

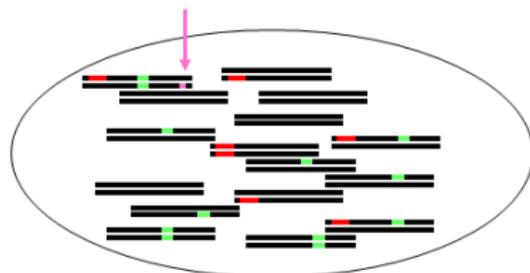
# Génétique d'association

INRA, Brigitte Mangin

Septembre 2016

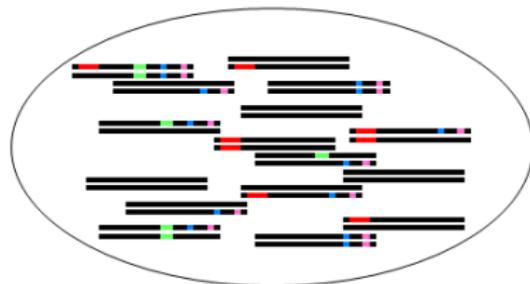
# Plan

- 1 **Introduction**
- 2 **Estimer le DL**
  - Les mesures usuelles
  - Les nouvelles mesures
- 3 **Tester l'association**
  - Les phénotypes "maladies"
  - Les phénotypes continus : le modèle le plus simple
  - Les phénotypes continus : corriger de la structure
  - Les phénotypes continus : corriger de la structure et de l'apparentement
  - Estimations et tests dans le modèle linéaire mixte
  - Seuil
- 4 **La covariance génétique**
  - Calculer  $\Sigma_G$ , souvent appelée kinship
  - Caractéristique d'une "bonne" kinship
- 5 **Pour finir**
  - Illustration

Contexte évolutif

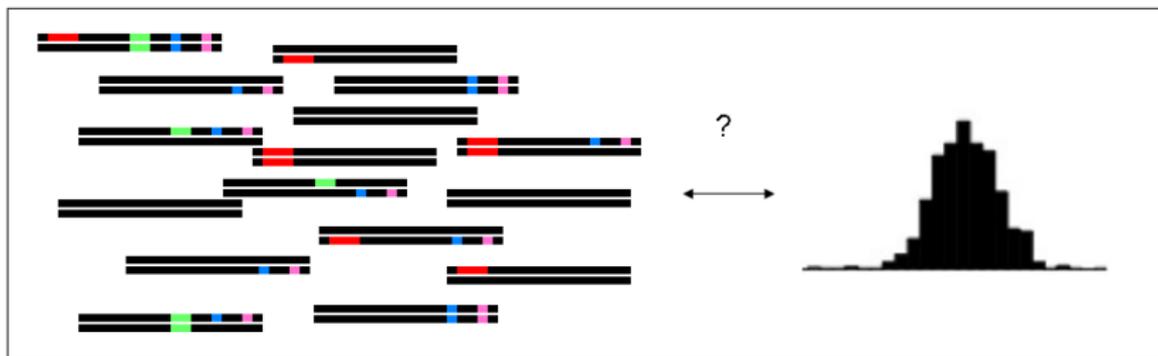
Soit une population a un temps  $t_0$  avec apparition d'une **nouvelle mutation**

↓  
 $T$  générations  
↓



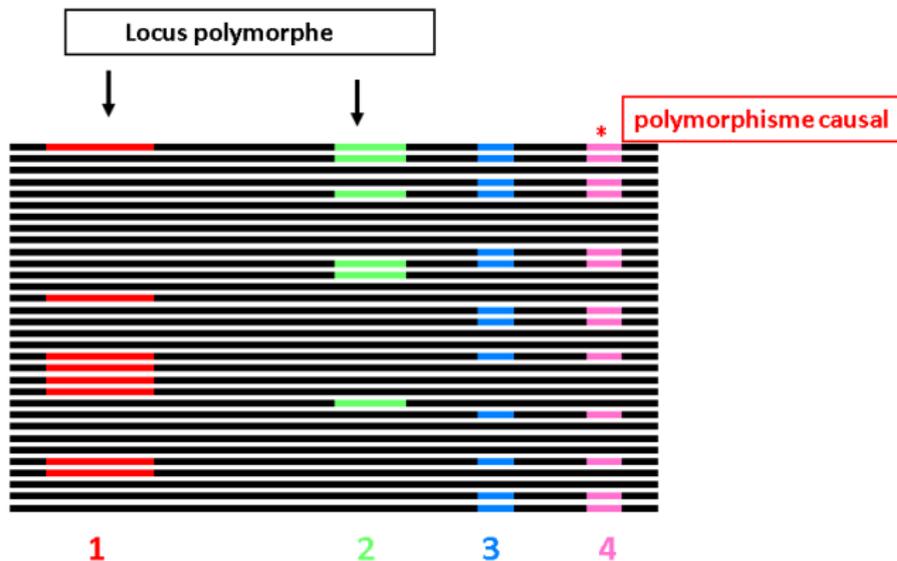
Au sein de cette même population a un temps  $(t_0 + T)$  ce site polymorphe a augmenté en fréquence

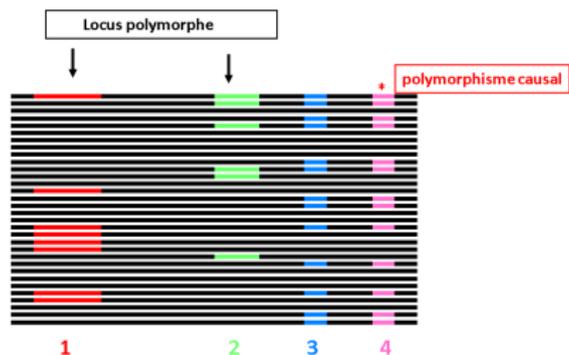
# Objectif



Quel est le polymorphisme causal ?

- Tester une différence d'effet de l'allèle muté par rapport à l'allèle sauvage
- en tout locus polymorphe





Les tests aux loci 3 et 4  
sont identiques.

C'est le fait du déséquilibre de liaison (DL), que l'on exploite  
pour

- réduire le génotypage (TAG SNP)
- en conservant la "couverture" du génome (ou de la région génomique)

# Définition du déséquilibre de liaison

## Wikipedia

On dit qu'il y a déséquilibre de liaison si la fréquence des gamètes porteurs des allèles de deux locus différents est différente du produit des fréquences des allèles.

C'est-à-dire s'il y a une association préférentielle entre deux allèles de deux locus.

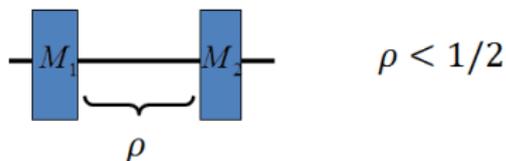
# Mathématiquement parlant

Pour deux marqueurs  $M_1$  et  $M_2$ , la définition actuelle assimile le déséquilibre de liaison au

déséquilibre gamétique  $\pi_{ij} \neq \pi_{i+} \pi_{+j}$

Loci		$M_2$		Total
	Allèles		$i$	
$M_1$				
	$j$		$\pi_{ij}$	$\pi_{+j}$
Total			$\pi_{i+}$	

C'est une notion différente de la liaison génétique des deux marqueurs.



# Les mesures usuelles : cas biallélique

$$D_{ij} = \pi_{ij} - \pi_{i+} \pi_{+j}$$

$$D \begin{cases} = D_{11} = D_{22} \\ = -D_{12} = -D_{21} \end{cases}$$

$$r^2 = \frac{D^2}{\pi_{1+} \pi_{+1} \pi_{2+} \pi_{+2}}$$

Mesure du  $\chi^2$

$$D' = \begin{cases} \frac{D}{\min(\pi_{1+} \pi_{+2}, \pi_{2+} \pi_{+1})} & \text{si } D > 0 \\ \frac{D}{\min(\pi_{1+} \pi_{+1}, \pi_{2+} \pi_{+2})} & \text{si } D < 0 \end{cases}$$

Mesure comprise entre -1 et +1  
Lewontin (1964)

## Un regard plus statistique

- $\Delta_{M_1,i}$  la dose d'allèle  $i$  au locus  $M_1$
- $\Delta_{M_2,j}$  la dose d'allèle  $j$  au locus  $M_2$

$$D_{ij} = \text{Cov}(\Delta_{M_1,i}, \Delta_{M_2,j}) \quad r_{ij}^2 = \text{Cor}^2(\Delta_{M_1,i}, \Delta_{M_2,j})$$

## Pour des haplotypes (phase connue)

$$\Delta_{M_1,i} = 0 \text{ ou } 1 \quad (\text{idem } \Delta_{M_2,j})$$

## Pour des génotypes (phase inconnue)

$$\Delta_{M_1,i} = 0, 1, 2 \quad (\text{idem } \Delta_{M_2,j})$$

Rogers & Huff, Genetics, 2009

# Estimation de la mesure $r^2$

Soit l'observation des doses alléliques pour un échantillon de taille  $N$

$$\Delta_{M_1,i} = \begin{pmatrix} \delta_{M_1,i,1} \\ \vdots \\ \delta_{M_1,i,n} \\ \vdots \\ \delta_{M_1,i,N} \end{pmatrix} \quad \Delta_{M_2,j} = \begin{pmatrix} \delta_{M_2,j,1} \\ \vdots \\ \delta_{M_2,j,n} \\ \vdots \\ \delta_{M_2,j,N} \end{pmatrix}$$

$$\hat{r}_{ij}^2 = \widehat{\text{Cor}}^2(\Delta_{M_1,i}, \Delta_{M_2,j})$$

où  $\widehat{\text{Cor}}$  est la corrélation empirique

# Exemple

Soit l'observation des doses alléliques pour un échantillon de taille  $N = 10$ . Le premier SNP a pour allèles A/T et le second T/A.

$$\Delta_{M_1,A} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 2 \\ 2 \\ 1 \\ 2 \\ 2 \\ 0 \end{pmatrix} \quad \Delta_{M_2,T} = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 2 \\ 2 \\ 1 \\ 2 \\ 1 \end{pmatrix}$$

$$\hat{r}_{ij}^2 = 0.54$$

dans R `cor(.)^2`

# Biais et solution

$\hat{r}^2$  est biaisé,  $E_{sp}(\hat{r}^2) \neq r^2$

- lorsque l'échantillon a une structure
- lorsque les individus sont fortement apparentés et d'apparentements contrastés

Des mesures qui corrigent ces biais  $r_S^2, r_V^2, r_{VS}^2$

Mangin et al., Heredity, 2011

- lorsque la structure est connue (ou estimée)  $S$
- lorsque l'apparentement est connu (ou estimé)  $V$

# Définition

## corrigée de la structure $S$

$$r_{S,ij}^2 = \text{Corr}^2(\Delta_{M_1,i}, \Delta_{M_2,j}; S)$$

## corrigée de l'apparentement $V$

$$r_{V,ij}^2 = \text{Cor}^2(V^{-1/2}\Delta_{M_1,i}, V^{-1/2}\Delta_{M_2,j})$$

## corrigée de la structure et de l'apparentement $S, V$

$$r_{VS,ij}^2 = \text{Corr}^2(V^{-1/2}\Delta_{M_1,i}, V^{-1/2}\Delta_{M_2,j}; S)$$

où  $\text{Corr}(X, Y; Z)$  dénote la corrélation partielle de  $X$  et  $Y$  lorsque  $Z$  est constant,  
ou encore la corrélation des résidus  $\epsilon_X$  et  $\epsilon_Y$  des régressions linéaires  $X = S\beta + \epsilon_X$  et  $Y = S\beta' + \epsilon_Y$

# Estimation des nouvelles mesures

Comme pour  $r^2$ , la corrélation est estimée par la corrélation empirique.

On utilise la matrice  $S$  de structure en  $K$  groupes de l'échantillon

$$S = \begin{bmatrix} S_{1,1} & \dots & S_{1,K} \\ \vdots & \vdots & \vdots \\ S_{n,1} & \dots & S_{n,K} \\ \vdots & \vdots & \vdots \\ S_{N,1} & \dots & S_{N,K} \end{bmatrix}$$

## Exemple

3 plants de maïs d'Europe et 2 plants de maïs des Etats Unis  
 $\Rightarrow$  2 groupes

$$S = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

## Les nouvelles mesures

On utilise la matrice  $V$  de variance-covariance de l'échantillon

$$V = \begin{bmatrix} V_{1,1} & \dots & V_{1,n} & \dots & V_{1,N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ V_{n,1} & \dots & V_{n,n} & \dots & V_{n,N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ V_{N,1} & \dots & V_{N,n} & \dots & V_{N,N} \end{bmatrix}$$

## Exemple

2 plants de maïs plein-frères non-apparentés à 3 plants de maïs non-apparentés.

$$V = \begin{bmatrix} 1 & 0.25 & 0 & 0 & 0 \\ 0.25 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$



## Exemple pour la correction de l'apparentement

On observe 110 abricotiers, dont les 50 premiers ont tous été clonés à partir d'un seul individu. Les autres sont tous non apparentés. Les deux marqueurs sont pour le premier C/T et le second G/A.

$$\Delta_{M_1, C}^t = \left\{ \underbrace{1, \dots, 1}_{50 \text{ cl\u00f4nes}}, \underbrace{1, \dots, 1}_{10}, \underbrace{0, \dots, 0}_{50} \right\} \quad \Delta_{M_2, G}^t = \left\{ \underbrace{1, \dots, 1}_{50 \text{ cl\u00f4nes}}, \underbrace{1, 0, \dots, 1, 0}_{30} \right\}$$

$$\hat{r}^2 = 0.217 \quad \hat{r}_V = 0.001$$

En ne gardant qu'un seul des cl\u00f4nes :  $\hat{r}^2 = 0.001$

# Bien estimer le DL, pourquoi ?

## Pour limiter le génotypage

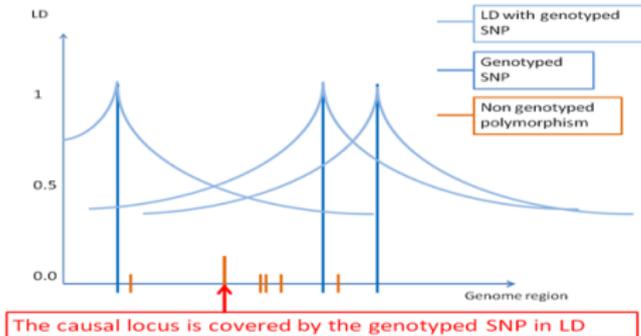
Deux SNP en fort DL apportent une information redondante, il n'est donc pas d'un grand intérêt de les génotyper tous les deux.

## Pour "couvrir" toute la région génomique d'intérêt

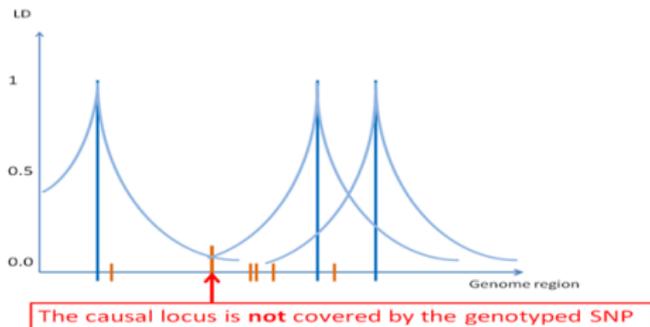
L'objectif est que tous les SNP non génotypés soient "couverts" par au moins un SNP génotypé en fort DL.

# Illustration

- Long extent of LD



- Short extent of LD



# Tester l'association

## les phénotypes de maladie

le dispositif cas-contrôle

## les phénotypes continus

- modèle simple
- modèle corrigé des effets de la structure
- modèle corrigé des effets de la structure et de l'apparentement

## Le dispositif cas-contrôle

Le test se calcule à partir de la table de comptage, aussi appelée table de contingence suivante :

Marqueur	$M_1$		Total
	$i$		
Allèles			
Cas malade		$N_{di}$	$N_{d+}$
Contrôle sain		$N_{si}$	$N_{s+}$
Total		$N_{+i}$	$N$

## Les phénotypes "maladies"

Utilisation de  $r^2$ 

Le phénotype "maladie" peut être vu comme un marqueur un peu particulier. Au locus  $l$  de la maladie, au lieu de la dose allélique d'un marqueur, on a

$$\Delta_l = \begin{cases} 1 & \text{si malade} \\ 0 & \text{si sain} \end{cases}$$

On estimera alors  $r_i^2 = \text{Cor}^2(\Delta_l, \Delta_{M_1,i})$  par

$$\hat{r}_i^2 = \widehat{\text{Cor}}^2(\Delta_l, \Delta_{M_1,i})$$

Sous l'hypothèse  $H_0$  : { pas d'association allèle  $i$  avec la maladie }

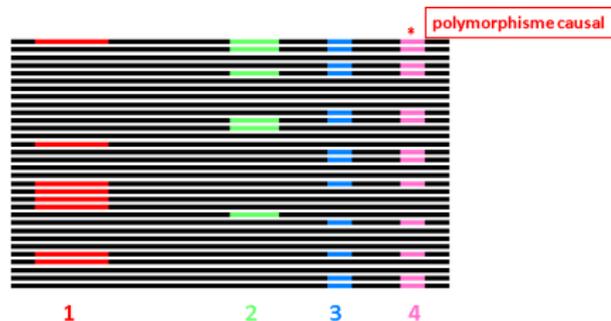
$$\hat{r}_i^2 \sim \chi^2(1)$$

## Les phénotypes "maladies"

Puissance et mesure  $r^2$ 

Pritchard &amp; Przeworski, Am. J. Hum.Genet., 2001

Pour des marqueurs bialléliques



Pour avoir la  
 même puissance,  
 qu'au  
 locus causal

$$r_{14}^2$$

$$r_{24}^2$$

$$r_{34}^2$$

Mesure  $r^2$  de DL

$$N_i = N_4 / r_{i4}^2$$

$$N_1$$

$$N_2$$

$$N_3$$

$$N_4$$

Nbre d'observations

Plus on s'éloigne du locus causal, plus il faut d'observations,  
 pour garder de la puissance de détection.

Les phénotypes continus : le modèle le plus simple

## Modèle linéaire à 1 facteur $\Rightarrow$ table ANOVA

### Effet du SNP fixe

Cas d'individus homozygotes, sans donnée manquante pour le

génotype :  $Y_n$   $\nearrow$   $\text{SNP}_n^l = 1$   $Y_{1k} = \mu + \theta^l + \epsilon_{1k}$

$\searrow$   $\text{SNP}_n^l = 0$   $Y_{0k} = \mu + \epsilon_{0k}$

- $\theta^l$  modélise la différence d'influence des deux allèles du SNP sur le phénotype  $Y$
- $\text{SNP}_n^l$  est la dose d'allèle A pour un SNP en  $l$  qui est A/T, par exemple

### Généralisation

Ce modèle se généralise aux cas de génotypes manquants inférés ou imputés, ainsi qu'aux individus hétérozygotes en choisissant un modèle additif et/ou dominant pour l'effet du SNP

Les phénotypes continus : le modèle le plus simple

## Les limites du modèle

- objectif de l'analyse : rechercher les SNPs qui sont **causaux**. Lorsqu'un SNP est fortement différencié entre deux groupes, il pourra être trouvé associé au phénotype sans être causal.
- postulat du modèle : les observations sont **indépendantes**. Ce n'est pas le cas lorsque les individus observés sont apparentés

### Solution

Pour répondre aux limites du modèle simple, le modèle se complexifie

- pour corriger de la structure
- pour corriger de l'apparentement

# Modèle linéaire à deux facteurs

## Effet du SNP fixe + structure fixe

Cas d'individus homozygotes, sans donnée manquante pour le

génotype :  $Y_n$  ↗  $\text{SNP}_n^l = 1$   $Y_n = \mu + S_n\beta + \theta^l + \epsilon_n$

↘  $\text{SNP}_n^l = 0$   $Y_n = \mu + S_n\beta + \epsilon_n$

- $S_n$  est la ligne correspondant à l'individu  $n$  dans la matrice de structure  $S$

Mêmes remarques pour la généralisation du modèle que pour le modèle simple

# Illustration

## Exemple

3 plants de maïs d'Europe et 2 plants de maïs des Etats Unis  $\Rightarrow$  2 groupes

$$S = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

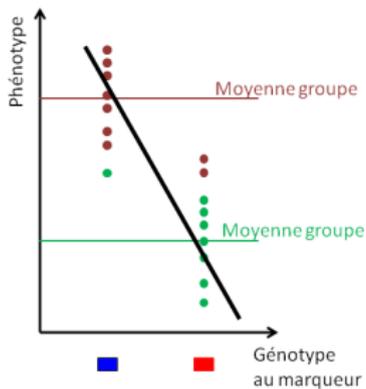
## Exemple pour $S_n\beta$

2 groupes : si  $n$  appartient au groupe 1,  $S_n\beta = \beta_1$   
 si  $n$  appartient au groupe 2,  $S_n\beta = \beta_2$

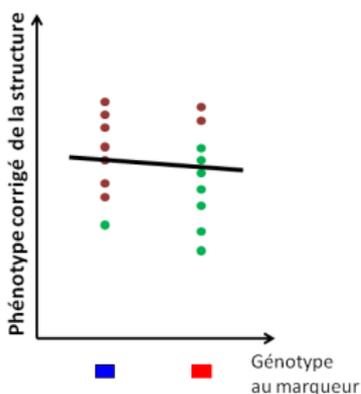
Ce modèle consiste à ne pas modéliser la moyenne du phénotype observé dans chacun des groupes de la même façon.

## Les phénotypes continus : corriger de la structure

## Illustration



Modèle simple

Modèle avec structure  
en groupe

La valeur des tests sur les marqueurs corrélés à la structure (ceux qui sont différenciés) diminue. Ils ne peuvent plus être trouvés comme associés.

Les phénotypes continus : corriger de la structure et de l'apparementement

## Modèle linéaire mixte

Yu et al., Nature Genetics, 2006

### Effet du SNP fixe + stucture fixe + covariance génétique

Cas d'individus homozygotes, sans donnée manquante pour le

génotype :  $Y_n$   $\nearrow$   $\text{SNP}_n^l = 1$   $Y_n = \mu + S_n\beta + \theta^l + G_n + \epsilon_n$

$\searrow$   $\text{SNP}_n^l = 0$   $Y_n = \mu + S_n\beta + G_n + \epsilon_n$

- $S_n$  est la ligne correspondant à l'individu  $n$  dans la matrice de structure  $S$
- $G_n$  est un valeur génétique de  $n$ . Soit  $G^t = (G_1, \dots, G_n, \dots, G_N)$ ,  $\text{Var}(G) = \sigma_G^2 \Sigma_G$

$$\text{Var}(Y) = \sigma_G^2 \Sigma_G + \sigma_\epsilon^2 \text{Id}$$

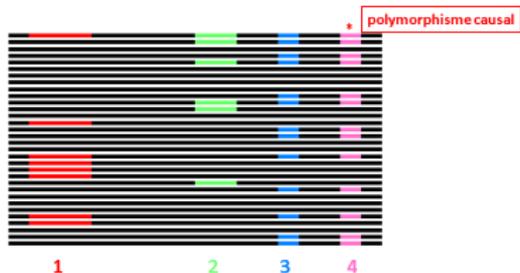
mêmes remarques pour la généralisation du modèle que pour le modèle simple

Les phénotypes continus : corriger de la structure et de l'apparentement

# Puissance et mesure $r_{VS}^2$

Mangin et al., Heredity, 2011

Pour des marqueurs bialléliques



$$r_{VS_{14}}^2$$

$$r_{VS_{24}}^2$$

$$r_{VS_{34}}^2$$

Mesure de  $r_{VS}^2$  DL

$$N_1$$

$$N_2$$

$$N_3$$

$$N_4$$

Nbre d'observations

Pour avoir la **même puissance**, qu'au locus causal

$$N_i = N_4 / r_{VS_{i4}}^2$$

avec

$$V = \sigma_G^2 \Sigma_G + \sigma_\epsilon^2 Id$$

# Estimation dans le modèle linéaire mixte

$$Y_n = \mu + S_n\beta + \text{SNP}_n^l\theta^l + G_n + \epsilon_n$$

## Deux types de paramètres à estimer

- les paramètres entrant dans la variance de  $Y$  (composantes de la variance)
- les paramètres des effets fixes

## Les composantes de la variance $\sigma_G^2$ et $\sigma_\epsilon^2$

Elles sont estimées par ML (Maximum Likelihood) ou REML (Restricted ML)

Si on parle de vraisemblance, c'est que  $Y_n$  a une loi connue. Cette loi est une Gaussienne  $\Rightarrow G_n$  et  $\epsilon_n$  sont aussi Gaussiens. C'est un postulat nécessaire pour le modèle mixte.

## Les composantes de la variance $\sigma_G^2$ et $\sigma_\epsilon^2$

Les estimateurs du maximum de vraisemblance sont des estimateurs biaisés (leur espérance n'est pas égale aux paramètres qu'il estiment).

## Le REML

Méthode qui consiste à estimer par maximum de vraisemblance mais après avoir projeté  $Y$  sur l'espace orthogonal pour  $V$  aux effets fixes

## La différence entre ML et REML

dans un cas simple, juste une question de dénominateur

Exemple :  $Y_n = \mu + \epsilon_n$  pour  $n = 1, \dots, N$

$$\hat{\sigma}_\epsilon^{2ML} = \frac{\sum (Y_n - \hat{\mu})^2}{N}$$

$$\hat{\sigma}_\epsilon^{2REML} = \frac{\sum (Y_n - \hat{\mu})^2}{N-1}$$

## Algorithme de ML et/ou REML

Il n'existe pas de formule analytique pour calculer les estimateurs du ML ou du REML.

Les algorithmes qui résolvent cette question de maximisation, atteignent le maximum par itérations successives. Ils prennent beaucoup de temps CPU, en particulier car la matrice de covariance génétique  $\Sigma_G$  doit être inversée. Et ils sont longs à converger.

Une autre approche consiste à ne pas maximiser la vraisemblance mais seulement une approximation de la vraisemblance, plus simple à maximiser. Cette approximation a cependant la propriété d'être équivalente asymptotiquement à la vraisemblance. Elle permet le "passage à l'échelle" c'est-à-dire tester des millions de SNP.

# Tester l'effet d'un SNP dans le modèle linéaire mixte

## test de Wald au marqueur $l$

Le principe est de faire comme si la variance des observations était connue et d'utiliser l'estimateur de  $\theta^l$  ainsi que sa variance, classiquement obtenus par les moindres carrés généralisés

Pour estimer les composantes de la variance on utilise le REML.

La variance "supposée connue" est  $\hat{V} = \hat{\sigma}_G^{2REML} \Sigma_G + \hat{\sigma}_\epsilon^{2REML} Id$

Le test de Wald au locus  $l$  :

$$\frac{(\hat{\theta}^l - \theta^l)^2}{\text{Var}(\hat{\theta}^l)}$$

suit asymptotiquement une loi de  $\chi^2(1)$  sous l'hypothèse

$H_0 : \{ \text{pas d'association} \}$

# Tester l'effet d'un SNP dans le modèle linéaire mixte

## test du rapport de vraisemblance (ML) au marqueur $l$

Pour faire ce test on utilise le ML jamais le REML

$$RV(l) = \frac{\text{maximum de la vraisemblance si } \theta^l = 0}{\text{maximum de la vraisemblance si un QTL est en } l}$$

$-2 \ln(RV)$  suit asymptotiquement une loi de  $\chi^2(1)$  sous l'hypothèse  $H_0 : \{ \text{pas d'association} \}$

## Mais on fait des millions de tests ....

### Tests multiples non indépendants

Que ce soit pour les phénotypes binaires ou continus, un test d'association est affectué par SNP. Chacun de ces tests sous l'hypothèse  $H_0 : \{ \text{pas d'association} \}$  est comparé un  $\chi^2$  à 1 degré de liberté. Mais d'un SNP à un autre, les tests ne sont pas indépendants.

### Bonferroni trop conservateur

⇒ On ne peut pas utiliser la correction de Bonferroni car elle conduit à un seuil de rejet très petit, donc à un test beaucoup trop conservateur, donc très très peu de puissance, donc pas de SNP trouvés associés.

## En pratique

### Bonferroni moins conservateur

⇒ Au lieu d'utiliser le nombre de SNP pour corriger la p-value du test, on calcule un équivalent du nombre de SNP indépendants.

Li & Ji, Heredity, 2005

### FDR

⇒ On peut aussi utiliser le FDR pour contrôler le taux de faux positifs.

### Permutation

⇒ On peut aussi faire des permutations, mais cela demande beaucoup de CPU. Avec de la parallélisation cela reste possible.

## Estimer la structure d'un échantillon

Voir la partie du cours de “génétique des populations”

# La matrice $\Sigma_G$

## Deux grandes méthodes

- Lorsque le pedigree est connu
- A l'aide des marqueurs

Lorsque l'on dispose de nombreux marqueurs, la deuxième méthode donne des résultats plus précis, en effet elle est capable de voir des différences entre deux plein-frères. On dit qu'elle "suit les aléas de la méiose".

Calculer  $\Sigma_G$ , souvent appelée kinship

# Lorsque le pedigree est connu

## Le coefficient d'apparentement (coancestry)

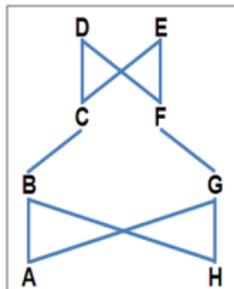
Weir et al., Nature Rev Genet, 2006

$$\rho_{ij} = \sum_a (1 - F_a) (1/2)^{n_a}$$

$a$  ancêtre commun de  $i$  et  $j$

$n_a$  le nbre d'individus du pedigree sur le chemin le plus court de  $i$  à  $j$  en passant par  $a$

$F_a$  coefficient de consanguinité de  $a$



Exemple

Si  $D$  et  $E$  non consanguins,  $\rho_{AF} = 2(1/2)^5$

Calculer  $\Sigma_G$ , souvent appelée kinship

# Lorsque la covariance génétique est estimée avec les marqueurs

## AIS

La première proposition consiste à compter le nombre espéré d'allèles identiques entre deux individus, c'est le coefficient AIS (alike in state) aussi appelée IBS (identity by state).

## C'est quoi le nombre espéré ?

		AA	AT	TT
Soit un SNP A/T	AA	2	1	0
	AT	1	$1/2(0+2)=1$	1
	TT	0	1	2

Calculer  $\Sigma_G$ , souvent appelée kinship

## Exemple

Deux individus,  $ind_1$  et  $ind_2$ , génotypés pour 10 marqueurs  $l$ .

Codage du génotype : 0 = homozygote pour l'allèle mutante, 2 = homozygote pour l'allèle du référent, 1 = hétérozygote

											somme	
$M_{ind_1} = ($	0,	0,	1,	1,	2,	1,	0,	0,	1,	2	)	
$M_{ind_2} = ($	0,	2,	1,	0,	2,	0,	0,	0,	1,	2	)	
$AIS^l = ($	2,	0,	1,	1,	2,	1,	2,	2,	1,	2	)	14

$$AIS_{(ind_1, ind_2)} = 14 / (2 \times 10) = 0.7$$

## Propriété de l'AIS

$$AIS_{(ind_1, ind_2)} = \frac{\sum_{l=1}^{\mathcal{L}} \left( M_{ind_1}^l M_{ind_2}^l + (2 - M_{ind_1}^l)(2 - M_{ind_2}^l) \right)}{4\mathcal{L}}$$

$$AIS_{(ind_1, ind_1)} = 1, \quad 0 \leq AIS_{(ind_1, ind_2)} \leq 1$$

Calculer  $\Sigma_G$ , souvent appelée kinship

## Avec les marqueurs, de nombreuses autres propositions

### Kinship de type "VanRanden"

$$K_{(ind_1, ind_2)} = \frac{\sum_{l=1}^{\mathcal{L}} M_{ind_1}^l M_{ind_2}^l}{C} = \frac{M_{ind_1}^t M_{ind_2}}{C}$$

où  $M_{ind_1}$  peut avoir été au préalable centré et/ou réduit par SNP pour l'échantillon d'étude.

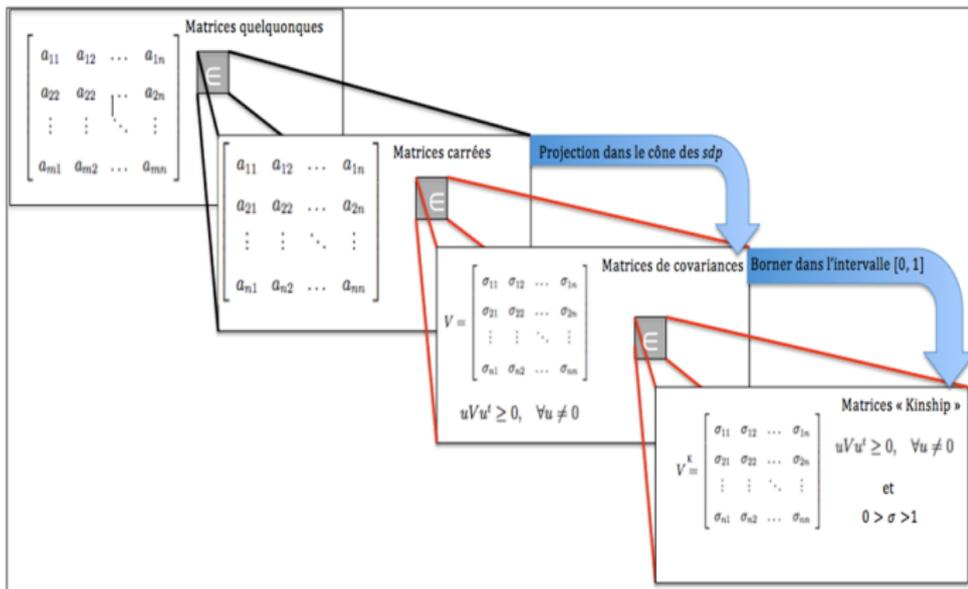
La constante  $C$  n'a aucune influence sur les résultats, elle peut être fixée à une valeur arbitraire.

Calculer  $\Sigma_G$ , souvent appelée kinship

## ... ou encore

Estimateur	Principe	Plusieurs Populations	Matrice Kinship
AIS	Proba (IBS)	✗	😊
BNO	Proba(IBD)=Proba (IBS)- correction	✓	😞
WAIS	Proba(IBD)=Proba (IBS)- correction	✓	😊
LOI	Proba(IBD)=Corrélation des fréquences alléliques	✗	😞
MIL	Max de vraisemblance	✗	😊

## Caractéristique d'une "bonne" kinship



Une matrice *semi-définie positive* (*sdp*) dont les éléments sont compris entre 0 et 1

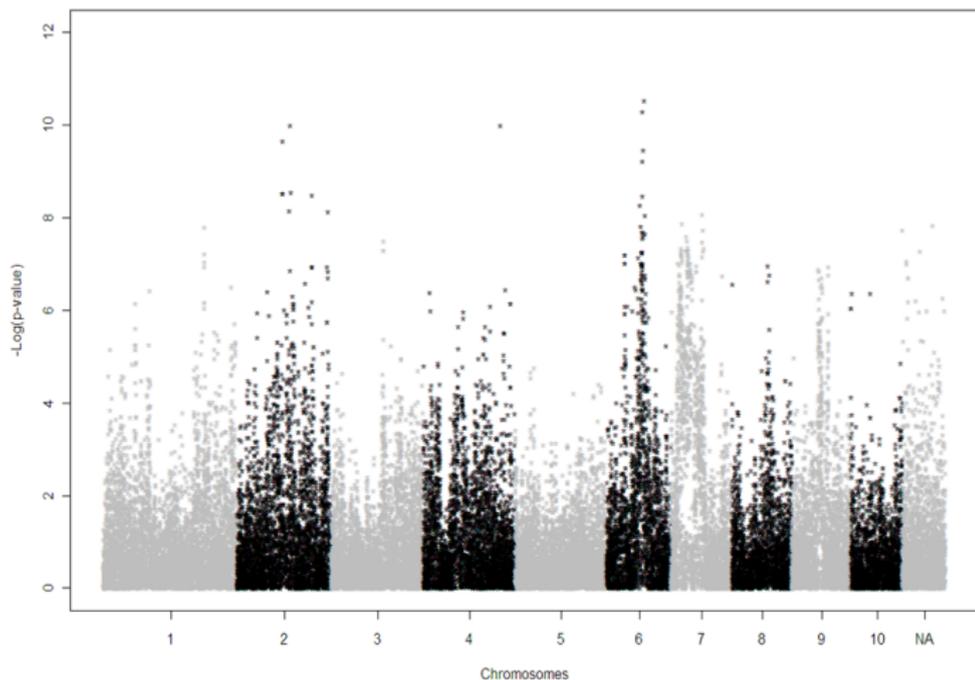
# Les logiciels

- ASREML, générique pour les modèles mixtes, dans R, maximise la vraisemblance, test du rapport de vraisemblance
- EMMA et suivants, spécifique de “association mapping”, dans R, maximise une approximation de la vraisemblance, test de Wald
- Tassel, spécifique de “association mapping”, java, propose la maximisation de la vraisemblance et de son approximation ainsi que le test du rapport de vraisemblance et celui de Wald
- Plink, spécifique de “association mapping”, pas de modèle mixte, plutôt spécifique des applications en génétique humaine

## Illustration

# Résultats du test d'association pour un modèle

## Manhattan plot



## Illustration

# Résultats du test pour plusieurs modèles

Différentes matrices de structure, différentes matrices de covariance génétique

