

Ressemblance entre apparentés et ressemblance moléculaire entre apparentés



Christèle Robert-Granié



avec la collaboration d'Andrés Legarra

La ressemblance entre apparentés

Les notions de **parenté** et de **consanguinité** sont anciennes et centrales en génétique des populations et en génétique quantitative.

Elles permettent :

- de décrire la constitution génétique d'une population
- de décrire avec précision les liaisons entre deux individus donnés et de quantifier la ressemblance qui peut exister pour leurs valeurs phénotypiques respectives pour un caractère quantitatif

Ces notions sont au cœur des méthodes d'évaluation génétique des reproducteurs ainsi que de celles visant à estimer la part de variation des caractères complexes qui est d'origine héréditaire.

Deux individus sont apparentés s'ils possèdent au moins un ancêtre commun

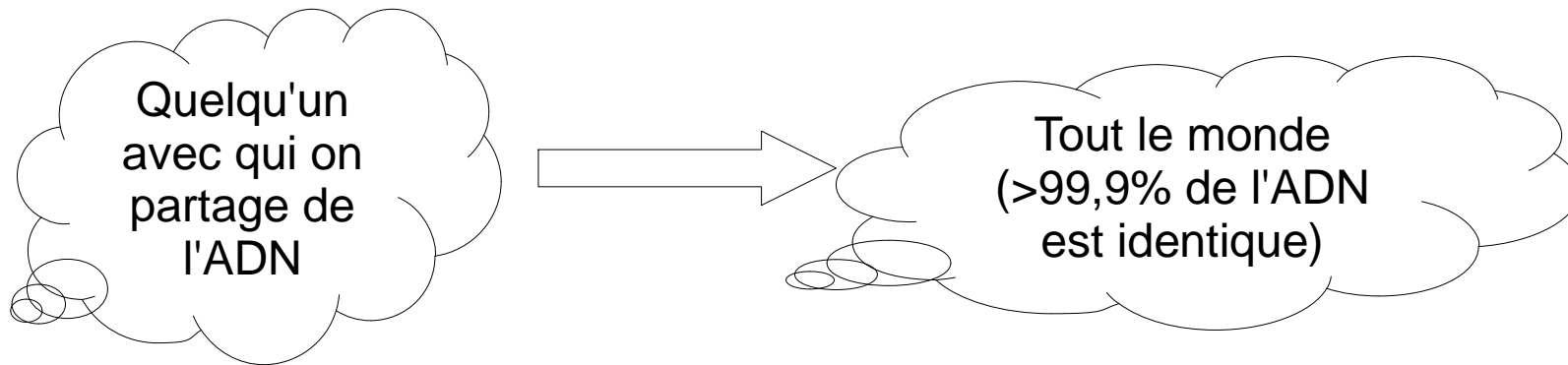
Un individu est consanguin si ses deux parents sont apparentés

La parenté concerne des paires d'individus. On est apparenté, ou non, à quelqu'un.

La consanguinité concerne des individus. On est consanguin ou on ne l'est.

La ressemblance entre apparentés

Qu'est ce qu'un apparenté?



Ce constat n'est pas très encourageant car nous voulons expliquer en quoi la génétique explique *la(les) différence(s)*

Pourquoi ?

comprendre les mécanismes évolutifs

pour sélectionner les meilleurs animaux (critère à définir)

pour prédire l'occurrence d'une maladie

La ressemblance entre apparentés

Qu'est ce qu'un apparenté?

Quelqu'un avec qui on partage de l'ADN

Tout le monde (>99,9% de l'ADN est identique)

Une personne avec qui on partage des allèles aux gènes qui expliquent la ressemblance phénotypique

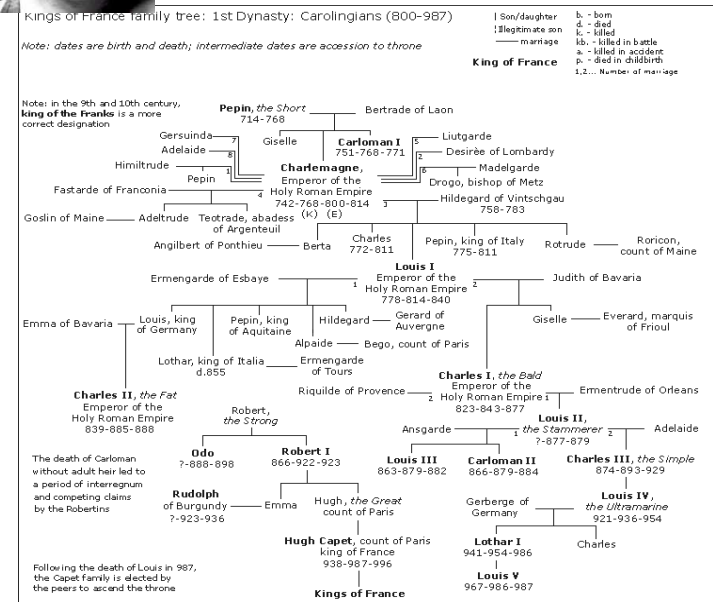
Partagent le type sanguin (O)

(1) Soit parce qu'on partage de vrais allèles aux gènes



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

- Mais les généalogies sont souvent incomplètes
- Et de toute manière, nous descendons tous du même individu(s)



La ressemblance entre apparentés

Clarification des notions intuitives

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

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

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

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

(3) Le coefficient de parenté

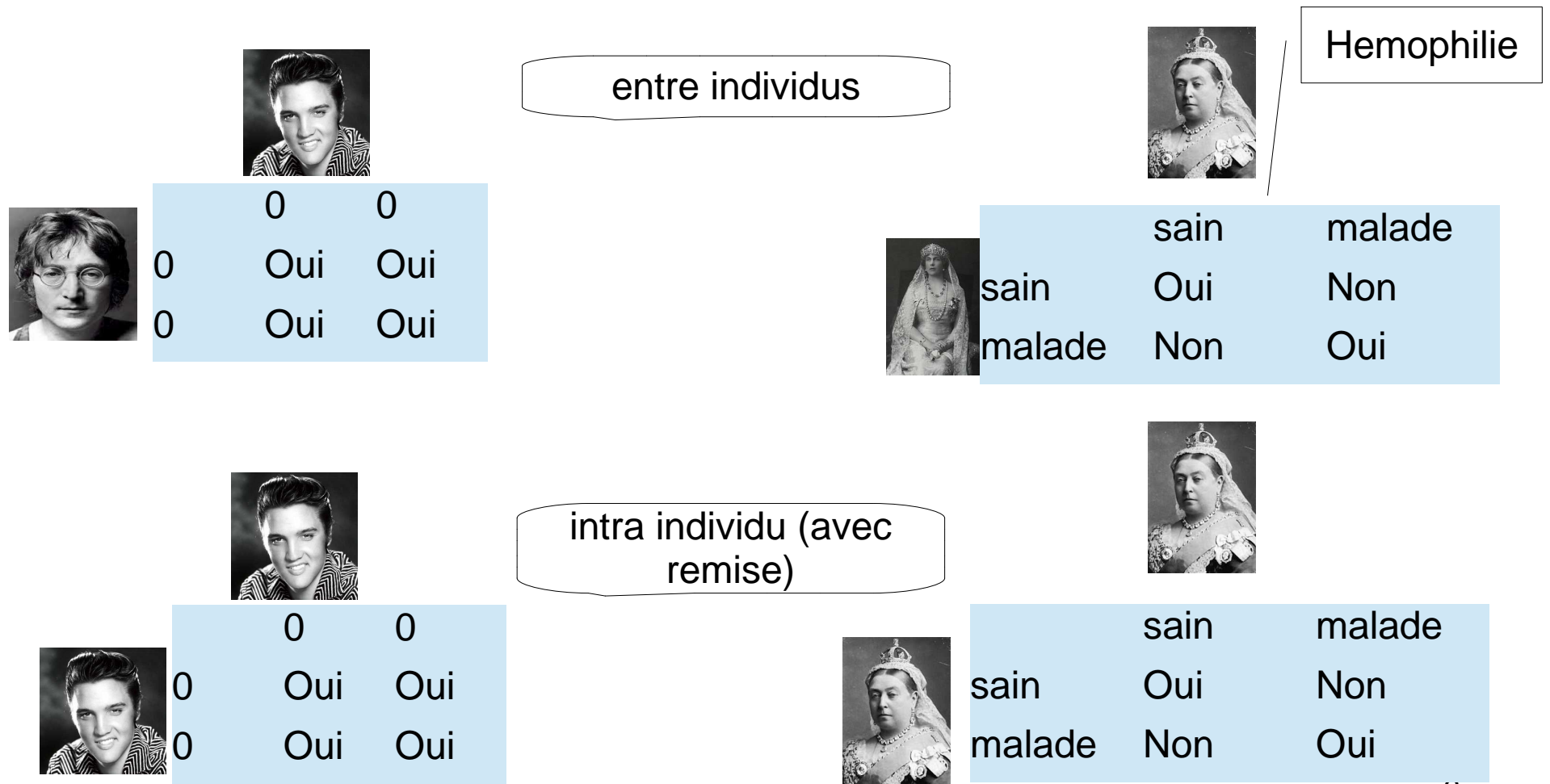
généralisation à un ensemble d'individus

(4) La relation génétique additive

sert à expliquer la ressemblance « phénotypique »

Identité par état

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

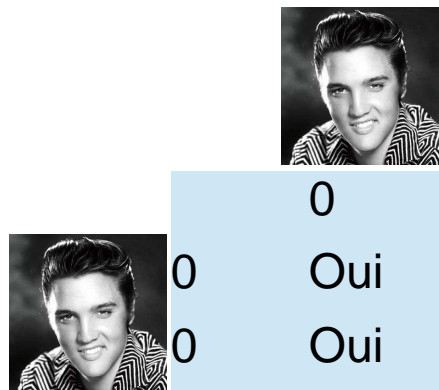


Identité par état

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 que ces 2 allèles soient identiques?



0 0
0 Oui Oui
0 Oui Oui

Proba=4/4=1



	sain	malade
sain	Oui	Non
malade	Non	Oui

Proba=2/4=0.5

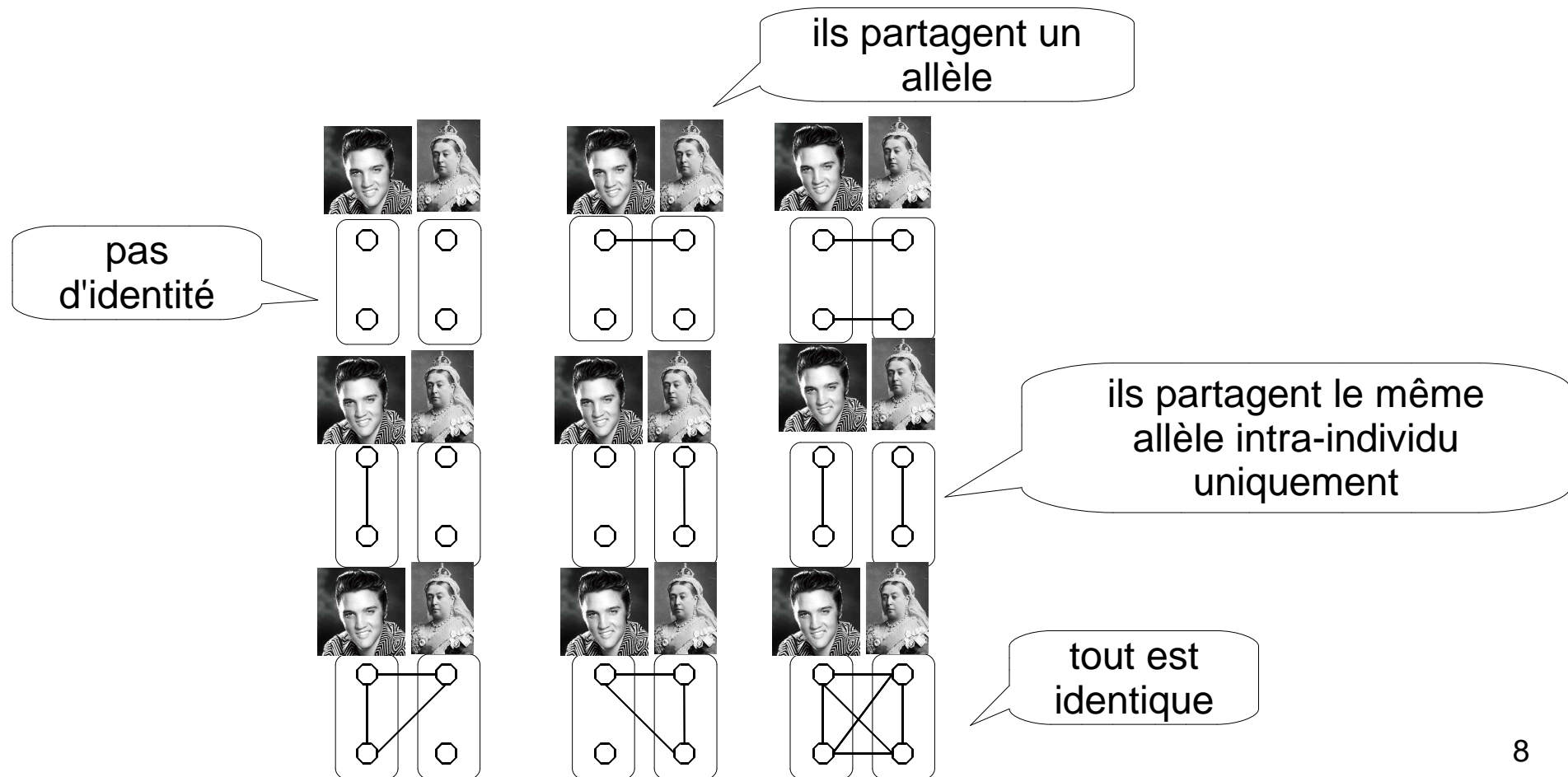
	A	B
C	Non	Non
B	Non	Oui

Proba=1/4=0.25

Identité par état

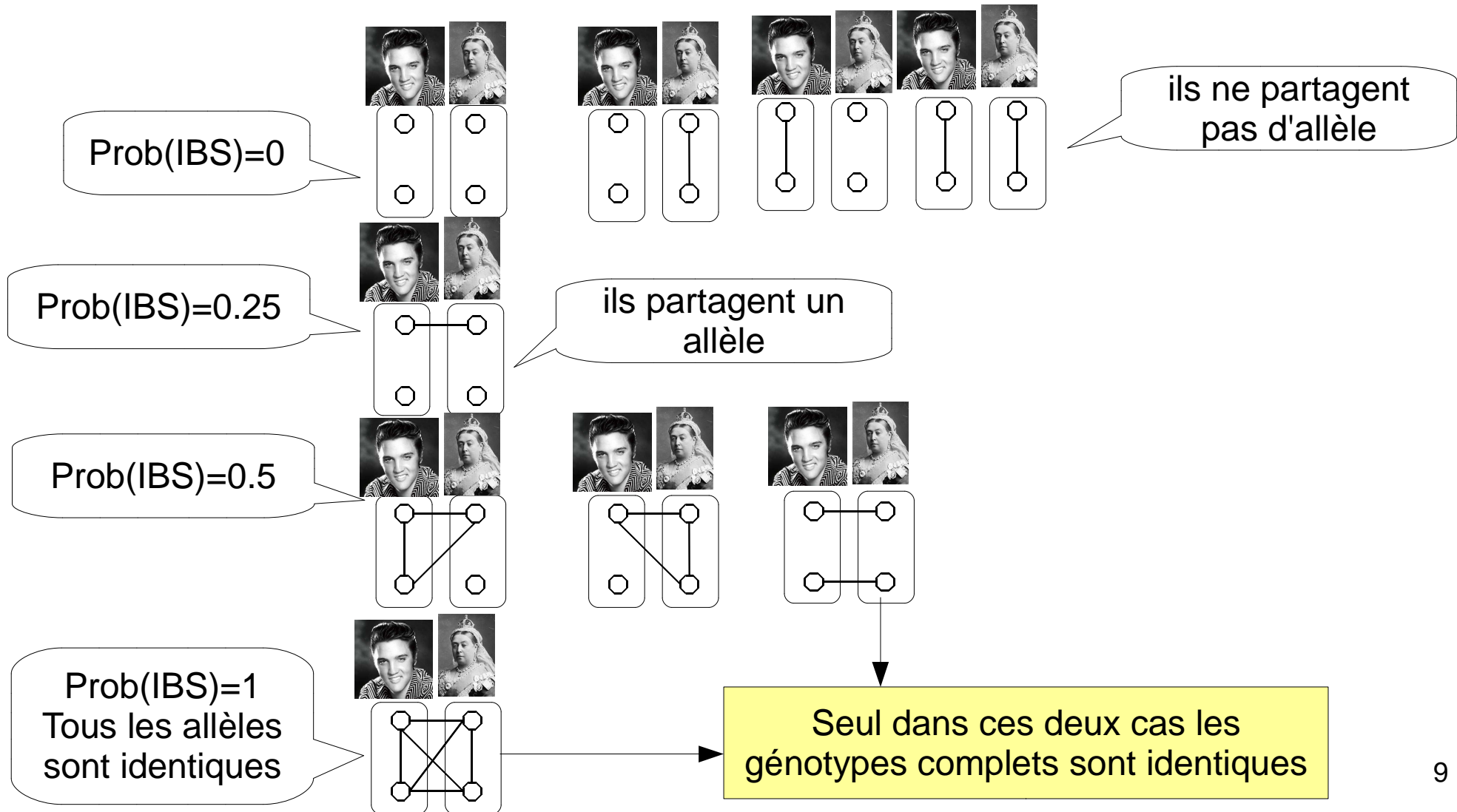
2- Etats d'identité

Quelles configurations d'identité peuvent prendre deux individus?



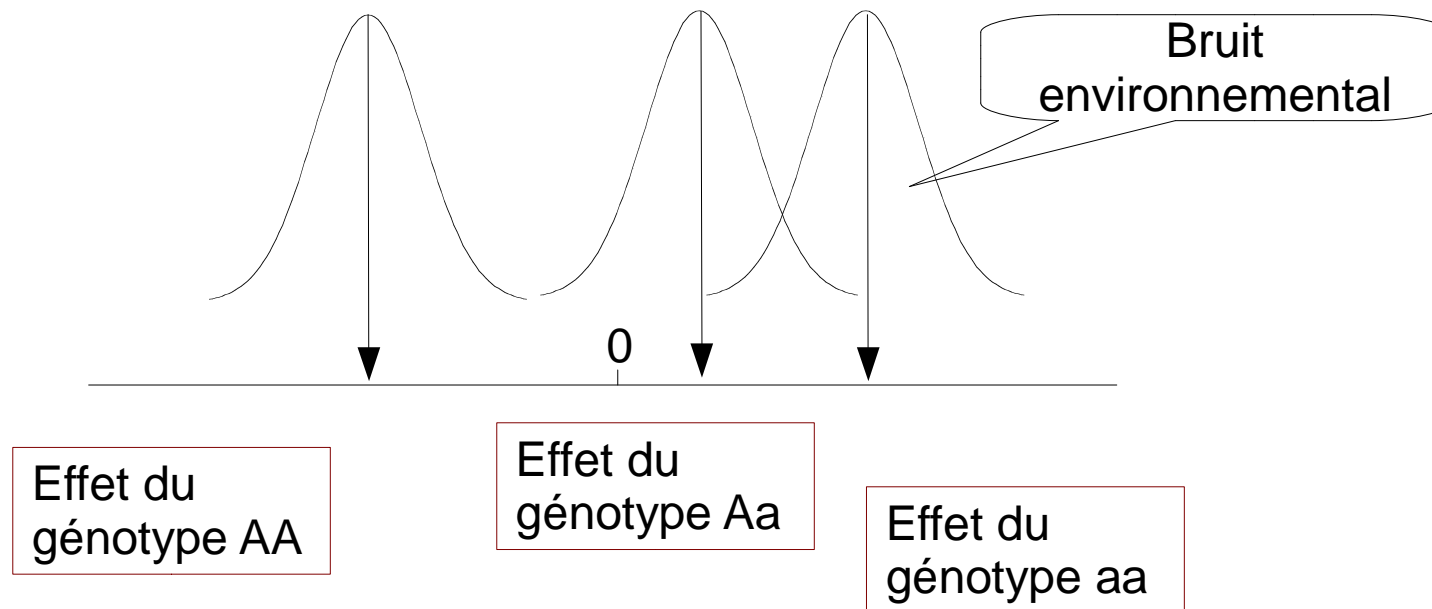
Identité par état

2- Etats d'identité



La ressemblance entre les phénotypes des apparentés

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



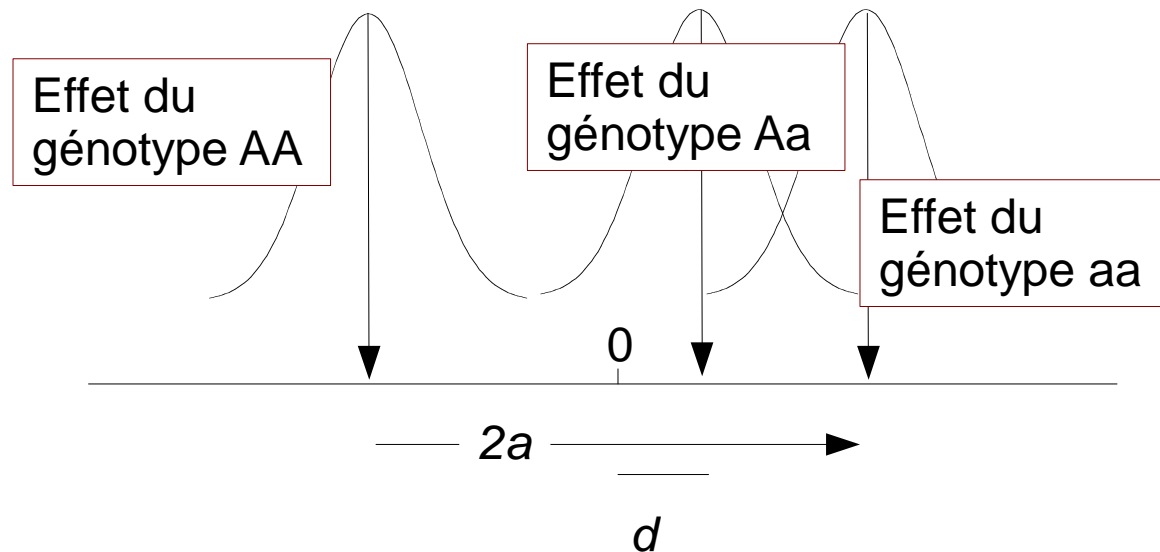
La ressemblance entre les phénotypes des apparentés

Parent-descendant

Génotype	A	D	G_{parent}	$G_{descendant}$
AA	$2\alpha_1 = 2q\alpha$	$\delta_{11} = -2q^2d$	$2q(\alpha - qd)$	$\alpha_1 = q\alpha$
Aa	$\alpha_1 + \alpha_2 = (q-p)\alpha$	$\delta_{12} = 2pqd$	$(q-p)\alpha + 2pqd$	$\frac{1}{2}\alpha_1 + \frac{1}{2}\alpha_2 = \frac{1}{2}(q-p)\alpha$
aa	$2\alpha_2 = -2p\alpha$	$\delta_{22} = -2p^2d$	$-2p(\alpha + pd)$	$\alpha_2 = -p\alpha$

$$\alpha = \alpha_1 - \alpha_2$$

La ressemblance entre les phénotypes des apparentés



On est surtout intéressé (et on peut regarder par moyennes statistiques) par l'effet *additif* qu'un génotype aura :

$$(2-2p)\alpha=2q\alpha$$

$$(1-2p)\alpha=(q-p)\alpha$$

$$-2p\alpha$$

Fonction de la quantité de copies de « A »

effet *dominant* : petit écart par rapport à l'effet sur le phénotype chez l'individu

$$-2q^2d$$

$$2pqd$$

$$-2p^2d$$

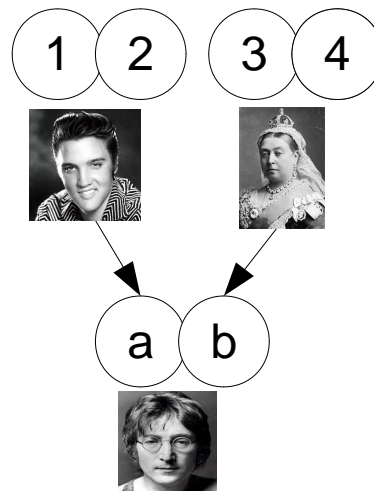
normalement (très) petit

La ressemblance entre les phénotypes des apparentés

Nous savons donc bien décrire les choses mais en pratique 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*



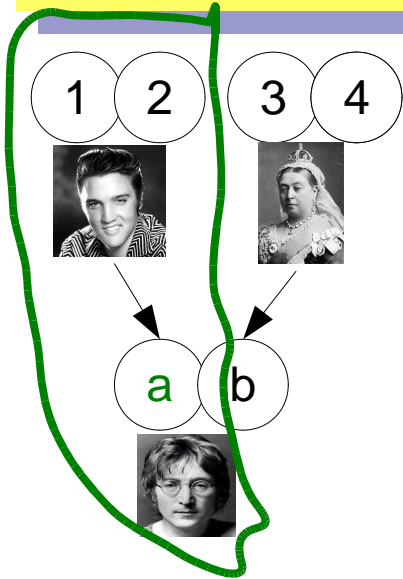
Supposons (c'est le plus fréquent) que les gènes sont transmis intacts (sans mutation)

Donc un allèle transmis est un allèle identique (la chaîne d'ADN n'est pas modifiée)

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} – 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

entre individu

	1	2	
a	Oui	Non	$\Phi_{ij}=2/8=0.25$
b	Non	Non	

	1	2	
a	Non	Oui	$\Phi_{ij}=2/8=0.25$
b	Non	Non	

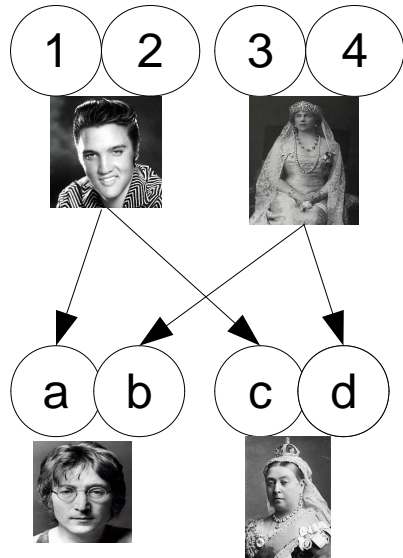
Cas père-fils

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

intra individu

	1	2	
1	Oui	Non	$\Phi_{ii}=2/4=0.5$
2	Non	Oui	

Probabilité d'identité par descendance



Événements possibles

		1	2
a		Oui	Non
	1	Non	Oui
	2		

		3	4
a		Non	Non
	1	Non	Non
	2		

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

		1	2
b		Non	Non
	3	Non	Non
	4		

		3	4
b		Oui	Non
	3	Non	Oui
	4		

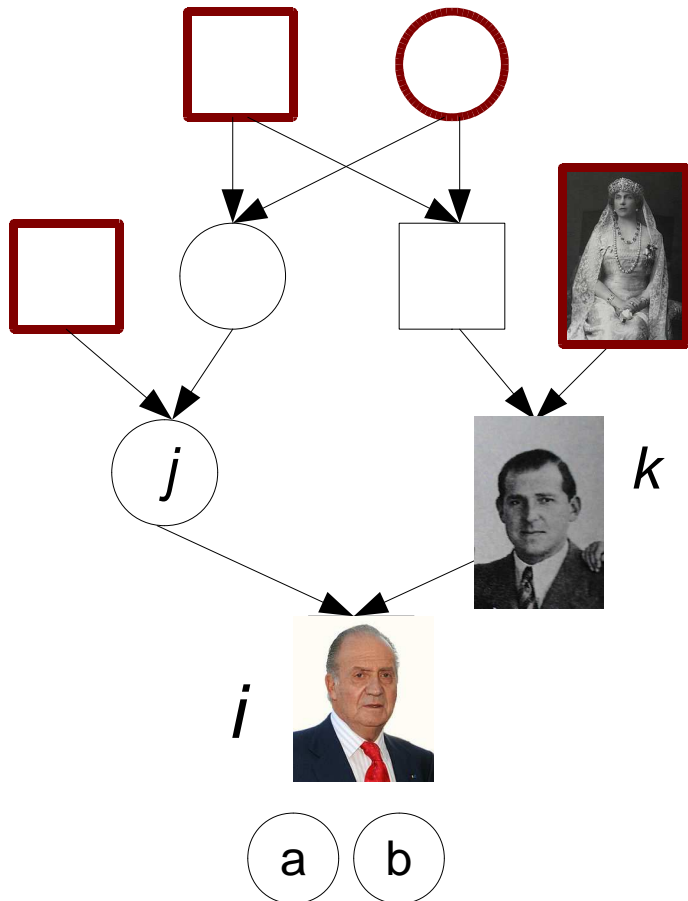
Cas plein frères

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

Probabilité d'identité par descendance

Et s'il y a de la consanguinité

Fondateurs



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

La différence avec le coefficient de parenté est que le tirage est fait avec remise (Φ_{ij}) ou pas (f_i). Et on a la relation suivante :

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

et aussi,

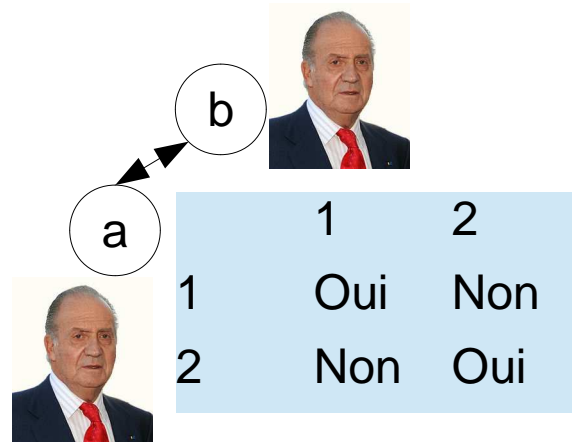
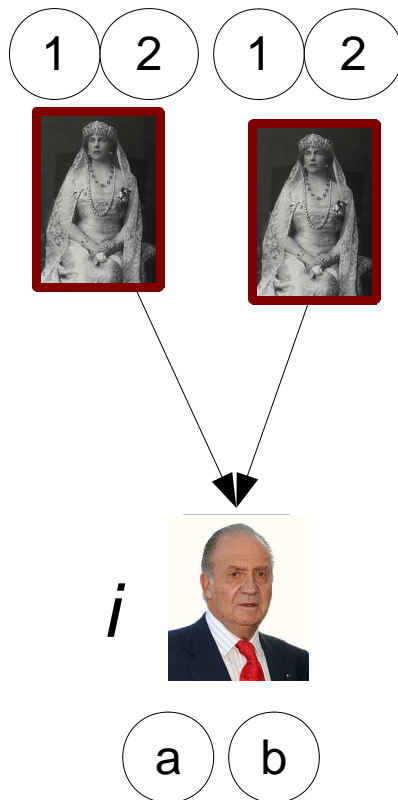
$$f_i=\Phi_{jk}$$

le coefficient de consanguinité d'un individu i est égal au coefficient de parenté entre ses 2 parents (j et k).

Probabilité d'identité par descendance

La consanguinité : cas simpliste

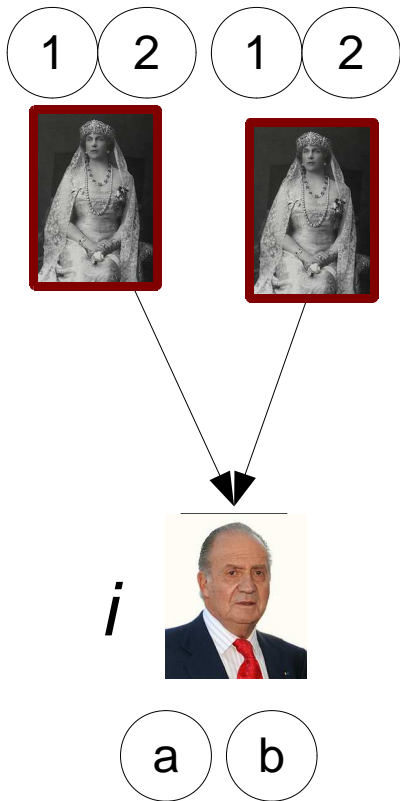
Fondateurs



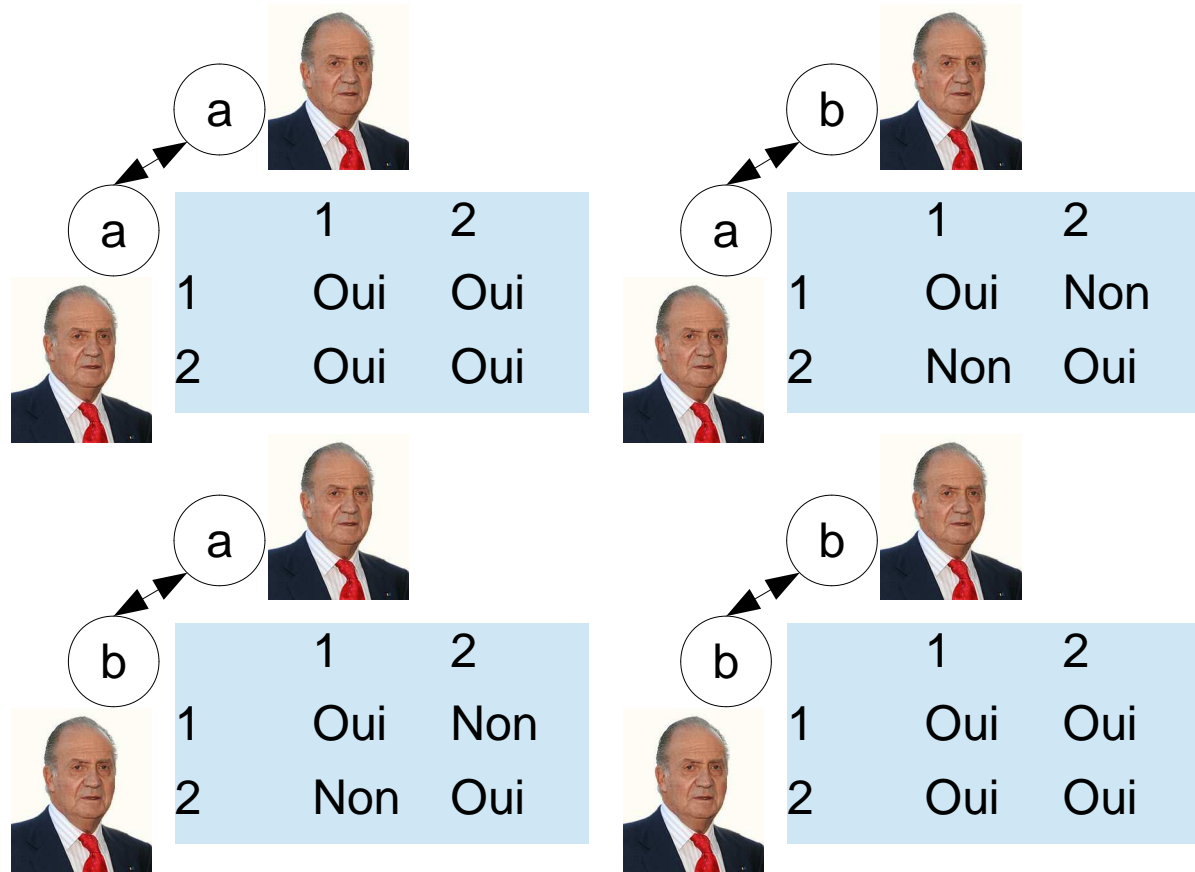
Pour cet individu,
 $f_i = \text{proba}(a==b)$
 $= 2/4 = 1/2 = 0.5$

Probabilité d'identité par descendance

Fondateurs



La consanguinité : cas simpliste



Pour cet individu,

$$\Phi_{ii} = \text{proba (2 allèles tirés au hasard chez } i \text{ avec remise sont identiques)} = 12/16 = 0.75 \quad 18$$

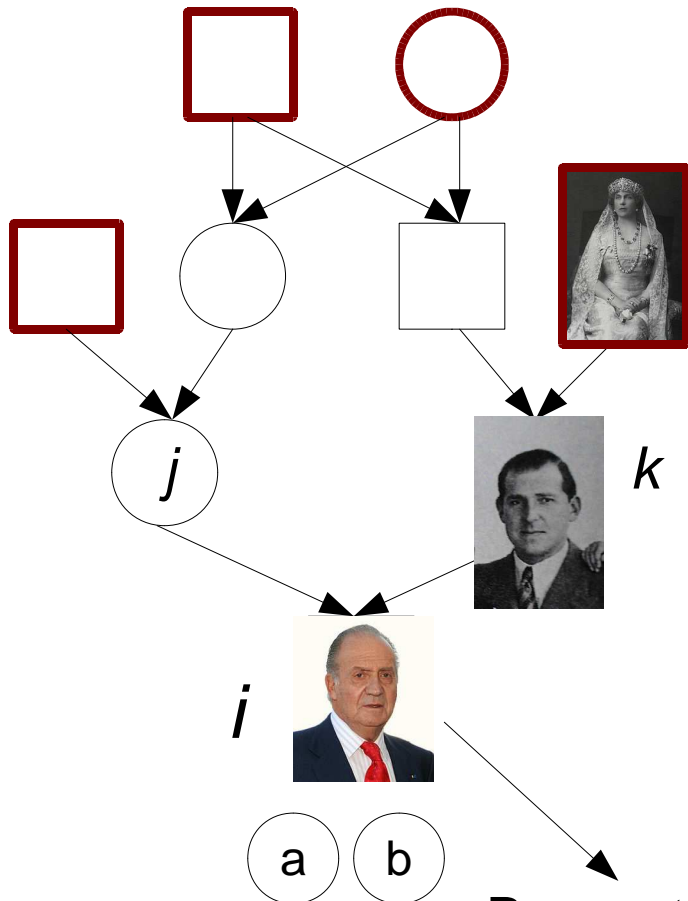
$$\text{On a bien } 1 + f_i = 2\Phi_{ii} \rightarrow 1 + 0.5 = 2 * 0.75$$

Probabilité d'identité par descendance

La consanguinité : cas un peu plus complexe

Fondateurs

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



La différence avec le coefficient de parenté est que le tirage est fait avec remise (Φ_{ij}) ou pas (f_j). Et on a la relation suivante :

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

et aussi,

$$f_i = \Phi_{jk}$$

le coefficient de consanguinité d'un individu i est égal au coefficient de parenté entre ses 2 parents (j et k).

Pour cet individu, $f_i = \Phi_{jk} = 1/16$ Cas de cousins germains

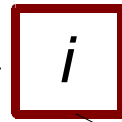
Tableau récapitulatif des quelques valeurs courantes de coefficients de parenté (sans consanguinité)

Lien de parenté	Probabilité	%
Parent-descendant	1/4	25
Grand-parent - petit(e)-fils(fille)	1/8	12,5
Oncle (tante) - neveu (nièce)	1/8	12,5
Vrais jumeaux	1/2	50
Pleins frères-soeurs	1/4	25
Demi frères-soeurs	1/8	12,5
Cousins germains	1/16	6,25

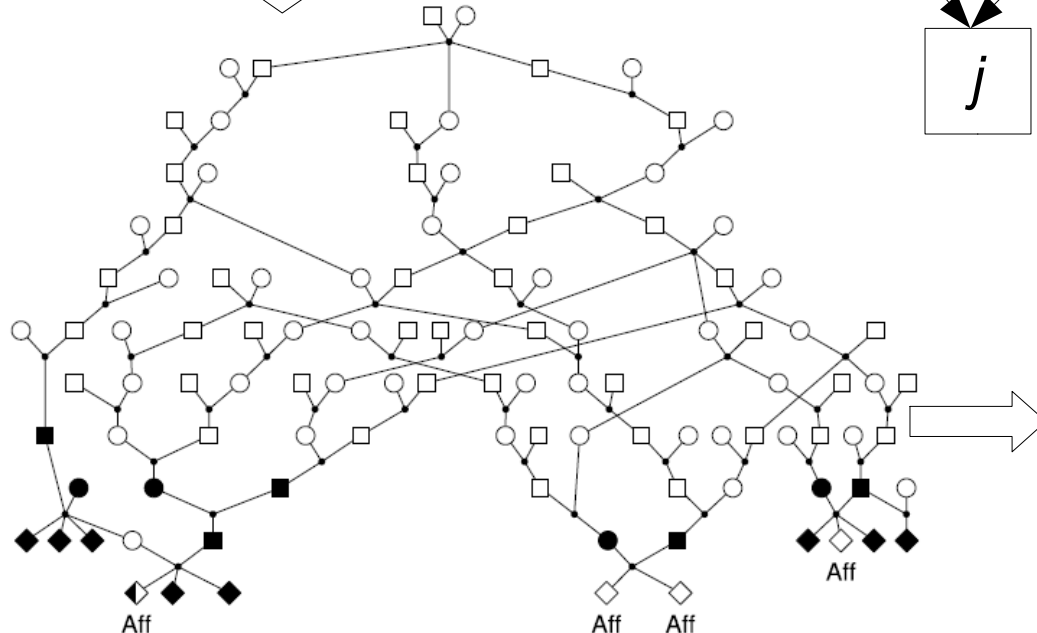
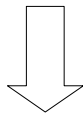
Probabilité d'identité par descendance

Le calcul

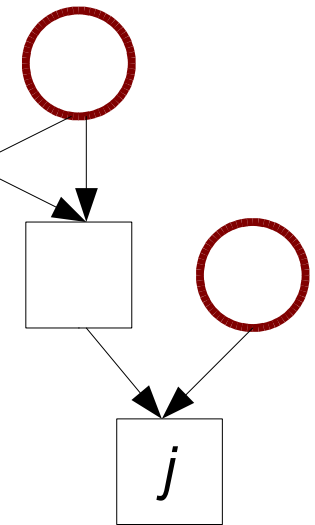
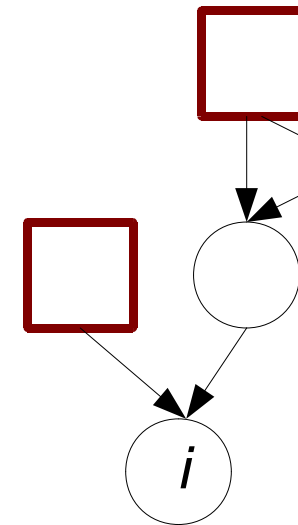
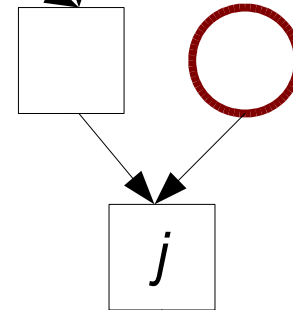
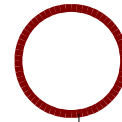
f et Φ peuvent se calculer à la main dans des cas simples...



Mais en général, il existe beaucoup de cas compliqués...



Fondateurs

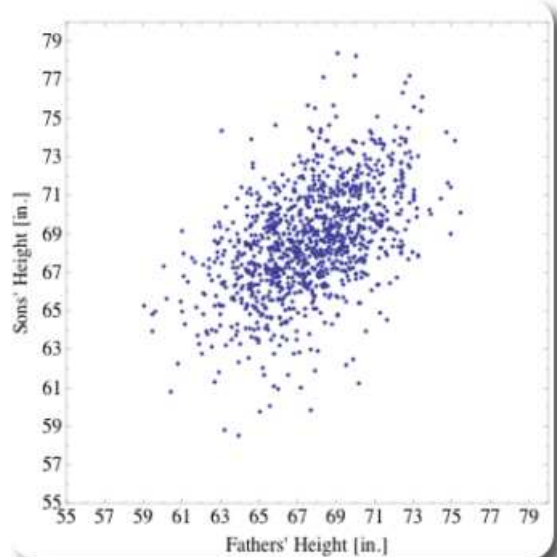
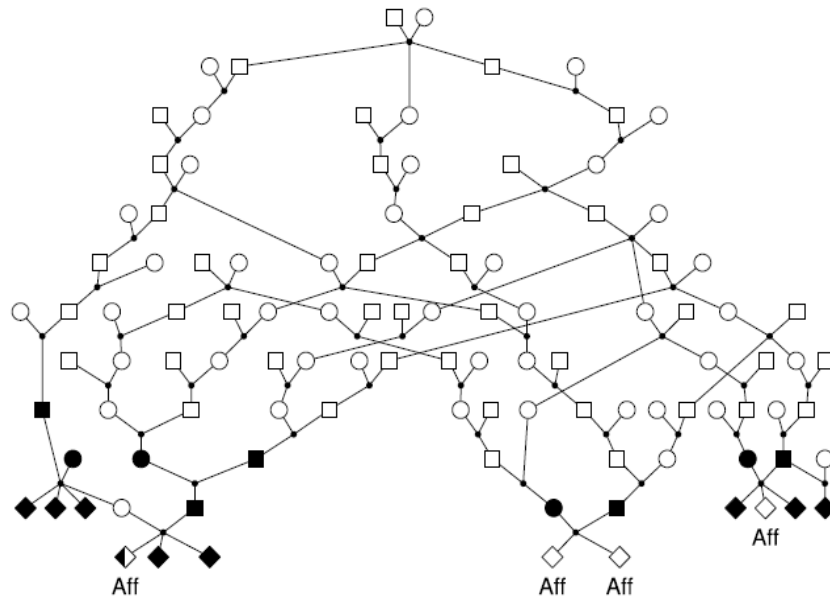


Heureusement il existe des méthodes qui permettent de calculer les f et Φ pour des généalogies très complexes

cf. exemple en TP

Probabilité d'identité par descendance

Utilisation



Quelle est l'utilité de tous ces coefficients ?

Le coefficient de consanguinité f_i sert :

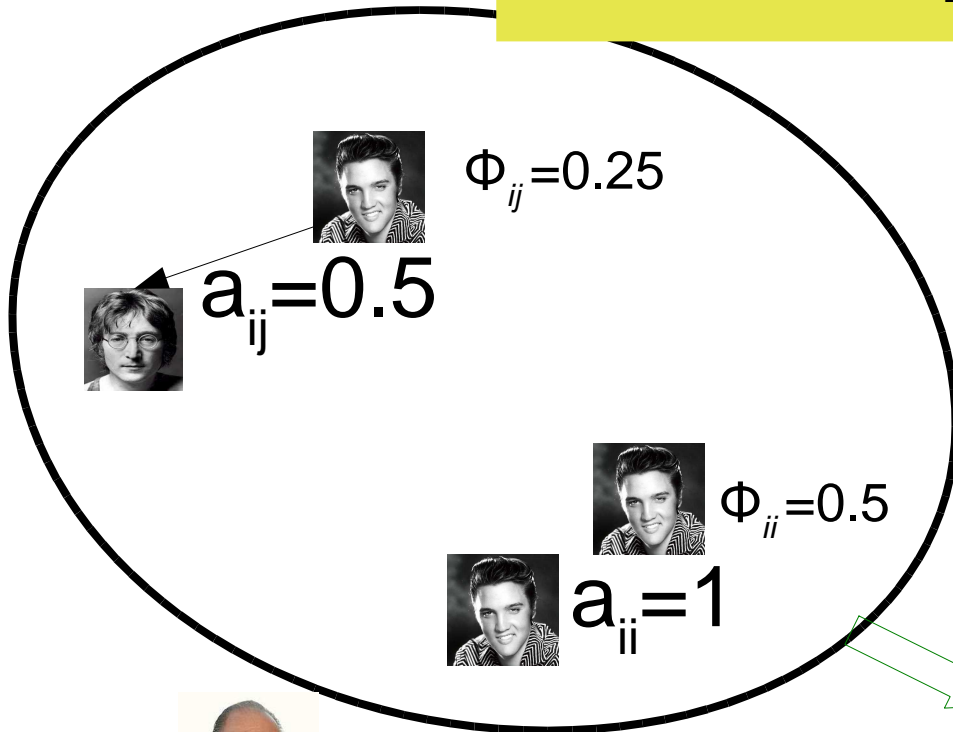
- à établir des risques d'être atteint des maladies récessives
- à gérer des populations rares (espèces protégées, Zoos)
- à gérer des populations pour éviter trop de consanguinité

Le coefficient de parenté Φ_{ij} sert :

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

Probabilité d'identité par descendance

La relation génétique additive



La relation additive est tout simplement

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

Et donc

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

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



$$a_{ii} = 1.0625$$

$$= 1 + f_i$$

$$= 1 + 1/16$$



$$a_{ij} = 0.5625$$

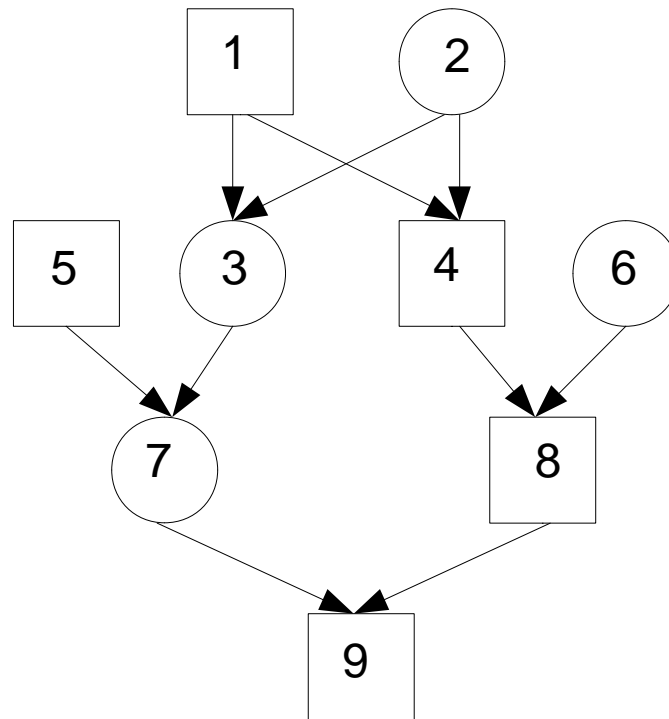
Père-fils avec consanguinité

Probabilité d'identité par descendance

Méthode tabulaire (Emik and Terrill 1949)

En pratique, on construit un fichier de pedigree avec 3 colonnes : individu, père, mère

1 0 0
2 0 0
3 1 2
4 1 2
5 0 0
6 0 0
7 5 3
8 4 6
9 8 7



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 file
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

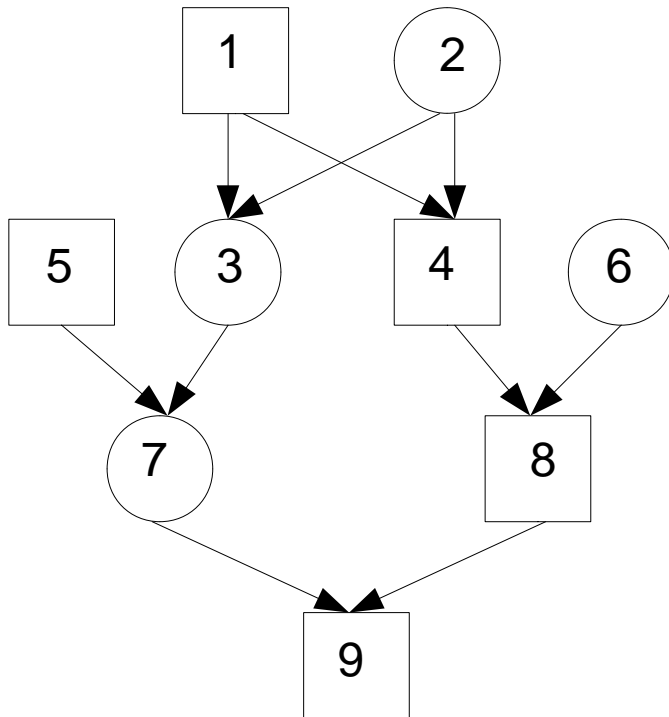
La matrice de parenté ou matrice de relations additives

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

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

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

Avec cette généalogie, reconstruire le tableau suivant

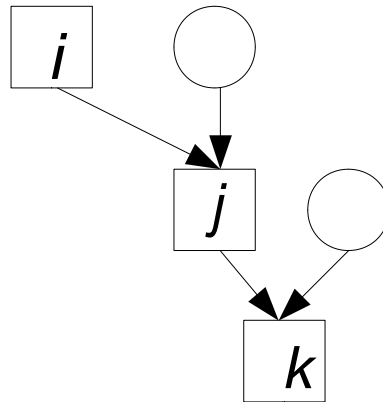


1	0	0,5	0,5	0	0	0,25	0,25	0,25
0	1	0,5	0,5	0	0	0,25	0,25	0,25
0,5	0,5	1	0,5	0	0	0,5	0,25	0,375
0,5	0,5	0,5	1	0	0	0,25	0,5	0,375
0	0	0	0	1	0	0,5	0	0,25
0	0	0	0	0	1	0	0,5	0,25
0,25	0,25	0,5	0,25	0,5	0	1	0,125	0,5625
0,25	0,25	0,25	0,5	0	0,5	0,125	1	0,5625
0,25	0,25	0,375	0,375	0,25	0,25	0,5625	0,5625	1,0625

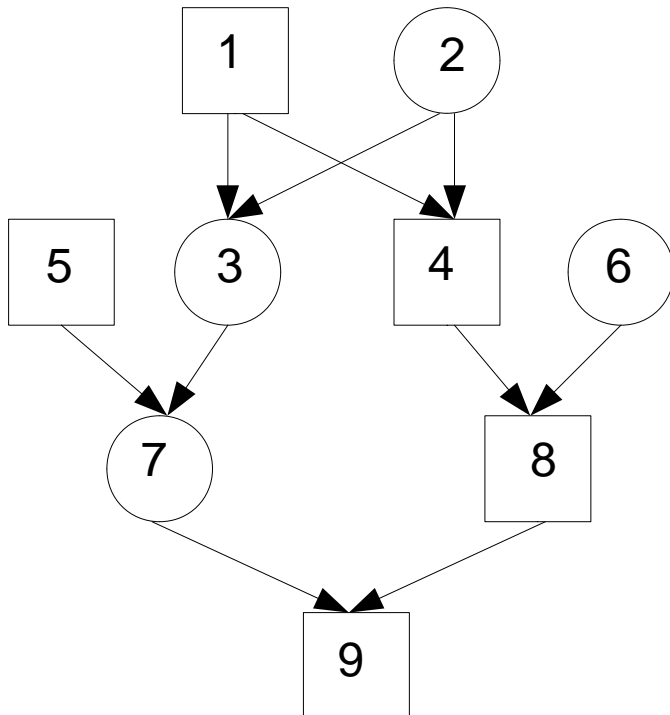
Probabilité d'identité par descendance

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



	10	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
0,5	0,5	1	0,5	0	0	0,5	0,25	0,375
0,5	0,5	0,5	1	0	0	0,25	0,5	0,375
0	0	0	0	1	0	0,5	0	0,25
0	0	0	0	0	1	0	0,5	0,25
	0,25	0,25	0,5	0,25	0,5	0	1	0,125
	0,25	0,25	0,25	0,5	0	0,5	0,125	1
	0,25	0,25	0,375	0,375	0,25	0,25	0,5625	0,5625

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

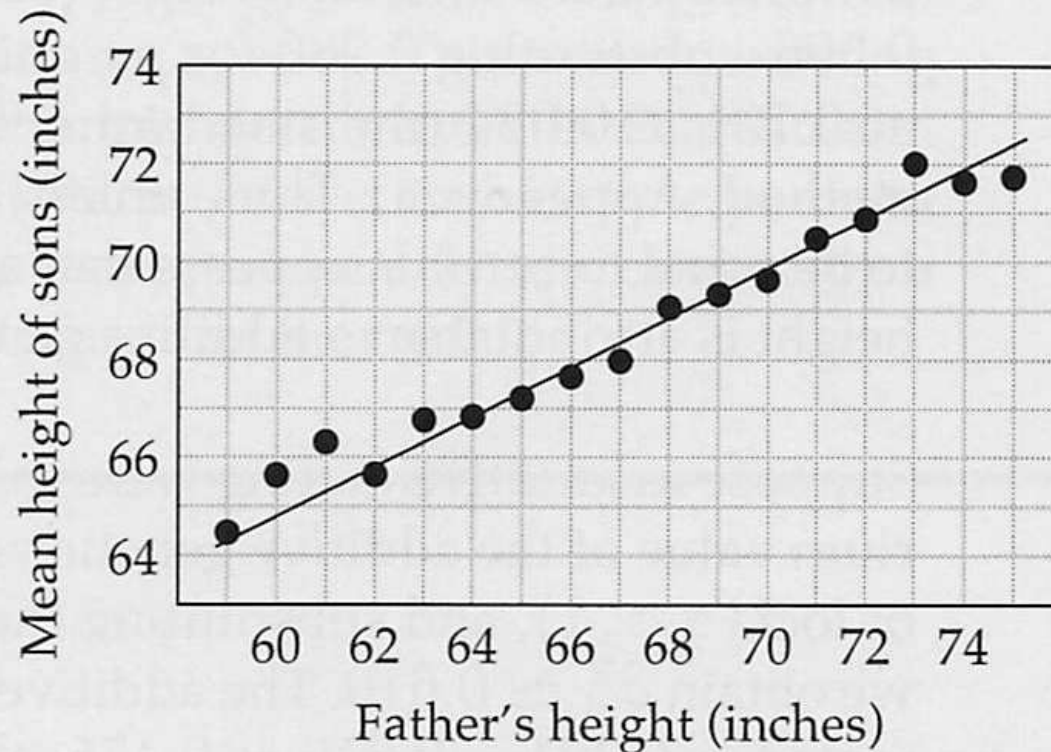
La covariance entre les phénotypes mesurés sur différent individus

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



Relation linéaire entre
taille des pères et taille de
leurs fils

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

La covariance entre les phénotypes mesurés sur différent individus

Répétition de l'expérience de Pearson (Gillespie)

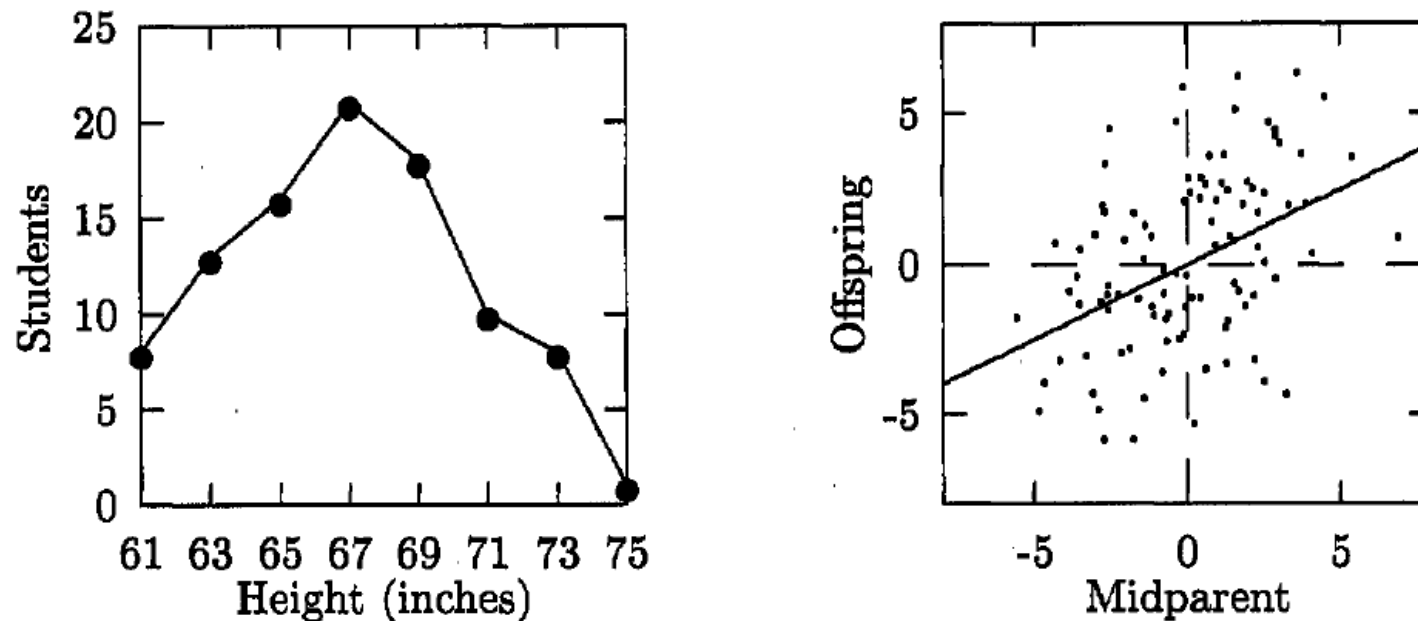


Figure 5.1: The left-hand figure is a histogram of the number of students of a particular height in an evolution class at UC Davis. The right-hand figure graphs the deviation of a student's height from the population mean against the deviation of the student's parents' average height from the population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

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

La covariance entre les phénotypes mesurés sur différent individus

...encore un exemple avec données sur des espèces sauvages...

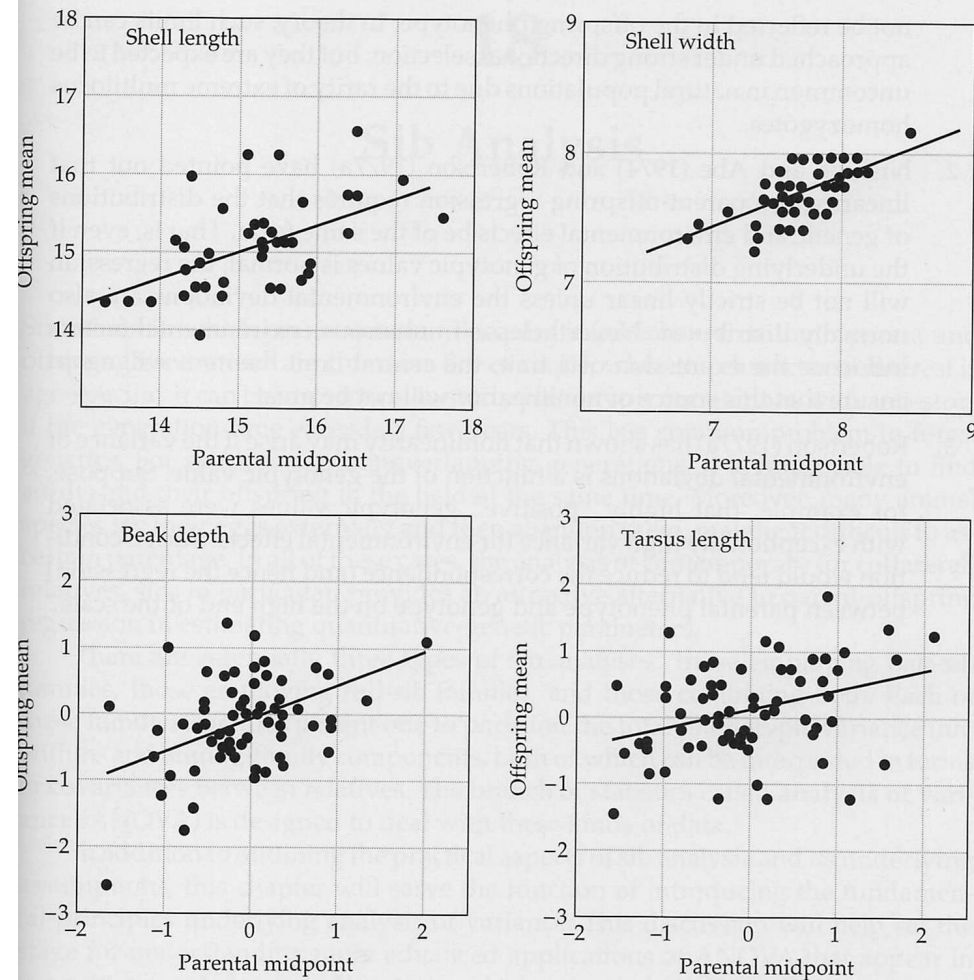
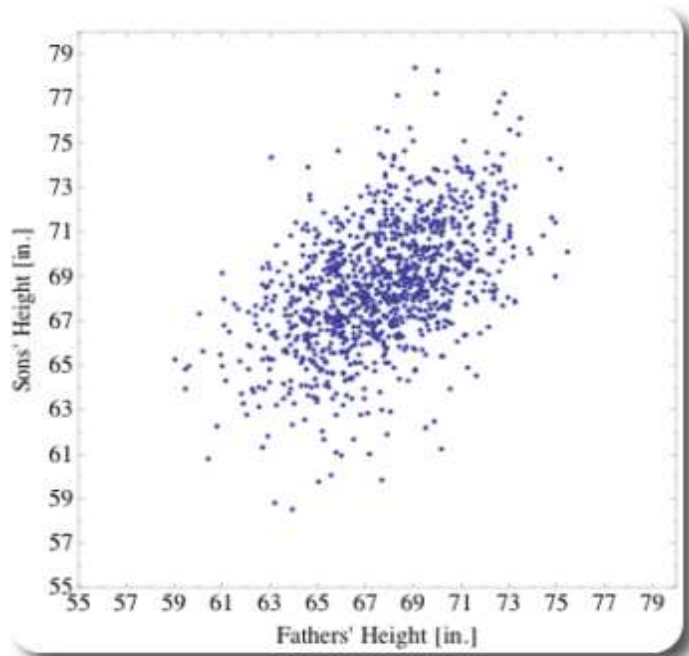


Figure 17.4 Examples of midparent-offspring regressions. **Top panels:** Land snail *Partula taeniata* (Murray and Clarke 1968). **Bottom panels:** Song sparrow *Melospiza melodia* (Smith and Zach 1979).

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

La covariance entre les phénotypes mesurés sur différent individus



Pearson a obtenu **une corrélation de 0.5 père-fils** pour la taille

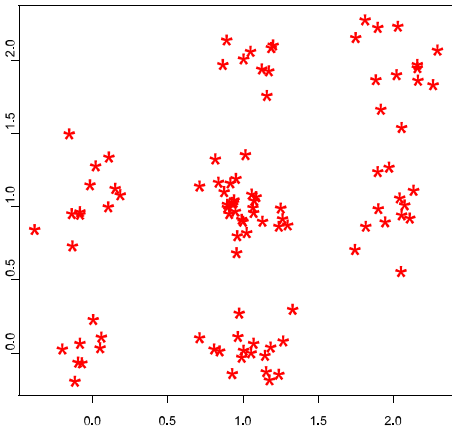
Les « mendéliennistes » trouvaient « louche » cette notion de corrélation car elle était en conflit avec la ségrégation mendélienne.

Mais en fait il n'y a pas de conflit de pensée, car ici on ne voit pas la ségrégation en raison de la multitude de gènes causaux.

Probabilité d'identité par descendance et ressemblance entre les phénotypes

Distribution des phénotypes, n loci causaux bi-alléliques inconnus, $h^2=0,95$

Phénotype du père
 $pheno_dad$

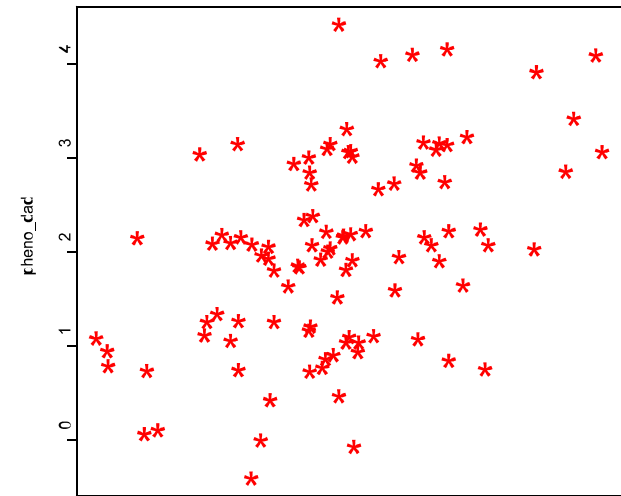


1 locus

On voit bien tous les génotypes possibles du père et fils

Phénotype du fils

$pheno_dad$

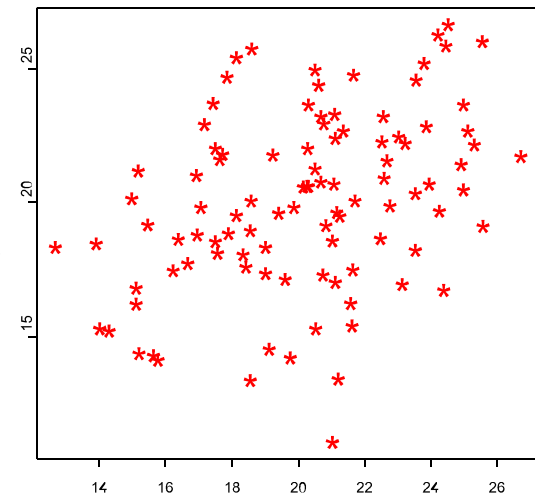


2 loci

Phénotype du fils

Phénotype du père

Phénotype du père
 $pheno_dad$



20 loci

On ne voit plus rien...
néanmoins la corrélation est
toujours de 0.5

Phénotype du fils

Pour mémoire coeff de parenté père-fils ($\Phi_{ij}=0.25$). Nous verrons plus loin le lien entre corrélation et coeff. de parenté.

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 37

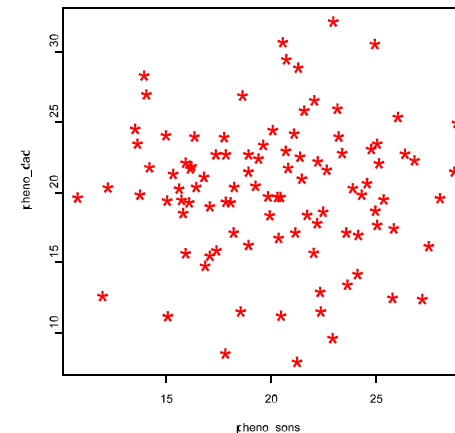
Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

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

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

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



Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

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

$$P_i = G_i + E_i$$

On peut aussi estimer la valeur génétique d'un individu i 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

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

$$\text{Var} \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_n \end{pmatrix} = \mathbf{A} V_A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix} V_A$$

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

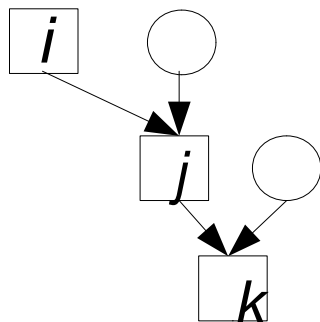
Probabilité d'identité par descendance et ressemblance entre les phénotypes

Covariance entre apparentés

La covariance entre apparentés généralisé

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

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



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

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

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

Pour les « plus statisticiens » : la manière d'introduire cette matrice → distribution normale multivariée

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

Probabilité d'identité par descendance et ressemblance entre les phénotypes

La dépression consanguine

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

Les individus consanguins sont moins viables (productifs, grands, etc.) que ne le sont pas les individus non consanguins.

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

Pour éviter ce phénomène : gestion de la consanguinité

- éviter les couples d'individus avec des apparentés proches

Probabilité d'identité par descendance et ressemblance entre les phénotypes

La dépression consanguine

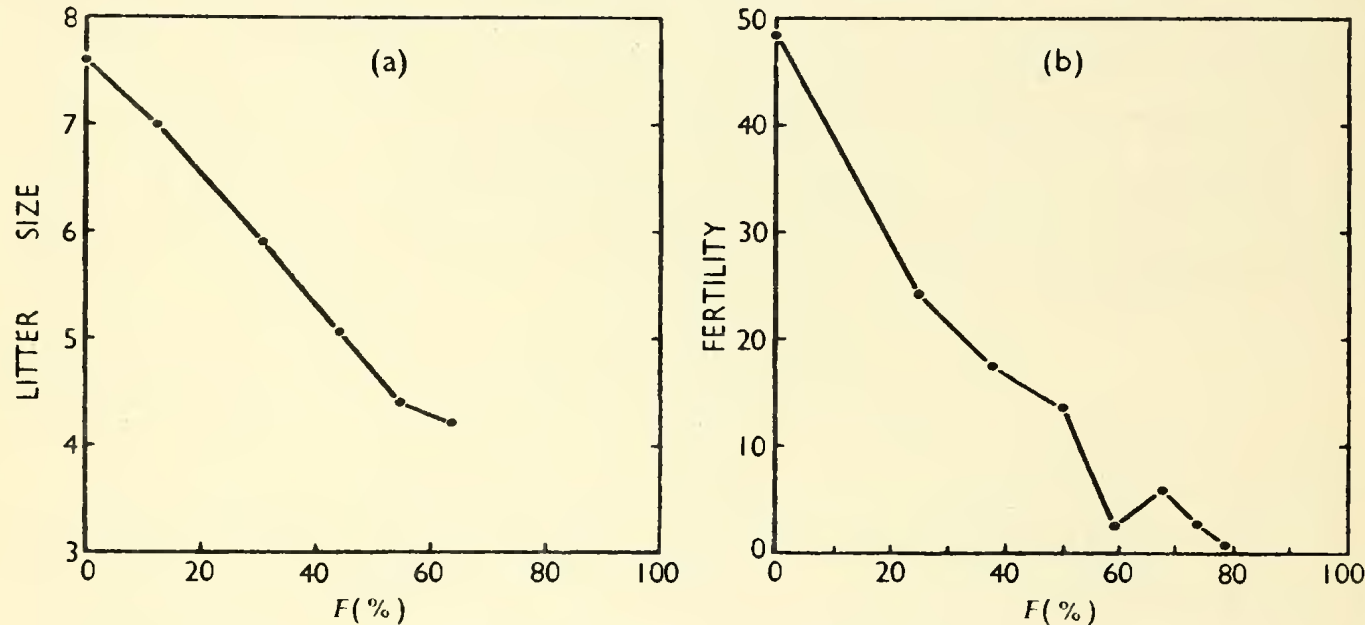


FIG. 14.1. Examples of inbreeding depression affecting fertility. (a) Litter-size in mice (original data). Mean number born alive in 1st litters, plotted against the coefficient of inbreeding of the litters. The first generation was by double-first-cousin mating; thereafter by full-sib mating. No selection 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.)

F : % de consanguinité

Utilisation de marqueurs ADN

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

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

Deux possibilités :

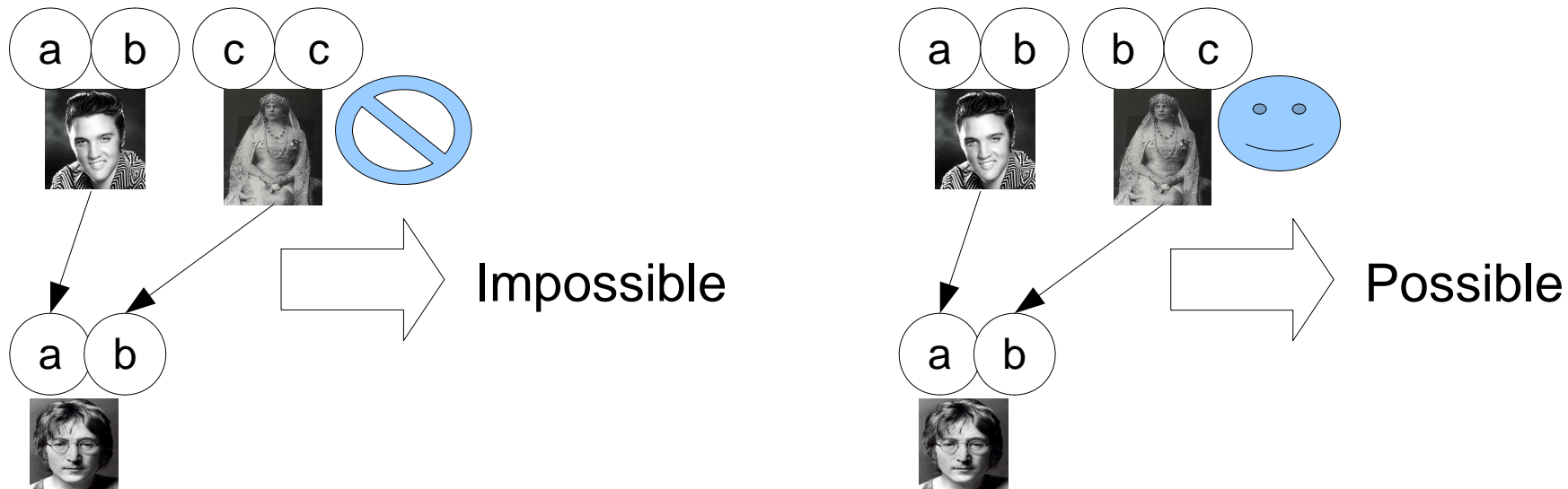
- On connaît une généalogie partielle ou incomplète → vérification de paternités
- On ne connaît rien :
 - estimation de la généalogie la plus probable
 - utilisation du parenté moléculaire

Solution : utilisation de marqueurs ADN

Utilisation de marqueurs ADN

Vérification 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 !!

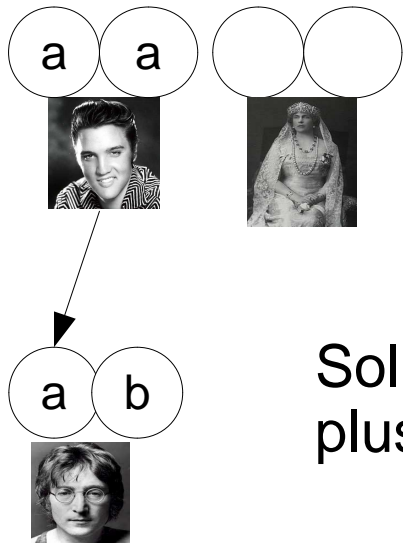
Utilisation de marqueurs ADN

Verification de paternités

Cas simplifié : de fait, « a » vient du père et « b » vient-il de cette mère ?

Il suffit que la « fausse » mère soit porteuse de l'allèle b pour qu'on ne l'exclut pas, avec une probabilité :

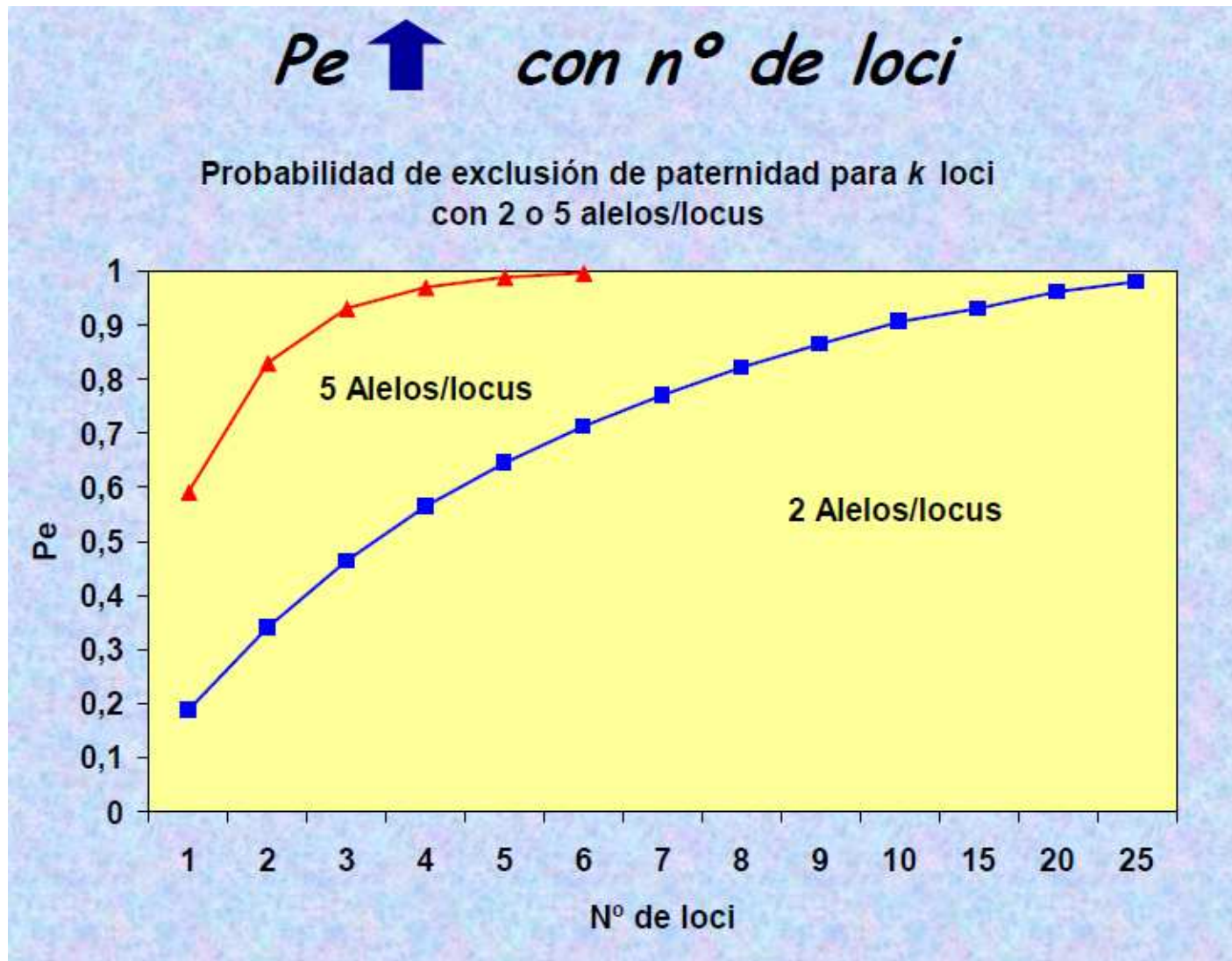
$$\Pr(\text{no exclusion} \mid \text{n'est pas mère}) = 2 * \text{freq}(b)$$



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, pour des espèces de rente, et aussi pour quelques populations sauvages

Utilisation de marqueurs ADN

Assignment de parentés ou reconstruction de la généalogie

Beaucoup plus difficile

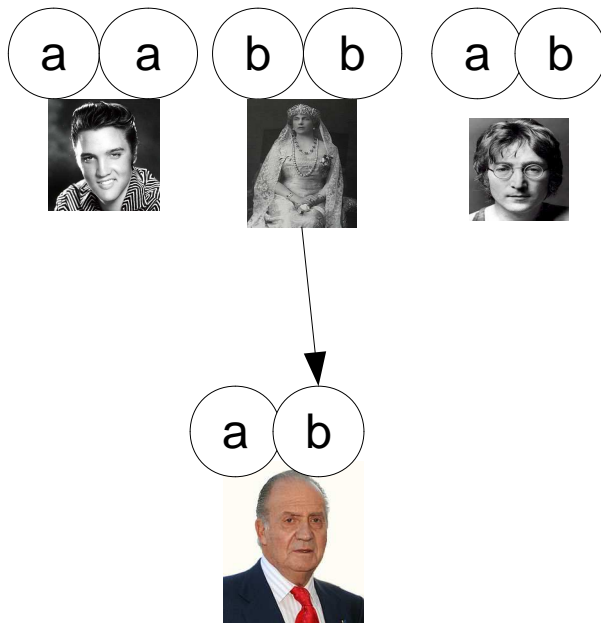
Sachant un ensemble d'individus et ses génotypes, il s'agit de reconstruire, soit les parentés « proches » (père-fille, père-fils, plein germains, demi-germains)

Cas assez difficile

Utilisation de marqueurs ADN

Assignment de parentés ou reconstruction de la généalogie

D'après ce schéma, qui est le papa le plus probable?

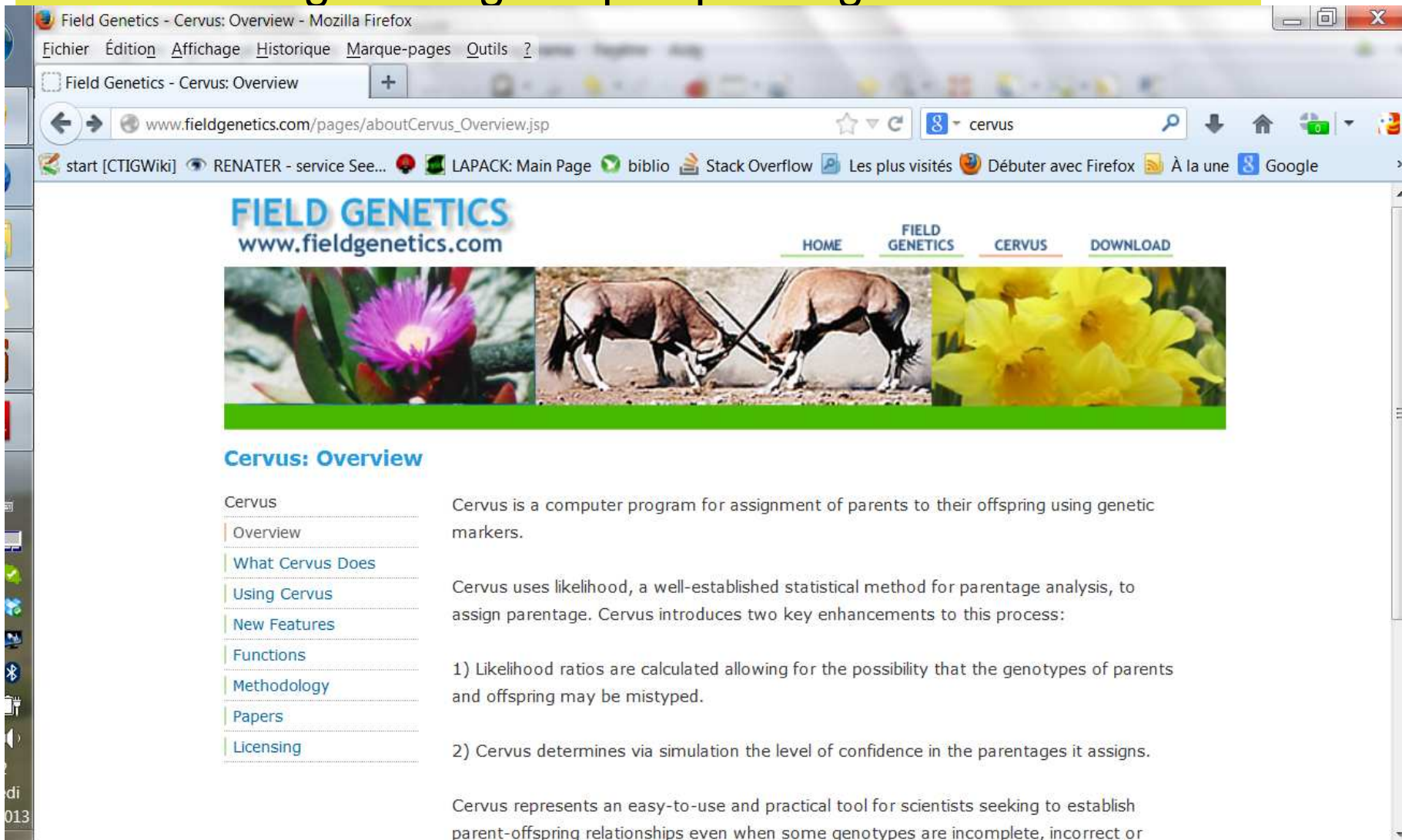


Elvis aurait certainement passé l'allèle « a »
Lennon aurait passé l'allèle « a » la moitié
des fois

$\text{Pr}(\text{Elvis papa})=2/3$
 $\text{Pr}(\text{Lennon papa})=1/3$

Utilisation de marqueurs ADN

Assignation de parentés ou reconstruction de la généalogie : quelques logiciels



The screenshot shows a Mozilla Firefox browser window displaying the 'Field Genetics - Cervus: Overview' page. The browser's address bar shows the URL 'www.fieldgenetics.com/pages/aboutCervus_Overview.jsp'. The website header includes the 'FIELD GENETICS' logo and the URL 'www.fieldgenetics.com'. A navigation menu contains links for 'HOME', 'FIELD GENETICS', 'CERVUS', and 'DOWNLOAD'. Below the navigation menu is a banner image featuring three panels: a purple flower, two donkeys facing each other, and a yellow flower. The main content area is titled 'Cervus: Overview' and contains the following text:

Cervus

- Overview
- What Cervus Does
- Using Cervus
- New Features
- Functions
- Methodology
- Papers
- Licensing

Cervus is a computer program for assignment of parents to their offspring using genetic markers.

Cervus uses likelihood, a well-established statistical method for parentage analysis, to assign parentage. Cervus introduces two key enhancements to this process:

- 1) Likelihood ratios are calculated allowing for the possibility that the genotypes of parents and offspring may be mistyped.
- 2) Cervus determines via simulation the level of confidence in the parentages it assigns.

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

Assignment de parentés ou reconstruction de la généalogie : quelques logiciels

Jesús Fernández - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

Jesús Fernández

dl.dropboxusercontent.com/u/5714008/Fernandez.htm

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

SOFTWARE

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](#).
[\[Download\]](#)

ADDRESS

RESEARCH

PROJECTS

[Past projects](#)

[Running projects](#)

PUBLICATIONS

[2011 – present](#)

[2006 - 2010](#)

[2001 - 2005](#)

[2000 and before](#)

SOFTWARE

CURRICULUM

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 risque de se tromper, il y aura des morceaux de la généalogie mal reconstruits.

Solution : utilisation des coefficients d'apparenté « moléculaires »

On ne va pas essayer de reconstruire une généalogie pour calculer les coefficients. On va calculer 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...
que 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 existe 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

$$\Phi_{Mij} = 2/4 = 0,5$$

On peut le moyenner sur plusieurs loci...

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

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

$$\Phi_{Mij} = 3,5/6 = 0,583$$

avec L : nb loci 56

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

$$\Phi_{Mij} = 2/4 = 0,5$$

Mais, deux individus peuvent être identiques « par état » sans 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

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

L : nb loci

$$\bar{g}_i = \frac{1}{L} \sum_L g_{ik}$$

$$Cov_{Mij} = 0,0694$$

$$g_{i1} \quad g_{j1}$$

$$\bar{g}_i = 2/6 = 1/3$$

$$\bar{g}_j = 4/6 = 2/3$$

La ressemblance “moléculaire” entre apparentés

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

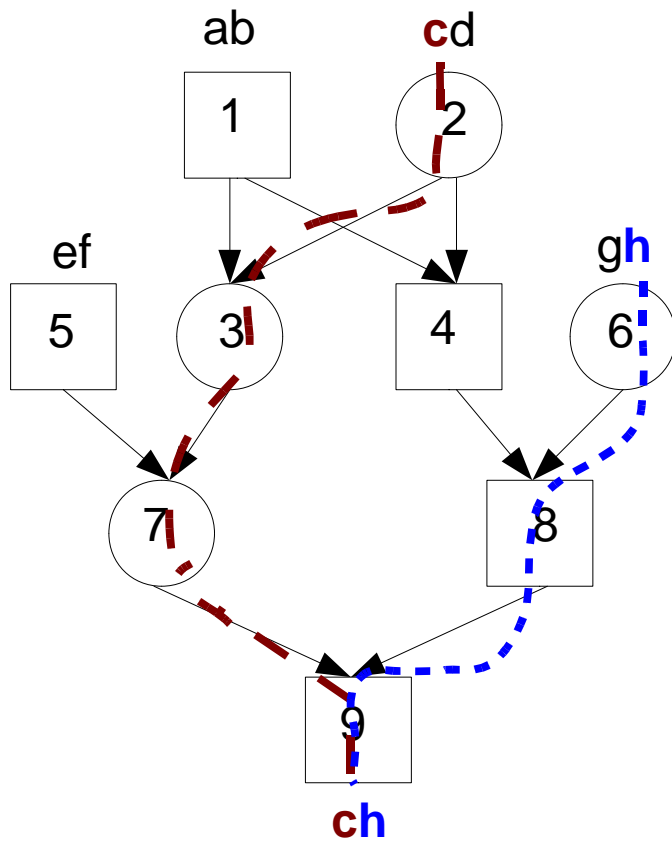
Supposons qu'il existe un marqueur pour lequel **tous les allèles sont différents chez les fondateurs**

Nous génotypons l'individu 9

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

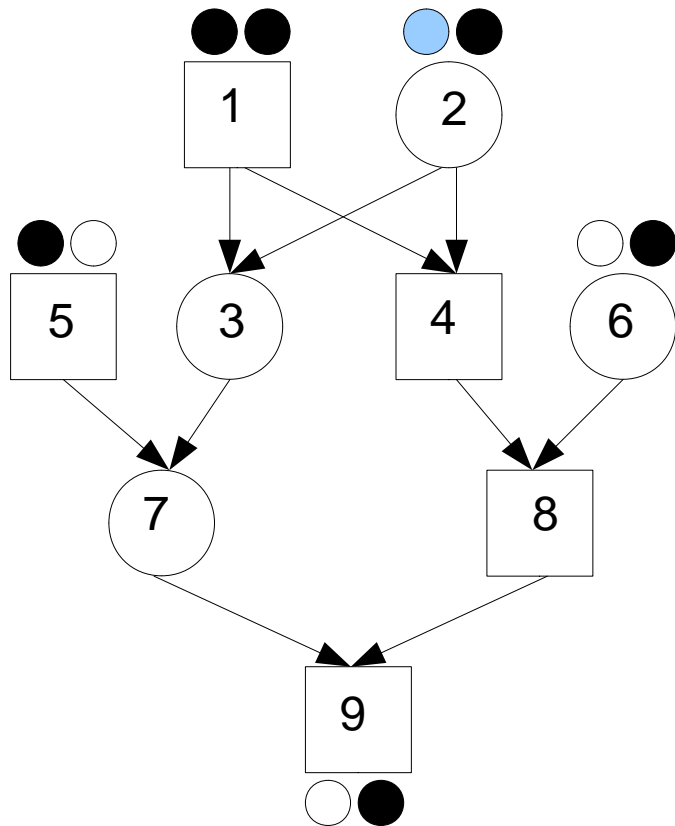
Cette identité par descendance « observé » est *plus exacte* que la probabilité d'identité par descendance (qui moyenne toutes les possibilités)

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



La ressemblance “moléculaire” entre apparentés

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



C'est bien si les marqueurs étaient complètement multialléliques

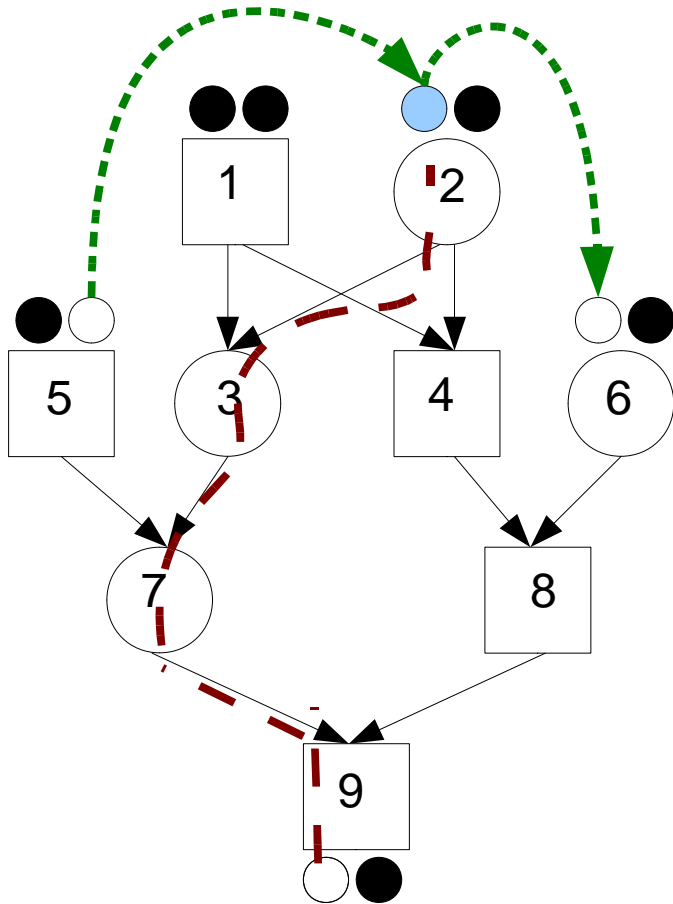
Comment faire pour des marqueurs bialléliques ?

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

On génotype à nouveau l'individu 9

La ressemblance “moléculaire” entre apparentés

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



On verra une identité par état que si :
a) on est identique **par descendance**

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

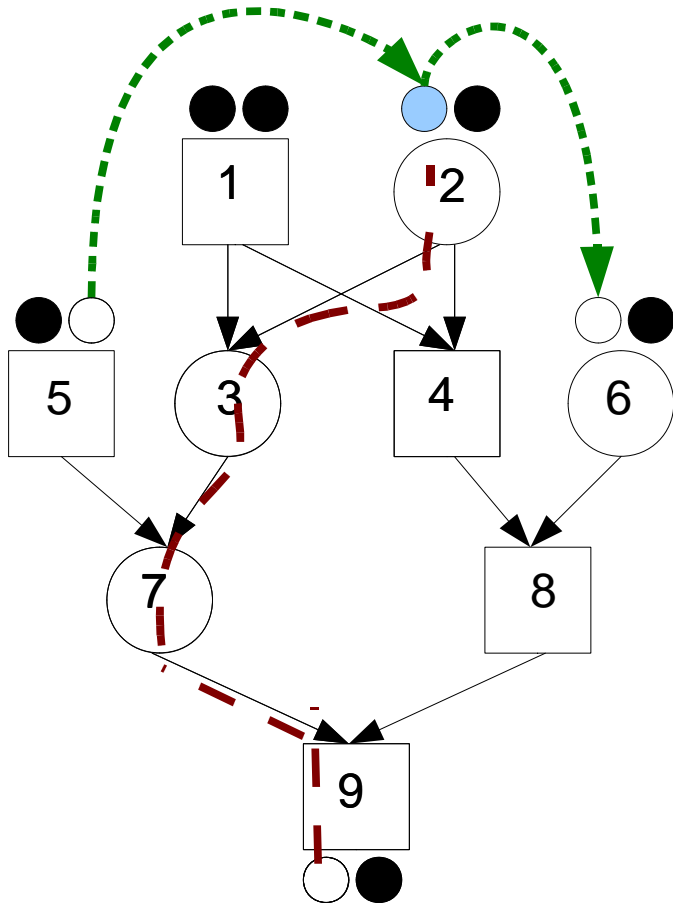
b) on n'est pas identique par descendance
mais **les fondateurs étaient identiques par état**

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

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

La ressemblance “moléculaire” entre apparentés

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



De même pour les covariances, on peut montrer que :

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

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

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

La ressemblance “moléculaire” entre apparentés

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

On remarque que la formule du coefficient de parenté moléculaire ressemble à des formules de génétique de population :

$$(1 - \phi_M) = (1 - \phi) \left(1 - \sum p_i^2\right)$$

$$(1 - F_{IT}) = (1 - F_{IS}) (1 - F_{ST}) \quad \leftarrow \text{cf. cours de génétique des population}$$

C'est un peu la même chose...

$$(1 - \phi_M) = (1 - \phi) \left(1 - \sum p_i^2\right)$$

Heterozigosité due à
identité par
descendance

Heterozigosité dans la
population de base

Heterozigosité
moléculaire

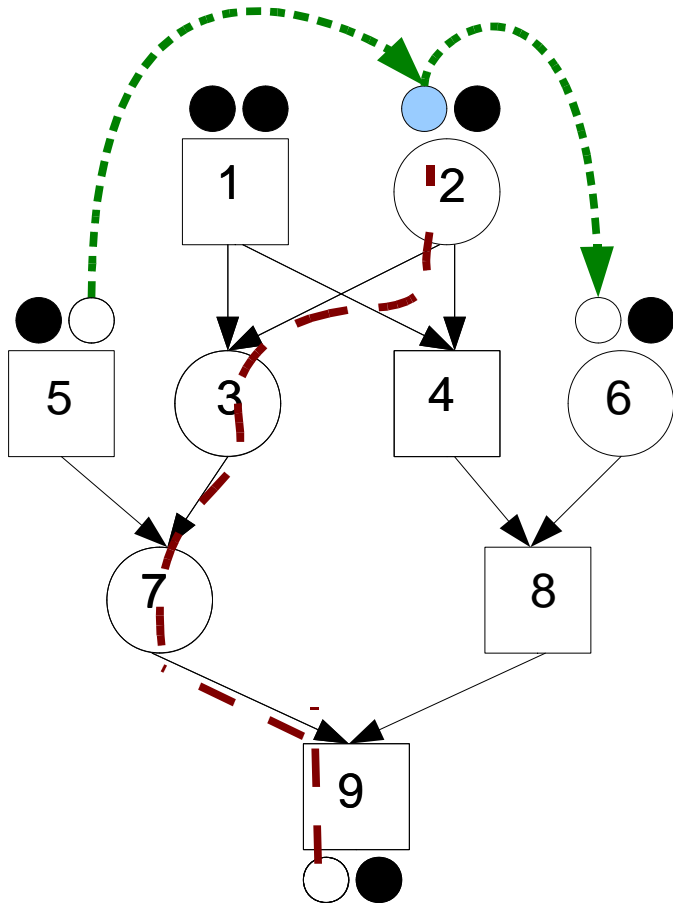
La ressemblance “moléculaire” entre apparentés

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

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

La ressemblance “moléculaire” entre apparentés

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



Trois estimateurs typiquement utilisés

$$\hat{\phi}_{i,j} = \frac{1}{L} \sum_L (\phi_{Mij,l} - p_l^2 - q_l^2) / 2p_l q_l \longrightarrow \text{Parenté moléculaire}$$

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

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

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 », donc **de définir p**
- Souvent, on n'a pas accès aux fréquences alléliques de cette population ancienne
- On prend les fréquences de la population actuelle
- Avec le cours du temps et la dérive , $p \rightarrow 0$ ou 1

- Quel effet ?

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

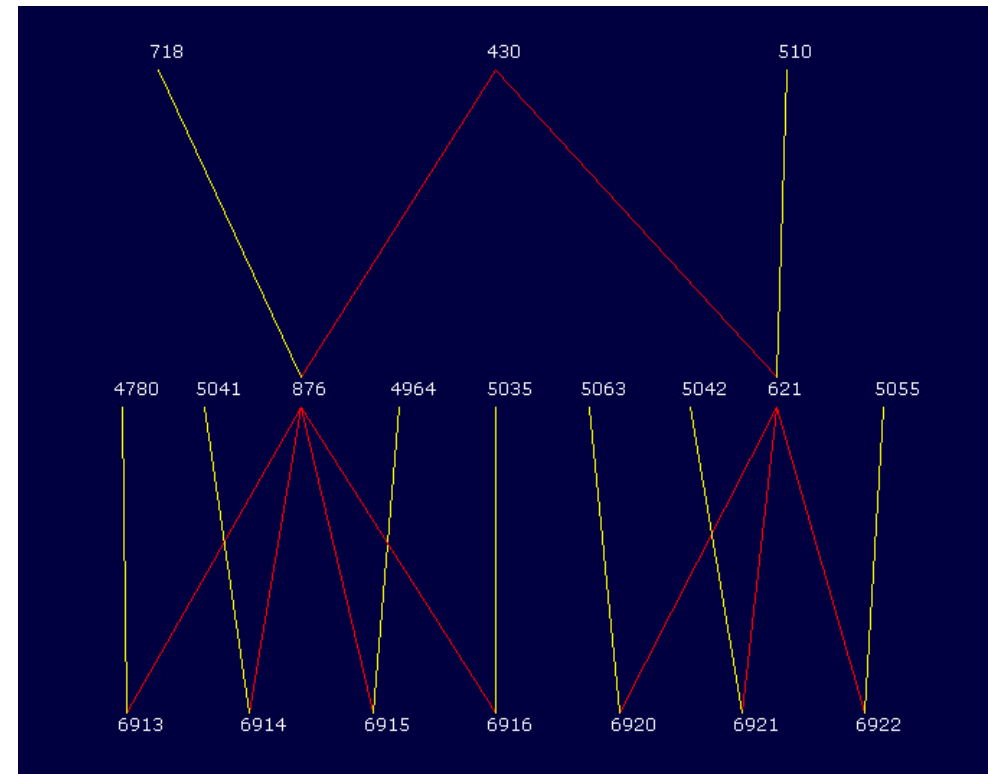
On sous-estime les coefficients de parenté : coefficients estimés négatifs
Mais ils sont très corrélés aux « vrais ».

La ressemblance “moléculaire” entre apparentés

Exemple réel

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

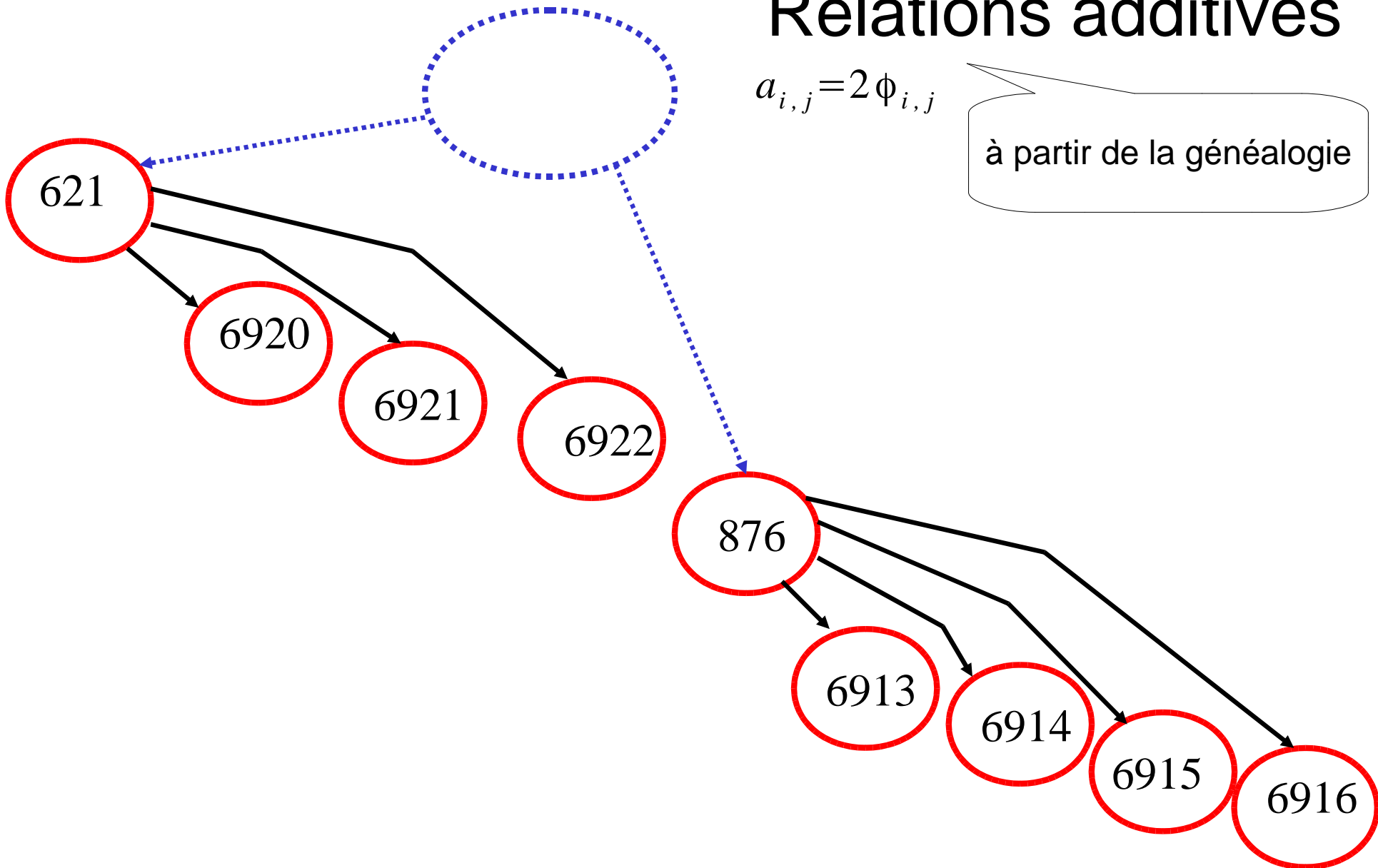
Voici une généalogie simplifiée



Relations additives

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

à partir de la généalogie

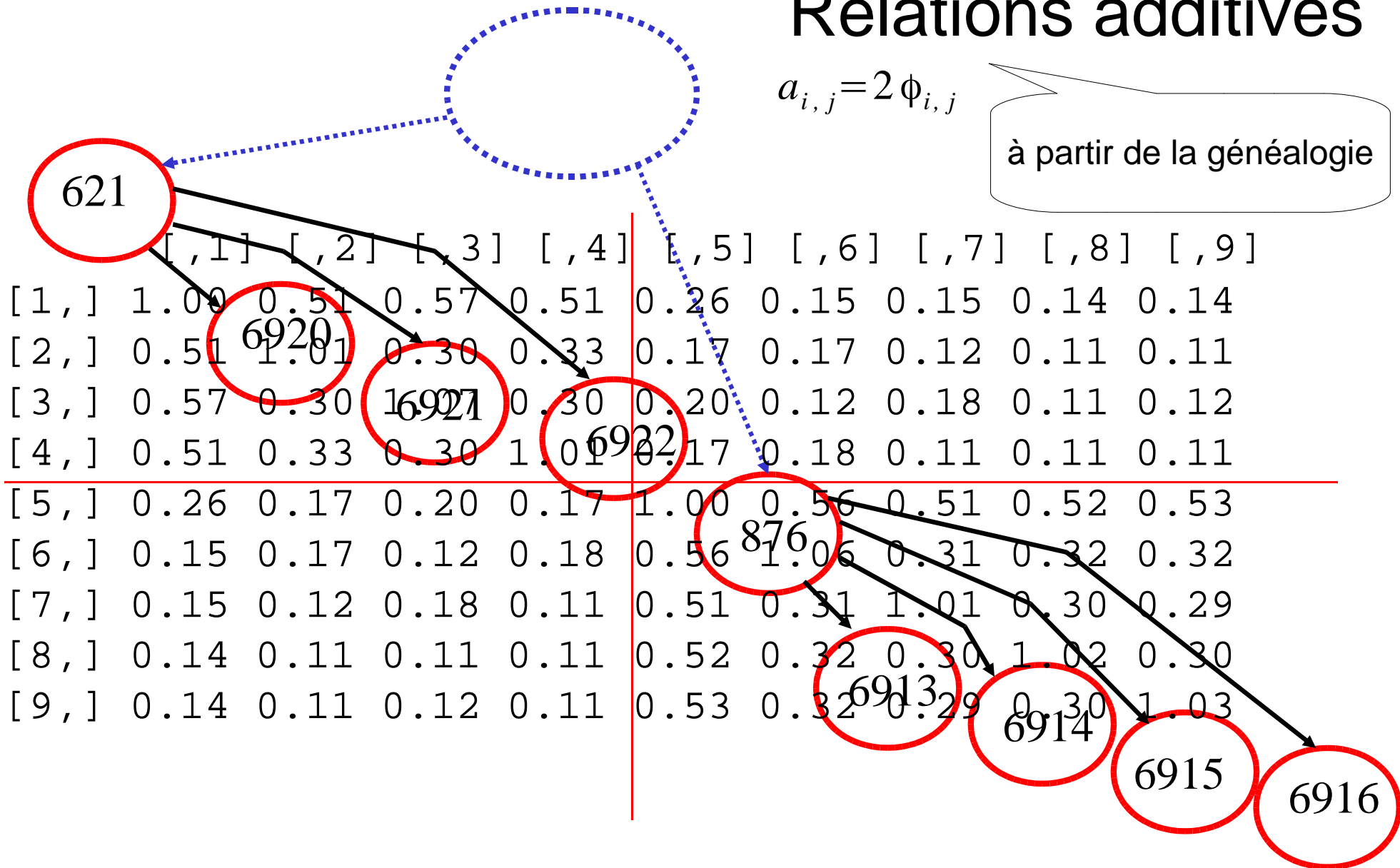


Il s'agit de deux familles de « demi-frères » issues des deux taureaux (621 et 876) qui sont des « demi-frères »

Relations additives

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

à partir de la généalogie



Il s'agit de deux familles de « demi-frères » issues des deux taureaux (621 et 876) 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 (1/8) 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 sur la diagonale

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

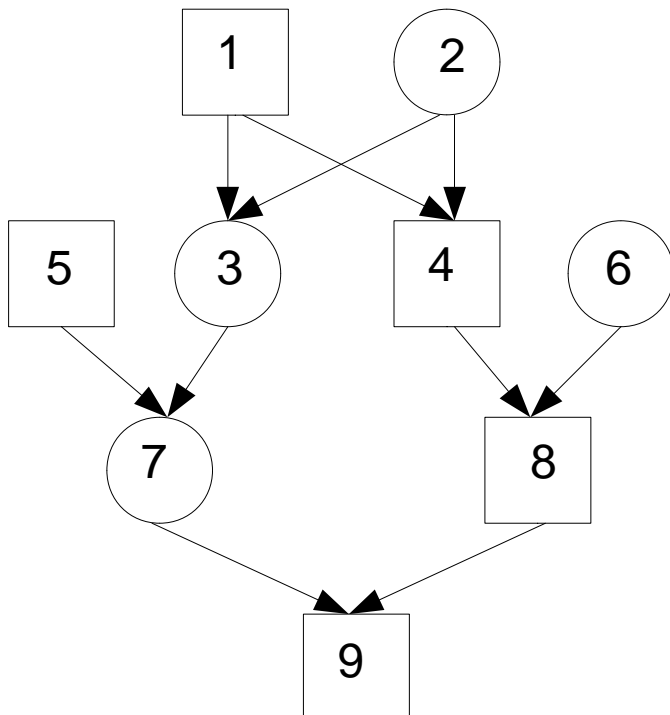
Les valeurs négatives indiquent que la valeur des fréquences p 's est mal estimée. Mais ce n'est pas très « important » car on décale tout d'une constante et la formule des covariances est

juste, même si $a_{ij} < 0$

$$\text{cov}(\mathbf{G}_i, \mathbf{G}_j) = \text{cov}(A_i, A_j) + \text{cov}(D_i, D_j) = \mathbf{a}_{ij} V_A + \mathbf{d}_{ij} V_D$$

La ressemblance entre apparentés

Resumé



Les coefficients de parenté permettent de quantifier la ressemblance *génétique* entre apparentés et donc la ressemblance *phénotypique*

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

On a des outils efficaces de les calculer, à partir de la généalogie ou 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. [// *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.
- **Compléments sur la génétique quantitative, en particulier la parenté et la consanguinité**
 - <http://www.agroparistech.fr/svs/genere/uvf/index.htm>