

# Cartographie génétique par analyse de liaison

INRA, Brigitte Mangin

Septembre 2016

## 1 Introduction

## 2 Modélisation

- Les phénotypes continus

## 3 Un QTL

- Le test du rapport de vraisemblance
- Quel seuil pour la détection d'un QTL ?

## 4 Pour aller plus loin

- D'autres modélisations
- Pedigrees complexes
- Les logiciels

# Introduction

## Qu'est ce que l'analyse de liaison ?

⇒ C'est l'étude de la **transmission** allélique parmi des individus **apparentés**

## Que recherche-t'on ?

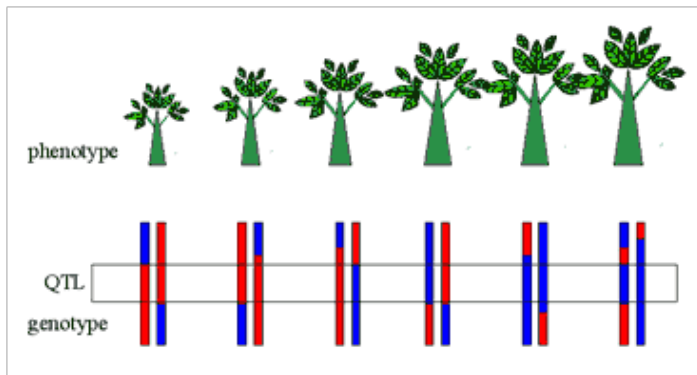
⇒ Des **gènes** causaux responsables d'une maladie, caractère binaire

ou

⇒ Des **QTLs**, c'est-à-dire des loci du génome qui expliquent la variabilité d'un caractère quantitatif.

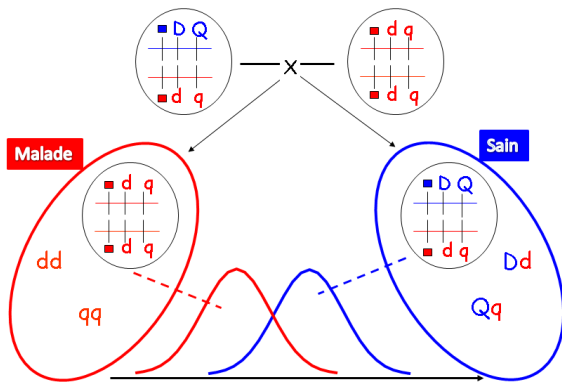
## Que cherche t-on à faire ?

⇒ lier le caractère binaire ou quantitatif (phénotype) à un endroit du génome (génotype au gène ou au QTL)



## Comment s'y prend-t'on ?

⇒ suit la transmission allélique à l'aide de marqueurs moléculaires



# Modélisation

## La modélisation pour quoi faire ?

⇒ “décrire” la relation entre le phénotype et le génotype

## En terme de probabilité “pénétrance”, phénotype binaire

$$f_{dd} = \text{Prob}(\text{malade} \mid dd)$$

$$f_{Dd} = \text{Prob}(\text{malade} \mid Dd)$$

$$f_{DD} = \text{Prob}(\text{malade} \mid DD)$$

## En terme de densité de probabilité $\mathcal{L}$ , phénotype continu $Y$

$$f_{qq} = \mathcal{L}(Y|qq)$$

$$f_{Qq} = \mathcal{L}(Y|Qq)$$

$$f_{QQ} = \mathcal{L}(Y|QQ)$$

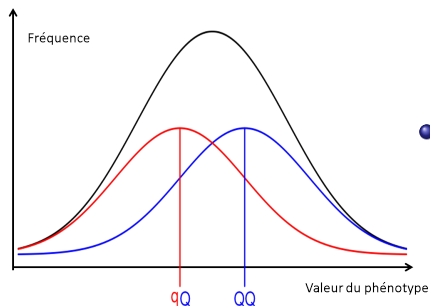
# Modèle QTL simple dans un pedigree simple

## Nous allons nous restreindre

- à une modélisation simple qui postule la présence d'un unique QTL
- à une descendance simple, le back-cross
- dans le cadre de la génétique mendélienne
  - back-cross :  $1/2$  qq ,  $1/2$  Qq ou  $1/2$  qQ ,  $1/2$  QQ

# Back-cross ou rétrocroisement

La distribution du phénotype (courbe noire)



- est un mélange 1/2:1/2 entre les distributions du phénotype pour chaque génotype
- les distributions du phénotype pour chaque génotype ont la même variance, mais pas la même moyenne

- et sont Gaussiennes



## Les phénotypes continus

## Données

$n$  observations quantitatives     $M$  marqueurs

$$Y_n$$

$$n = 1, \dots, N$$

$$M_n^m$$

$$m = 1, \dots, M$$

## Paramètres du modèle

$\sigma^2$     variance résiduelle

$\mu_{qQ}$     effet moyen du génotype  
qQ au QTL

$\mu_{QQ}$     effet moyen du génotype  
QQ au QTL

## Modèle de mélange

$$Y_n = \begin{cases} \mu_{qQ} + \epsilon_n & \text{si } n \text{ a le génotype } qQ \text{ au QTL} \\ \mu_{QQ} + \epsilon_n & \text{si } n \text{ a le génotype } QQ \text{ au QTL} \end{cases}$$

# Les paramètres génétiques

## Back-cross

fond génétique  
effet de substitution allélique

$$\mu = \frac{\mu_{QQ} + \mu_{Qq}}{2}$$
$$\alpha = \mu_{QQ} - \mu_{Qq}$$

ou

$$\alpha = \mu_{qq} - \mu_{Qq}$$

## et les marqueurs à quoi y servent ?

- à **connaître** le génotype au QTL
  - lorsque le QTL et le marqueur sont au même locus
- à **inférer** le génotype au QTL, “pseudo-marqueur”
  - en calculant la probabilité que l'individu  $n$  soit de génotype  $qQ$  ou  $QQ$  sachant ses marqueurs et la carte génétique
- ou à **imputer** le génotype au QTL, “pseudo-marquage”
  - en tirant aléatoirement le génotype au QTL avec les probabilités que l'individu  $n$  soit de génotype  $qQ$  ou  $QQ$  sachant ses marqueurs et la carte génétique

# Comment détecter un QTL ?

## Détecter un QTL c'est :

- prendre un risque et affirmer qu'il y a un QTL ⇒ **c'est donc une procédure de test**
  - Quelle est l'hypothèse nulle ?
  - Quelle est l'hypothèse alternative ?
  - Quelle statistique de test ?
- le localiser sur le génome en répétant le test tout le long du génome ⇒ **c'est donc une procédure de tests multiples**
  - Quel est le seuil de rejet de l'hypothèse nulle ?
- estimer les paramètres du modèle

# Tester l'effet du QTL

Hypothèse nulle  $H_0 = \{ \text{tous les génotypes au QTL ont la même moyenne} \}$

Hypothèse alternative  $H_1 = \{ \text{l'un au moins des génotypes au QTL a une moyenne différente des autres} \}$

## Pour le Back-cross

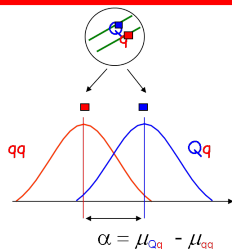
$$H_0 = \{ \mu_{QQ} = \mu_{Qq} \} \text{ versus } H_1 = \{ \mu_{QQ} \neq \mu_{Qq} \}$$

ou

$$H_0 = \{ \alpha = 0 \} \text{ versus } H_1 = \{ \alpha \neq 0 \}$$

# Le cas idéal : un marqueur sur le locus du QTL

## Modèle linéaire : comparaison de moyennes ou ANOVA

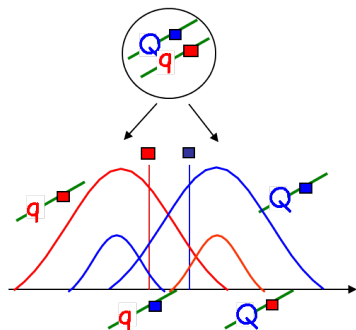

 $Y_n$ 

$$g = 1 \quad Y_{1k} = \mu + \frac{\alpha}{2} + \epsilon_{1k}$$

$$g = 2 \quad Y_{2k} = \mu - \frac{\alpha}{2} + \epsilon_{2k}$$

Source de variation	ddl	SCE	CM = $\frac{SCE}{ddl}$	F
QTL/marqueur	1	$SCE_M = \sum_g n_g (Y_{g.} - Y_{..})^2$	CM <sub>M</sub>	$\frac{CM_M}{CM_R}$
résiduelle	N - 1	$SCE_R = \sum_{gk} (Y_{gk} - Y_{g.})^2$	CM <sub>R</sub>	

# Le cas non idéal



⇒ c'est la méthode appelée "l'interval mapping"

- Se contenter de faire la comparaison de moyennes entre les ■ et les ■ c'est perdre de la puissance.
- Les marqueurs flanquants le locus QTL vont être utilisés pour inférer le génotype aux QTL

# “l'interval mapping”

## au locus $l$

Le test du rapport du maximum de vraisemblance au locus  $l$

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

*Lander et Botstein, Genetics, 1989*

## La courbe du $LOD$

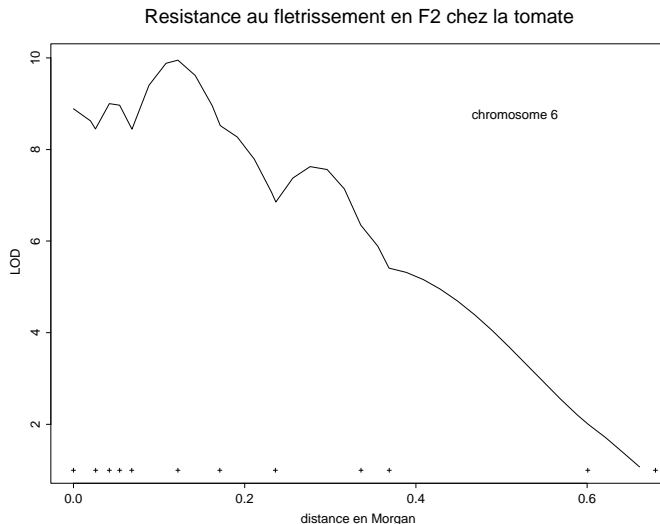
C'est la courbe obtenue en faisant varier  $l$  le long de la carte génétique et en transformant  $RV(l)$  par la fonction  $-\log_{10}()$

## Le $LOD$ test

C'est le locus où cette courbe est maximum



# Un exemple de courbe de *LOD*



Quel seuil pour la détection d'un QTL ?

## Seuil de rejet du *LOD* test

Loi "asymptotique" sous  $H_0 = \{\text{pas de QTL en } l\}$

Au locus  $l$ , pour le back-cross, le  $LOD(l)/0.217$  suit une loi de  $\chi^2(1)$

donc si  $LOD(l) > 3.84 \times 0.217$  on rejette l'absence de QTL en  $l$  au risque 5%

**MAIS on fait un test à chaque locus  $l$**

... la question difficile des tests multiples non indépendants

Le seuil dépend

du risque global choisi  
 du nombre de groupes de liaison  
 du nombre de marqueurs  
 et de leur localisation  
 du type de descendance

## Quel seuil pour la détection d'un QTL ?

⇒ formule analytique (étude asymptotique)

*Rebaï et al, Genetics, 1994*

⇒ simulation | par Monte-Carlo sous  $H_0$   
par permutation des  $Y_n$

*Churchill et Doerge, Genetics, 1994*

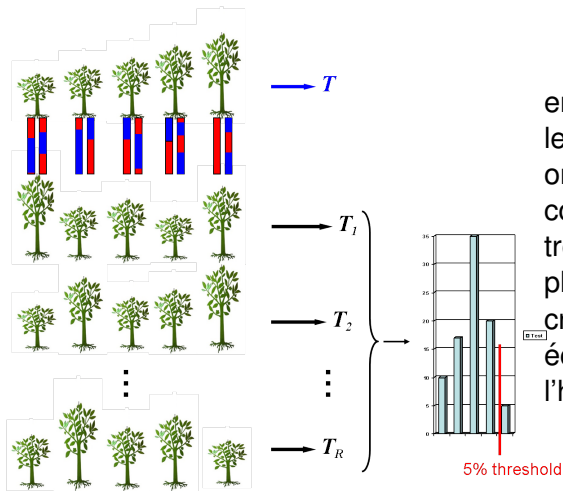
Quel seuil pour la détection d'un QTL ?

## Seuil pour le *LOD* test par simulations

		longueur du groupe de liaison : 1 M		
		intervalle entre les marqueurs		
type	risque	20 cM	10 cM	$\leq 1$ cM
BC	5 %	1.5	1.6	1.8
BC	1 %	2.2	2.3	2.6
F <sub>2</sub>	5 %	2.2	2.3	2.6
F <sub>2</sub>	1 %	2.9	3.1	3.4
		longueur du groupe de liaison : 2 M		
BC	5 %	1.8	1.9	2.1
BC	1 %	2.5	2.6	3.0
F <sub>2</sub>	5 %	2.4	2.6	3.0
F <sub>2</sub>	1%	3.2	3.4	3.8

Quel seuil pour la détection d'un QTL ?

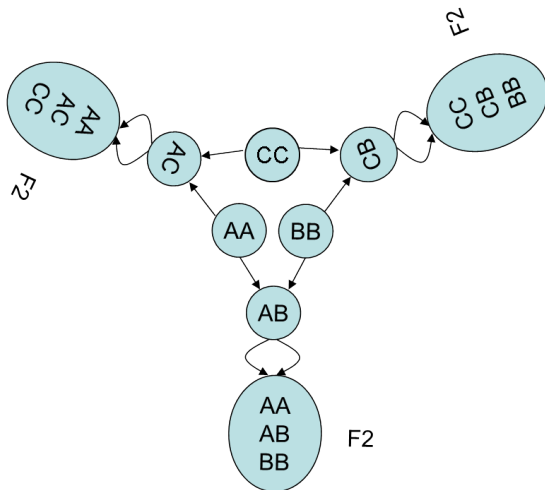
# Principe des permutations



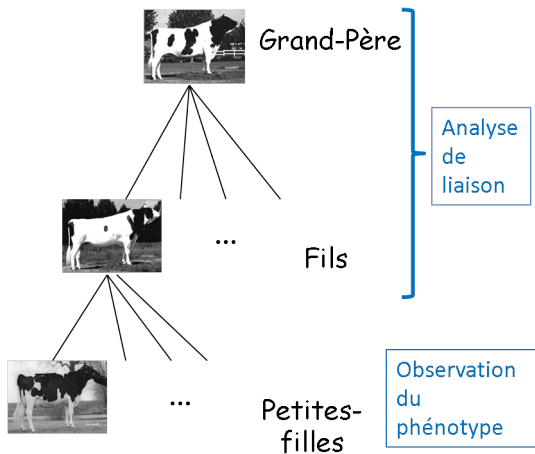
en permutant les phénotypes, on casse la corrélation entre génotype et phénotype, on crée donc un échantillon sous l'hypothèse nulle

- Plusieurs QTL
- Effet du QTL aléatoire (animaux)
- Effet polygénique aléatoire (animaux) ou effet fond génétique fixe (plantes)
- Effet d'épistasie (interaction) entre QTLs
- Effet d'épistasie (interaction) entre un QTL et le fond génétique
- QTL en pléiotropie
- ....

# Diallèle de lignées homozygotes

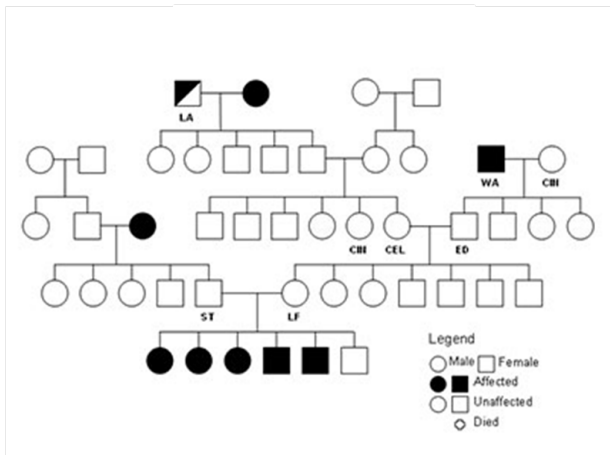


# Familles de demi-frères





# Pedigree humain



# Pour les QTL

- l'ancêtre : MAPMAKER/QTL
- le clé en main : QTL carthographe
- le payant : mapQTL
- le futur : R/qtl
- “les spécialisés”
  - pour les pedigrees et modèles animaux : QTLMap
  - pour plusieurs descendance : MCQTL