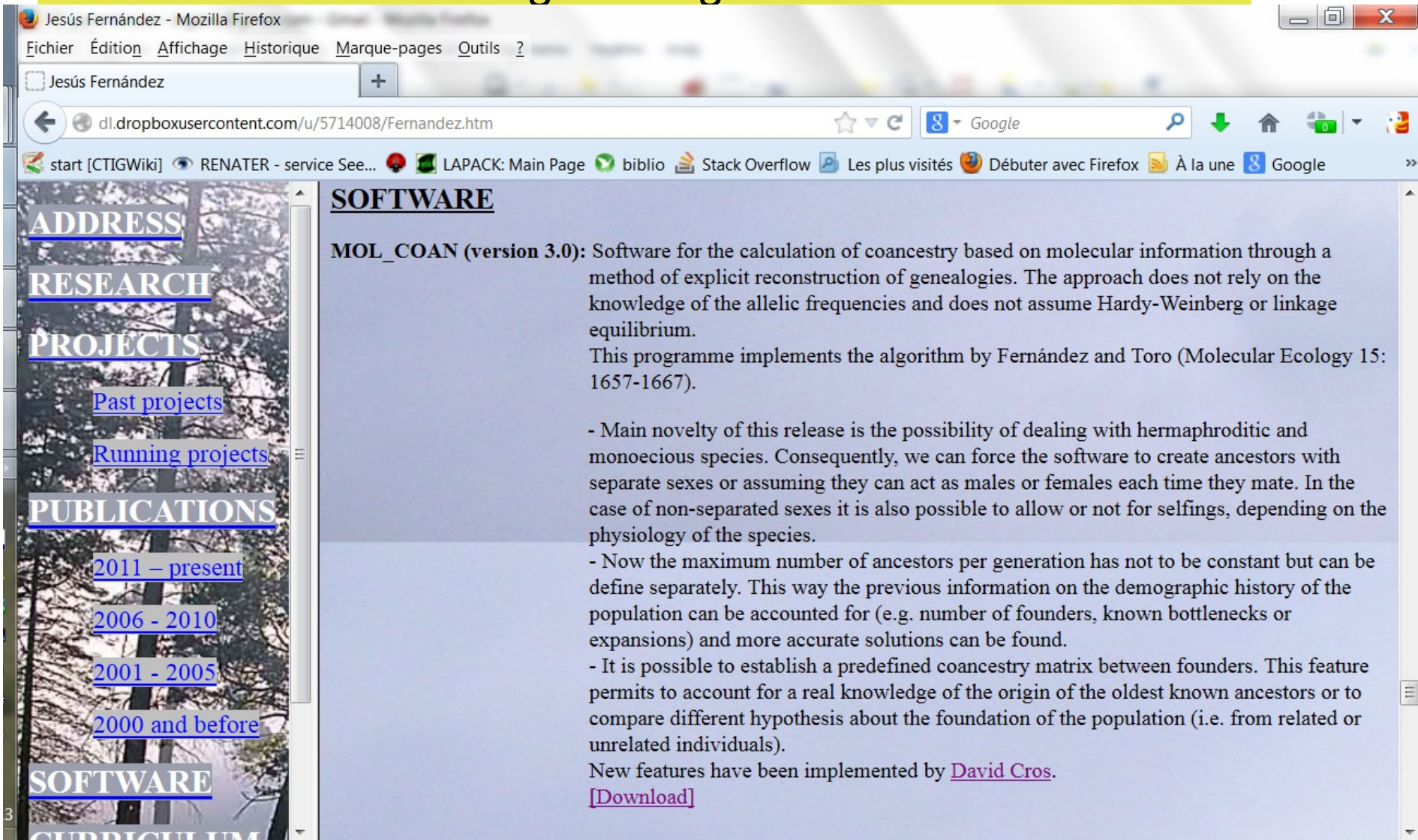
Cervus: Overview

Cervus	Cervus is a computer program for assignment of parents to their offspring using genetic markers.
Overview	
What Cervus Does	Cervus uses likelihood, a well-established statistical method for parentage analysis, to assign parentage. Cervus introduces two key enhancements to this process:
Using Cervus	1) Likelihood ratios are calculated allowing for the possibility that the genotypes of parents and offspring may be mistyped.
New Features	2) Cervus determines via simulation the level of confidence in the parentages it assigns.
Functions	
Methodology	
Papers	
Licensing	

Cervus represents an easy-to-use and practical tool for scientists seeking to establish parent-offspring relationships even when some genotypes are incomplete, incorrect or

Utilisation de marqueurs ADN

Assignation de parentés ou reconstruction de la généalogie



The screenshot shows a Mozilla Firefox browser window. The address bar displays the URL dl.dropboxusercontent.com/u/5714008/Fernandez.htm. The browser's menu bar includes 'Fichier', 'Édition', 'Affichage', 'Historique', 'Marque-pages', and 'Outils'. The sidebar on the left contains a tree view with the following items: 'ADDRESS', 'RESEARCH', 'PROJECTS', 'PUBLICATIONS', and 'SOFTWARE'. The 'PROJECTS' item is expanded, showing sub-items: 'Past projects', 'Running projects', '2011 - present', '2006 - 2010', '2001 - 2005', and '2000 and before'. The main content area is titled 'SOFTWARE' and contains the following text:

MOL_COAN (version 3.0): Software for the calculation of coancestry based on molecular information through a method of explicit reconstruction of genealogies. The approach does not rely on the knowledge of the allelic frequencies and does not assume Hardy-Weinberg or linkage equilibrium. This programme implements the algorithm by Fernández and Toro (Molecular Ecology 15: 1657-1667).

- Main novelty of this release is the possibility of dealing with hermaphroditic and monoecious species. Consequently, we can force the software to create ancestors with separate sexes or assuming they can act as males or females each time they mate. In the case of non-separated sexes it is also possible to allow or not for selfings, depending on the physiology of the species.
- Now the maximum number of ancestors per generation has not to be constant but can be define separately. This way the previous information on the demographic history of the population can be accounted for (e.g. number of founders, known bottlenecks or expansions) and more accurate solutions can be found.
- It is possible to establish a predefined coancestry matrix between founders. This feature permits to account for a real knowledge of the origin of the oldest known ancestors or to compare different hypothesis about the foundation of the population (i.e. from related or unrelated individuals).

New features have been implemented by [David Cros](#).
[\[Download\]](#)

Utilisation de marqueurs ADN

Utilisation du parenté moléculaire

On ne peut pas (facilement) reconstruire une généalogie pour 1000 individus avec 10,000 marqueurs...

De plus, on va se tromper (il y aura des bouts de la généalogie mal reconstruits)

Solution : utilisation des coefficients d'apparenté « moléculaires »

On n'essaye pas de reconstruire une généalogie pour calculer les coefficients. On calcule les coefficients avec les marqueurs sans passer par une généalogie.

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire »

Si on n'a pas de généalogies...
quoi faire ?

On a montré que le (coefficient de)
parenté est indicatif du coefficient
d'identité par état au gène(s) causal(aux)

Sont-elles apparentées ?



Il existent des gènes *non causaux* que
l'on peut voir (génotyper) : les
marqueurs moléculaires (SNPs,
microsatellites, et autres).



	sain	malade
sain	Oui	Non
malade	Non	Oui

Donc nous pouvons – en principe –
faire le chemin inverse IBS → IBD

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire »

Soit un indicateur de l'identité par état, le «coefficient de parenté moléculaire » ou « coefficient d'identité par état » Φ_{Mij}



	A	a
A	Oui	Non
a	Non	Oui

$$\Phi_{Mij} = 0,5$$

On peut le moyenner sur plusieurs loci...

	Individual i	Individual j	Φ_{Mij}
Locus 1	AA	AA	1
Locus 2	Bb	Bb	0,5
Locus 3	Cc	CC	0,5
Locus 4	dd	Dd	0,5
Locus 5	ee	ee	1
Locus /	ll	LL	0

$$\varphi_{Mij} = \frac{\sum_L \varphi_{l(i,j)}}{L}$$

$$\Phi_{Mij} = 0,583$$

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire »

Soit un indicateur de l'identité par état, le
«coefficient de parenté moléculaire » ou
« coefficient d'identité par état » Φ_{Mij}



	A	a
A	Oui	Non
a	Non	Oui

$$\Phi_{Mij} = 0,5$$

Mais, deux individus peuvent être identiques « par état » dans l'être « par descendance ».

Exemple : Elvis et moi sommes porteurs de l'allèle OO au groupe sanguin

Certes, nous sommes possiblement reliés, mais ça peut remonter au néolithique...

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire »

On peut définir aussi (seulement à partir de plusieurs loci) une « covariance moléculaire »

#fois/2 de l'allèle de référence (ex, « A »)

	Individual i	Individual j	dosage génique i	dosage génique j
Locus 1	AA	AA	1	1
Locus 2	Bb	Bb	0,5	0,5
Locus 3	Cc	CC	0,5	1
Locus 4	dd	Dd	0	0,5
Locus 5	ee	ee	0	0
Locus /	ll	LL	0	1

g_{il}

g_{jl}

$$Cov_{Mij} = Cov(g_{i,...}, g_{j,...}) = \frac{1}{L} \sum_L (g_{ik} - \bar{g}_i)(g_{jk} - \bar{g}_j)$$

$$Cov_{Mij} = 0,083$$

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs et la parenté par descendance

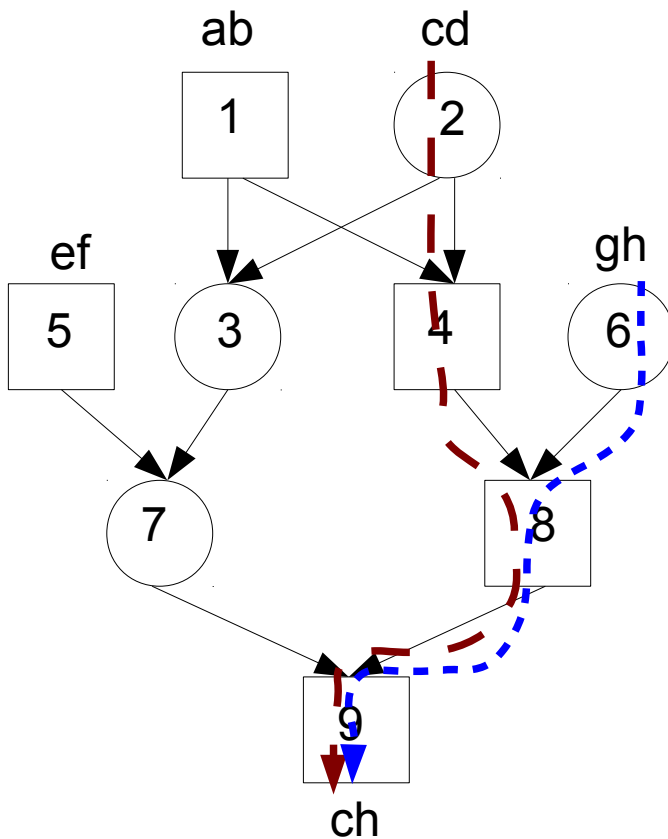
Supposons que il y a un marqueur pour lequel tous les allèles sont différents chez les fondateurs

Nous génotypons 9

Les allèles chez 9 sont identiques par état *que* s'ils sont identiques par descendance

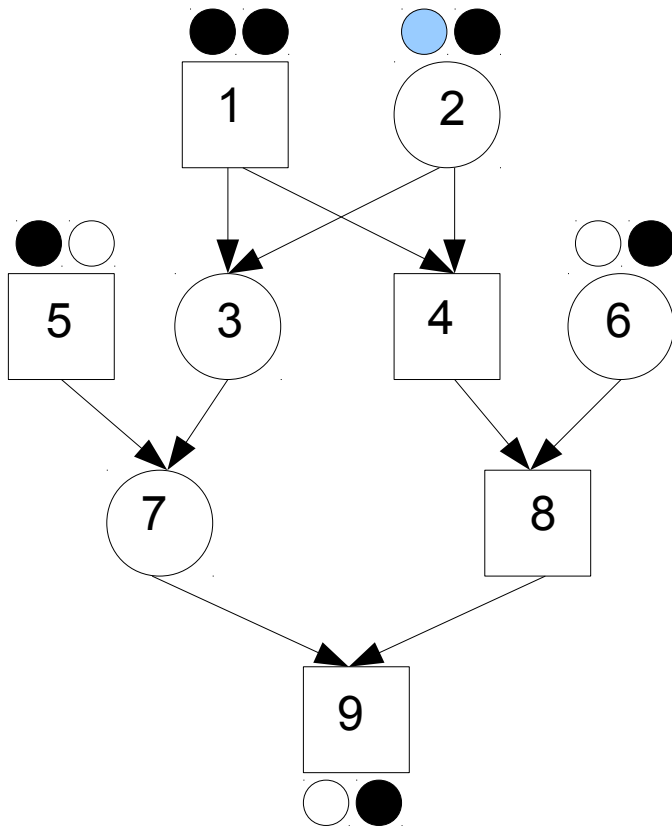
Cette identité par descendance « observé » est *plus exacte* que la probabilité d'identité par descendance (qui moyenne toutes les possibilités)

La probabilité par descendance est exacte comme moyenne d'un nombre infinie de loci.



La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs et la parenté par descendance



C'était bien s'il y avait des marqueurs complètement multialléliques

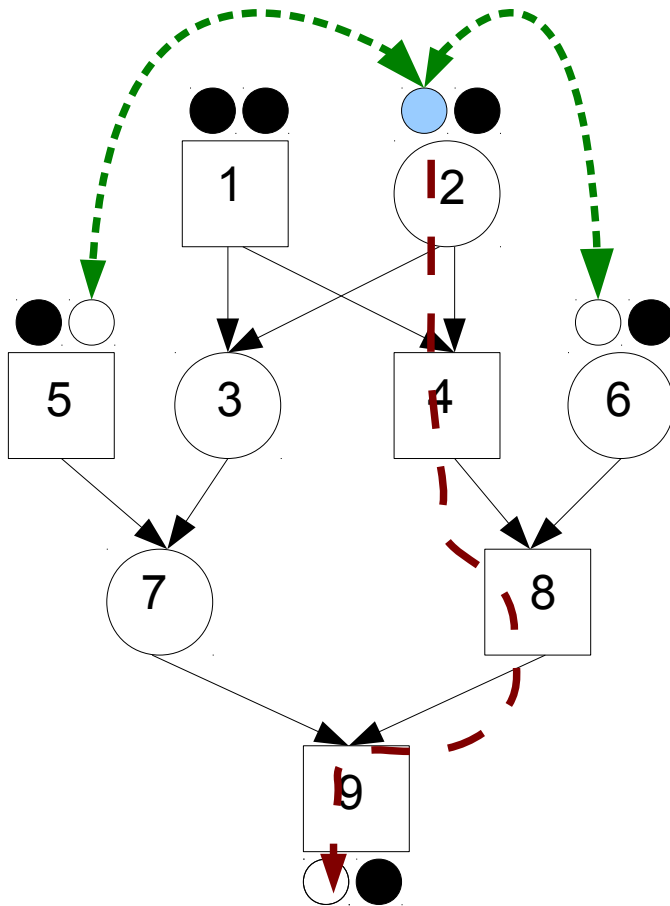
Comment faire pour marqueurs bialléliques ?

Supposons que l'on met au hasard des étiquettes (« A » ou « a ») aux allèles fondateurs avec une certaine fréquence p

On génotype à nouveau 9

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs et la parenté par descendance



On voit une identité par état que si :

a) on est identique **par descendance**

$$\varphi_{Mij} \approx \varphi_{i,j} + \underbrace{p^2 + q^2}$$

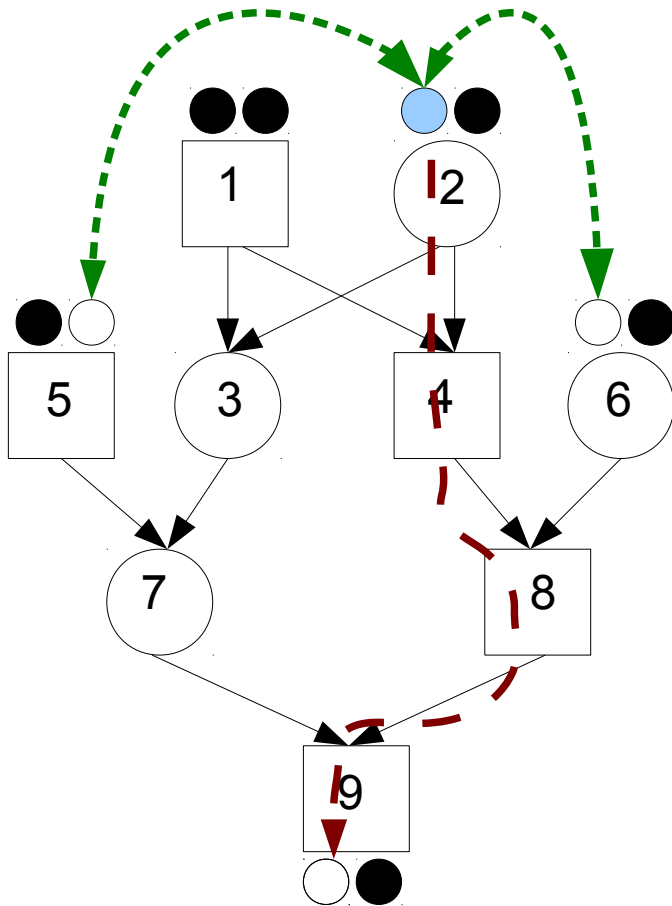
b) on n'est pas identique par descendance
mais **les fondateurs étaient identiques par état**

Donc on peut estimer le coefficient de parenté à partir du coefficient de parenté moléculaire

$$\hat{\varphi}_{i,j} = \frac{\varphi_{Mij} - p^2 - q^2}{2pq}$$

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs et la parenté par descendance



De même pour les covariances, l'on peut montrer que :

$$Cov_{Mij} = \varphi_{i,j} pq$$

Donc on peut estimer le coefficient de parenté à partir de la covariance moléculaire

$$\hat{\varphi}_{i,j} = \frac{Cov_{Mij}}{pq}$$

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs et la parenté par descendance

Étrangement, la formule pour le coefficient de parenté moléculaire se relie avec les formules de génétique de population :

$$(1 - \varphi_M) = (1 - \varphi) \left(1 - \sum p_i^2\right)$$

$$(1 - F_{IT}) = (1 - F_{IS}) (1 - F_{ST})$$

C'est un peu la même chose...

$$(1 - \varphi_M) = (1 - \varphi) \left(1 - \sum p_i^2\right)$$

Heterozigosité due à
identité par
descendance

Heterozigosité dans la
population de base

Heterozigosité
moléculaire

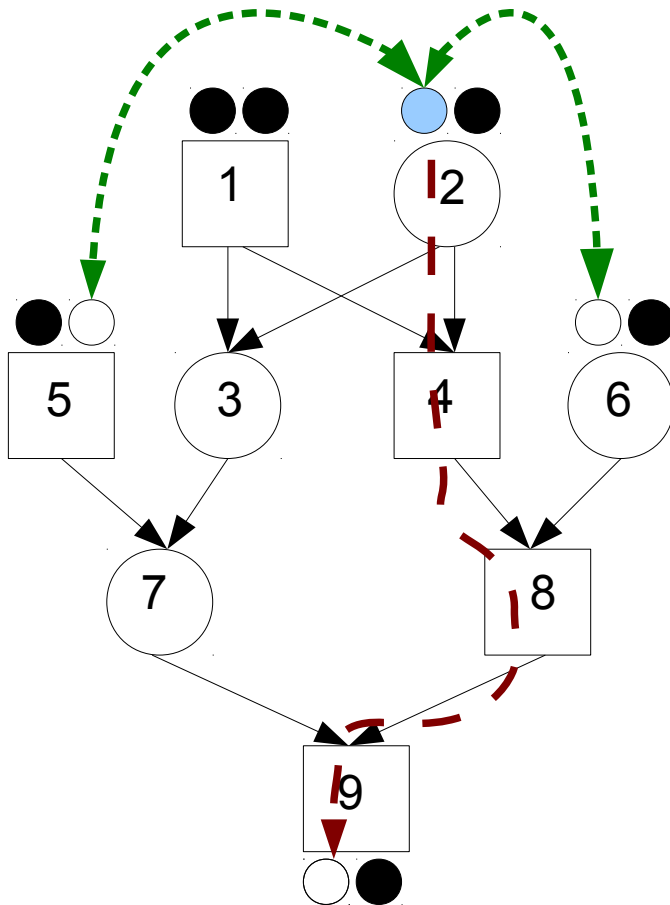
La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs et la parenté par descendance

$$(1 - \varphi_M) = (1 - \varphi) \left(1 - \sum p_i^2\right)$$
$$\varphi = \frac{\varphi_M - \sum p_i^2}{1 - \sum p_i^2} = \frac{\varphi_M - p^2 - q^2}{2pq}$$

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs et la parenté par descendance



Trois estimateurs typiquement utilisés

$$\hat{\varphi}_{i,j} = \frac{1}{L} \sum_L (\varphi_{Mij,l} - p_l^2 - q_l^2) / 2pq \rightarrow \text{Parenté moléculaire}$$

$$\hat{\varphi}_{i,j} = \frac{1}{L} \sum_L \frac{(g_{i,l} - p_l)(g_{k,l} - p_l)}{p_l q_l}$$

Covariance moléculaire

Pour que le calcul soit juste il faut beaucoup de marqueurs bialléliques ou plusieurs marqueurs multialléliques!

La ressemblance “moléculaire” entre apparentés

Le parenté « moléculaire » aux marqueurs

Problèmes typiques rencontrés

-
- Il est difficile de définir quelle est la population « de base »
- Souvent, on n'a pas accès aux fréquences alléliques de cette population ancienne
- On prends les fréquences de la population actuelle
- Avec le cours du temps et la dérive , $p \rightarrow 0$ ou 1
-
- Quel effet ?
-

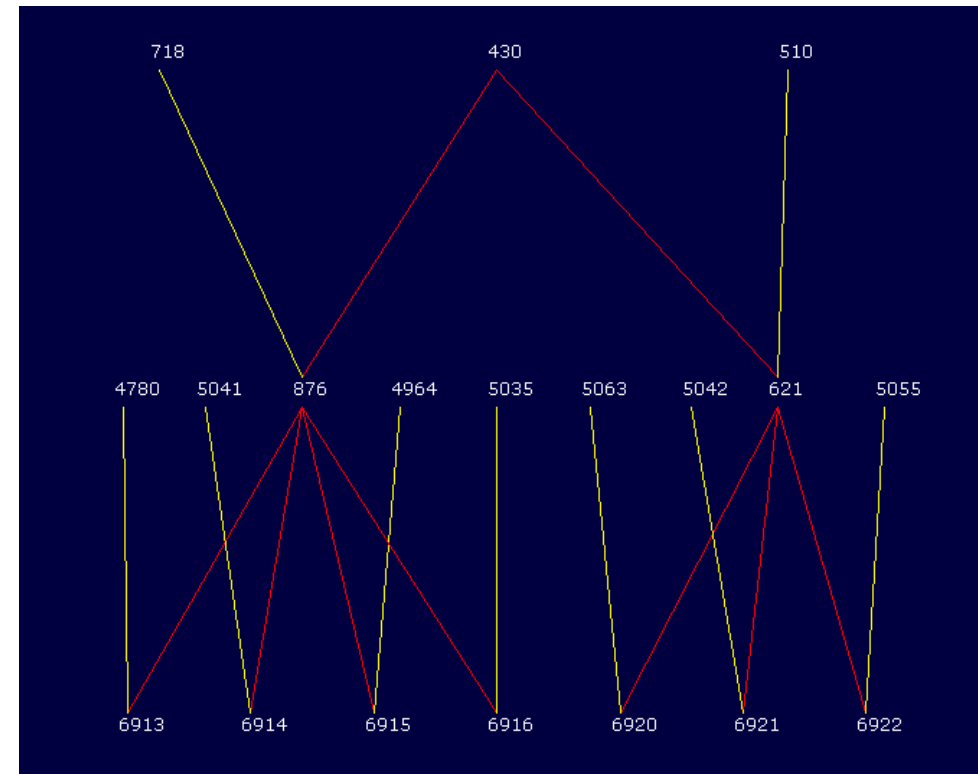
$$\hat{\varphi}_{i,j} = \frac{1}{L} \sum_L (\varphi_{Mij,l} - p_l^2 - q_l^2) / 2pq$$

On sous-estime les coefficients de parenté : coefficients estimés négatifs
Mais ils sont très corrélés aux « vrais » : utiles à tous les effets

La ressemblance “moléculaire” entre apparentés

Exemple réel

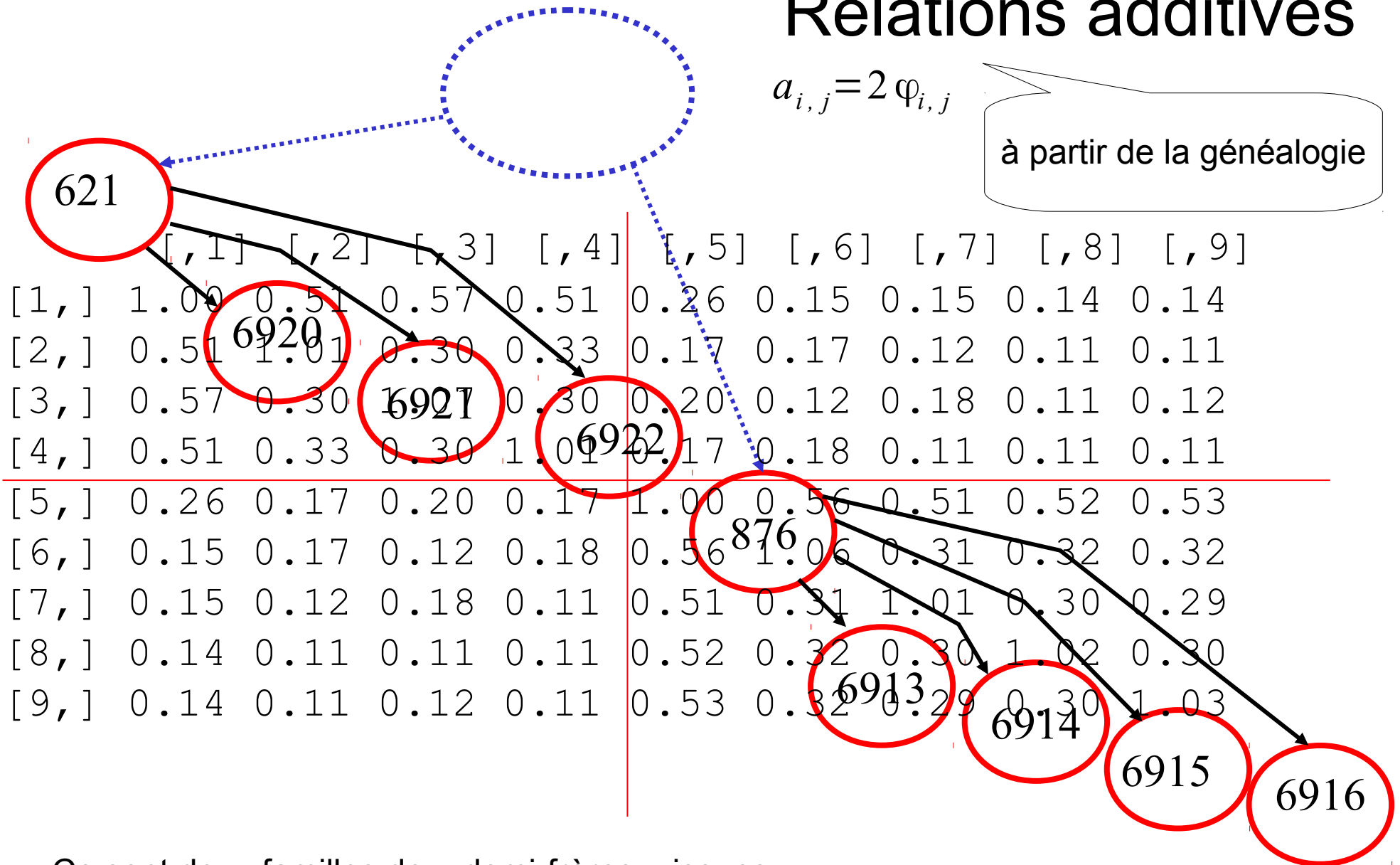
- 9 taureaux (race Montbéliarde) d'un ensemble de 1827, ~50000 SNPs
- Généalogie très complexe
Voici une généalogie simplifiée



Relations additives

$$a_{i,j} = 2 \varphi_{i,j}$$

à partir de la généalogie



Ce sont deux familles de « demi-frères » issues des deux taureaux qui sont des « demi-frères »

Relations additives

Légère consanguinité

à partir de la généalogie

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]
[1,]	1.00	0.51	0.57	0.51	0.26	0.15	0.15	0.14	0.14
[2,]	0.51	1.01	0.30	0.33	0.17	0.17	0.12	0.11	0.11
[3,]	0.57	0.30	1.07	0.30	0.20	0.12	0.18	0.11	0.12
[4,]	0.51	0.33	0.30	1.01	0.17	0.18	0.11	0.11	0.11
[5,]	0.26	0.17	0.20	0.17	1.00	0.56	0.51	0.52	0.53
[6,]	0.15	0.17	0.12	0.18	0.56	1.06	0.31	0.32	0.32
[7,]	0.15	0.12	0.18	0.11	0.51	0.31	1.01	0.30	0.29
[8,]	0.14	0.11	0.11	0.11	0.52	0.32	0.30	1.02	0.30
[9,]	0.14	0.11	0.12	0.11	0.53	0.32	0.29	0.30	1.03

0.125 comme attendu pour des demi-cousins

Relations additives

estimés avec des marqueurs $\hat{a}_{i,j} = 2 \hat{\phi}_{i,j}$

Coefficients autour de 1 dans la diagonal

Des relations négatives !

	[, 1]	[, 2]	[, 3]	[, 4]	[, 5]	[, 6]	[, 7]	[, 8]	[, 9]
[1,]	0.91	0.44	0.47	0.42	0.14	0.05	0.05	0.02	0.11
[2,]	0.44	1.01	0.20	0.27	0.02	0.06	-0.04	-0.04	0.04
[3,]	0.47	0.20	0.98	0.21	0.07	0.00	0.08	-0.02	0.05
[4,]	0.42	0.27	0.21	0.96	0.02	-0.01	-0.02	0.01	0.04
[5,]	0.14	0.02	0.07	0.02	0.81	0.37	0.33	0.35	0.39
[6,]	0.05	0.06	0.00	-0.01	0.37	0.94	0.16	0.15	0.20
[7,]	0.05	-0.04	0.08	-0.02	0.33	0.16	0.88	0.15	0.19
[8,]	0.02	-0.04	-0.02	0.01	0.35	0.15	0.15	0.88	0.18
[9,]	0.11	0.04	0.05	0.04	0.39	0.20	0.19	0.18	0.94

Les valeurs négatives indiquent que le valeur des p 's est mal estimé. Mais c'est pas méchant car on décale tout d'une constante et la formule des covariances est juste,

même si $a_{ij} < 0$

$$\text{cov}(\mathbf{G}_i, \mathbf{G}_j) = \text{cov}(\mathbf{A}_i, \mathbf{A}_j) + \text{cov}(\mathbf{D}_i, \mathbf{D}_j) = \mathbf{a}_{ij} \mathbf{V}_A + \mathbf{d}_{ij} \mathbf{V}_D$$

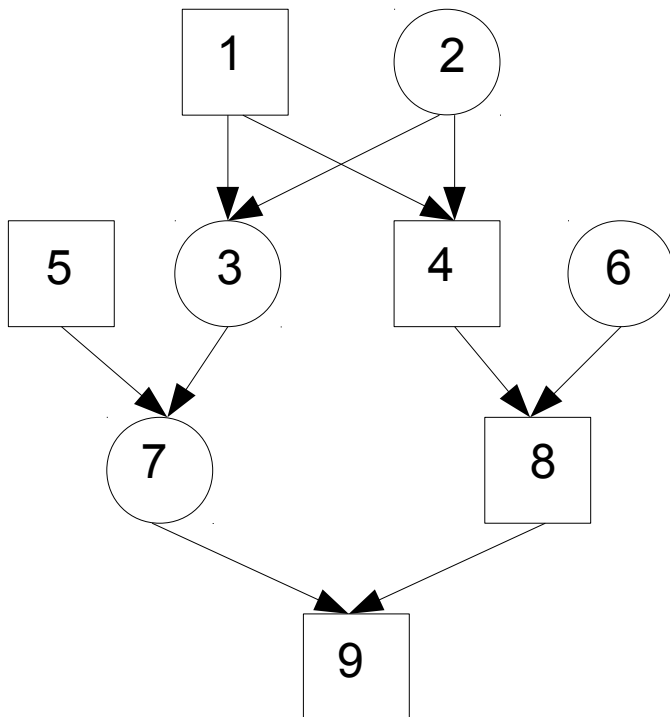
La ressemblance entre apparentés

Resumé

Les coefficients de parenté servent à quantifier la ressemblance *génétique* entre apparentés et donc sa ressemblance *phénotypique*

Ils sont bien définis comme probabilités d'être identiques, soit par état, soit par descendance,

On a des moyens efficaces de les calculer, tant à partir des généalogies comme à partir des marqueurs moléculaires



La ressemblance entre apparentés

Bibliographie

- Minvielle F : Principes d'amélioration génétique des animaux domestiques, INRA, 1990
- Malécot G: Les mathématiques de l'hérédité Paris: Masson; 1948
- Falconer D, Mackay T: Introduction to quantitative genetics New York: Longman; 1996. [Il existe une version en français]
-
- Cockerham C: Variance of gene frequencies. Evolution 1969, 23:72-84.
- Ritland K: Estimators for pairwise relatedness and individual inbreeding coefficients. Genet Res 1996, 67:175-185.
- Astle W, Balding D: Population structure and cryptic relatedness in genetic association studies. Stat Sci 2009, 24:451-471.