

# Calcul des seuils pour la détection de QTL

INRA - Toulouse

Mars 2011

# Objectifs

## Calcul des seuils pour la détection de QTL

### Objectifs

#### Calcul de seuils

Le modèle  
La Statistique  
Covariance  
Seuils  
Les Simulations  
Données réelles

#### Estimation des effets

Les estimations

#### Avantages de la méthode

- Contrôle de l'erreur de type I pour la détection de QTL.
  - Calcul des seuils en déterminant la distribution du processus sous  $H_0$ .
  - Estimation des effets en utilisant la distribution du processus sous  $H_1$ .
  - Lissage du processus.

# Le modèle

Calcul des  
seuils pour la  
détection de  
QTL

$$y_{ij}(t) = \mu(t) + m_i(t) + \epsilon_{ij}(t)$$

$y_{ij}(t)$  : Phénotype de l'individu  $j$  avec le génotype  $i$   
pour un SNP  $t$

$i = 1, 2, 3$ ;  $j = 1, \dots, n$ ;  $t = 1, \dots, nSNP$

$m_i(t)$  : Effet du génotype  $i$  pour un SNP  $t$ .

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

On souhaite tester :

$$H_0 : m_i(t) = 0 \forall t$$

vs.

$$H_1 : \exists \text{ au moins un } t \text{ tel que } m_i(t) \neq 0$$

Objectifs

Calcul de  
seuils

Le modèle  
La Statistique  
Covariance  
Seuils  
Les Simulations  
Données réelles

Estimation  
des effets  
Les estimations

Avantages de  
la méthode

# La Statistique

## Calcul des seuils pour la détection de QTL

### Objectifs

#### Calcul de seuils

Le modèle

#### La Statistique

Covariance

Seuils

Les Simulations

Données réelles

#### Estimation des effets

Les estimations

#### Avantages de la méthode

$$Z(t) = \begin{bmatrix} \frac{y_{1\cdot} - \mu}{\sigma/\sqrt{n_1}} \\ \frac{y_{2\cdot} - \mu}{\sigma/\sqrt{n_2}} \\ \frac{y_{3\cdot} - \mu}{\sigma/\sqrt{n_3}} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, I \right) \text{ sous } H_0$$

$$S(t) = \|Z(t)\|^2 \sim \chi^2_{(3)}$$

# Covariance entre les $S(t)$

Calcul des  
seuils pour la  
détection de  
QTL

Objectifs

Calcul de  
seuils

Le modèle  
La Statistique  
Covariance

Seuils  
Les Simulations  
Données réelles

Estimation  
des effets  
Les estimations

Avantages de  
la méthode

$$\text{cov} (Z(t), Z(\tilde{t})) = v(t, \tilde{t})_{3 \times 3}$$

$$v_{kl}(t, \tilde{t}) = \frac{n_{kl}(t, \tilde{t})}{\sqrt{n_k(t)} \sqrt{n_l(\tilde{t})}}$$

avec :

$n_{kl}(t, \tilde{t})$  : nombre d'individus avec le génotype  $k$  en  $t$  et  $l$  en  $\tilde{t}$

$n_k(t)$  : nombre d'individus avec le génotype  $k$  en  $t$

$n_l(\tilde{t})$  : nombre d'individus avec le génotype  $l$  en  $\tilde{t}$

- **Bonferroni** :  $\frac{\alpha}{nSNP}$  ne prend pas en compte la corrélation entre les SNP

- **Propositions de seuils :**

- **Loi du sup** : (Muller et al. 2010)

- 10000 simulations d'une loi  $N(0, V)$
- pour chaque simulation on prend le

$$\max (\|S(t_1)\|^2 + \|S(t_2)\|^2 + \dots + \|S(t_{nSNP})\|^2)$$

- on trace l'histogramme de ces 10000 valeurs pour connaître la distribution du sup : le seuil est le quantile à 95 % de cette distribution.

- **Integrale** :

- La même démarche que pour le sup mais dans ce cas on prend la somme des valeurs à la place du max.

# Les simulations

## Calcul des seuils pour la détection de QTL

### Objectifs

#### Calcul de seuils

Le modèle  
La Statistique  
Covariance  
Seuils

#### **Les Simulations** Données réelles

#### Estimation des effets

Les estimations

#### Avantages de la méthode

- 1000 simultaions
- 1000 individus
- 23 SNP
- 1 et 3 QTLs

# Les simulations

Calcul des seuils pour la détection de QTL

Objectifs

Calcul de seuils

Le modèle

La Statistique

Covariance

Seuils

**Les Simulations**

Données réelles

Estimation des effets

Les estimations

Avantages de la méthode

	Position	Sup	Int	Bonf
0 QTL	2	58	57	34
1 QTL	2	221	92	146
	2	908	454	866
	2	1000	999	1000
	12	275	191	216
	12	867	693	820
3 QTL	12	1000	999	1000
	2,12,22	1000	1000	1000
	2,12,22	953	926	818
	10,12,22	924	886	888

# Données réelles

Calcul des seuils pour la détection de QTL

- Données GENEQUIN (chromosome 12)
- 525 individus
- 1794 SNP

Objectifs

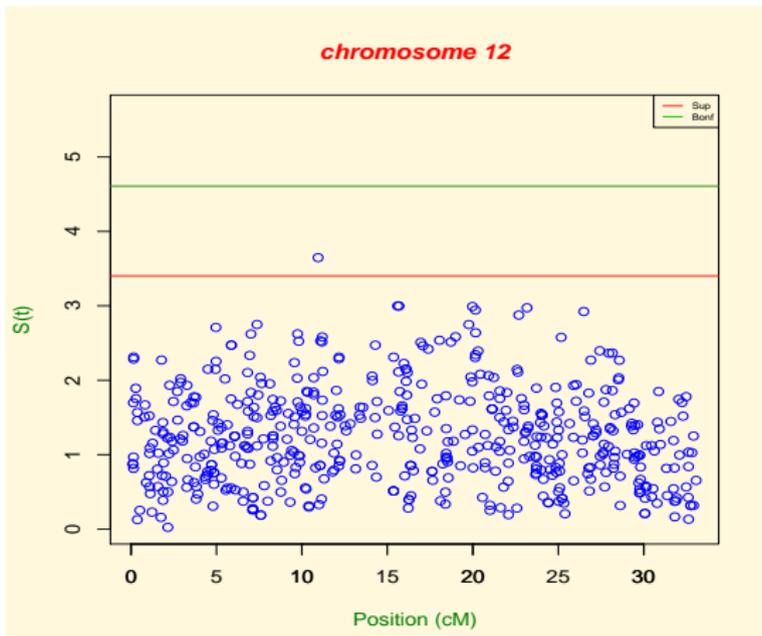
Calcul de seuils

Le modèle  
La Statistique  
Covariance  
Seuils  
Les Simulations  
Données réelles

Estimation des effets

Les estimations

Avantages de la méthode



# Les estimations

Calcul des  
seuils pour la  
détection de  
QTL

Objectifs

Calcul de  
seuils

Le modèle  
La Statistique  
Covariance  
Seuils  
Les Simulations  
Données réelles

Estimation  
des effets

Les estimations

Avantages de  
la méthode

$H_1$  : existe un QTL en  $t_l$

$$Z(t_l) = \begin{bmatrix} \frac{y_{1\cdot}(t_l) - \mu}{\sigma / \sqrt{n_1(t_l)}} \\ \frac{y_{2\cdot}(t_l) - \mu}{\sigma / \sqrt{n_2(t_l)}} \\ \frac{y_{3\cdot}(t_l) - \mu}{\sigma / \sqrt{n_3(t_l)}} \end{bmatrix} \sim N \left( \begin{bmatrix} \frac{m_1(t_l)}{\sigma / \sqrt{n_1(t_l)}} \\ \frac{m_2(t_l)}{\sigma / \sqrt{n_2(t_l)}} \\ \frac{m_3(t_l)}{\sigma / \sqrt{n_3(t_l)}} \end{bmatrix}, I \right) \sim N(\vec{m}, I)$$

$$Z(t) \sim N(v(t, t_l) \vec{m}(t_l), I)$$

# Les estimations

$H_1$  : existe un QTL en  $t_1, t_2, \dots, t_{nSNP}$

$$Z(t_l) = \begin{bmatrix} \frac{y_{1\cdot}(t_l) - \mu}{\sigma / \sqrt{n_1(t_l)}} \\ \frac{y_{2\cdot}(t_l) - \mu}{\sigma / \sqrt{n_2(t_l)}} \\ \frac{y_{3\cdot}(t_l) - \mu}{\sigma / \sqrt{n_3(t_l)}} \end{bmatrix} \sim N \left( \sum_{k=1}^{nSNP} v(t_l, t_k) m(\vec{t}_k), I \right)$$

$Z \sim N(V\vec{m}, V)$  et  $V^{-\frac{1}{2}}Z \sim N(V^{\frac{1}{2}}\vec{m}, I)$

On veut estimer  $\vec{m}$  tel que :

$$\left\| V^{-\frac{1}{2}}Z - V^{\frac{1}{2}}\vec{m} \right\|^2 \text{ soit minima}$$

avec la contrainte :  $A\vec{m} = 0$

# Les estimations

Calcul des seuils pour la détection de QTL

Objectifs

Calcul de seuils

Le modèle  
La Statistique  
Covariance  
Seuils  
Les Simulations  
Données réelles

Estimation des effets

Les estimations

Avantages de la méthode

avec

$$A = \begin{bmatrix} \frac{\sigma}{\sqrt{n_1(1)}} & \frac{\sigma}{\sqrt{n_2(1)}} & \frac{\sigma}{\sqrt{n_3(1)}} & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma}{\sqrt{n_1(2)}} & \frac{\sigma}{\sqrt{n_2(2)}} & \frac{\sigma}{\sqrt{n_3(2)}} & \dots & 0 & 0 & 0 \\ & & \vdots & & \ddots & & & & & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & \frac{\sigma}{\sqrt{n_1(n_{snp})}} & \frac{\sigma}{\sqrt{n_2(n)}} & \frac{\sigma}{\sqrt{n_3(n)}} \end{bmatrix}$$

ce qui conduit à :

$$\hat{\vec{m}} = V^{-1} \left( I - A^T (AV^{-1}A^T)^{-1} AV^{-1} \right) Z$$

# Les estimations

Calcul des seuils pour la détection de QTL

Objectifs

Calcul de seuils

Le modèle

La Statistique

Covariance

Seuils

Les Simulations

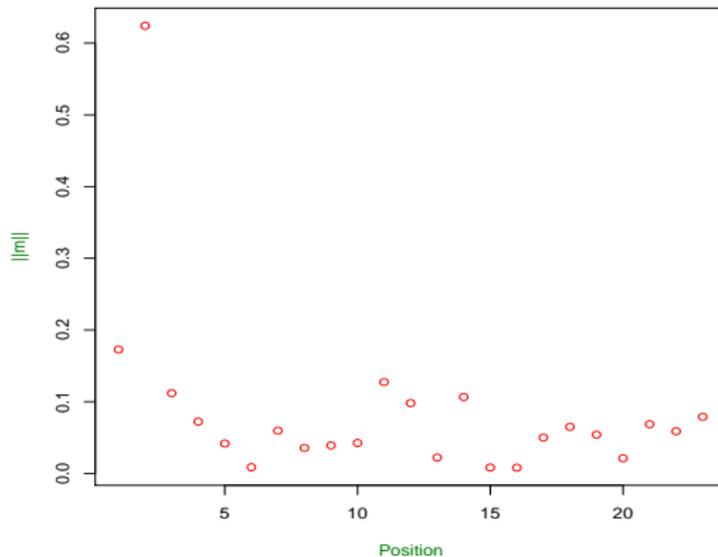
Données réelles

Estimation des effets

**Les estimations**

Avantages de la méthode

*Effets QTL*



# Avantages

## Calcul des seuils pour la détection de QTL

### Objectifs

#### Calcul de seuils

- Le modèle
- La Statistique
- Covariance
- Seuils
- Les Simulations
- Données réelles

#### Estimation des effets

- Les estimations

#### Avantages de la méthode

- Obtention d'un système bien déterminé de  $3 nSNP$  équations et  $3 nSNP$  inconnus
- Les interactions n'ont pas d'influence sur l'estimation des effets