Cartographie de gènes à caractères quantitatifs par déséquilibre de liaison

Simon Boitard

INRA Unité BIAT (Biométrie et Intelligence Artificielle Toulouse)
MIA (Mathématiques et Informatique Appliquées) et Génétique Animale

Université Paul Sabatier Toulouse III

Directeurs : Brigitte Mangin et Jean-Marc Azaïs

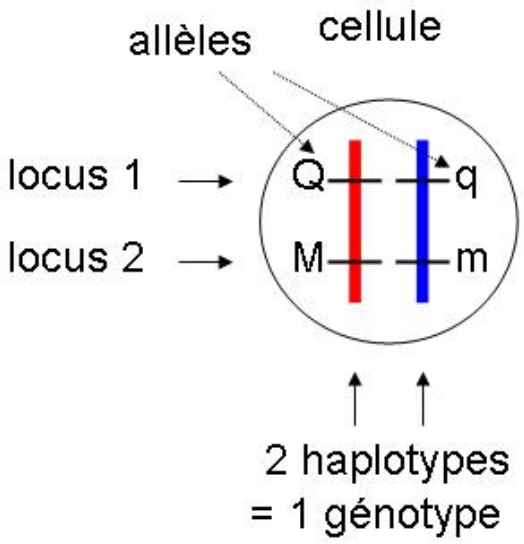
Contexte

- Enjeu : Identifier les gènes ayant une influence sur un caractère quantitatif d'intérêt (pourcentage de gras d'une viande, taux de collagène ...).
  $\longrightarrow$ **Gènes à caractères quantitatifs (QTL)**.
- Objectif : développer des *méthodes statistiques* pour la cartographie de gènes.
- Cartographie par **analyse de liaison** ou **déséquilibre de liaison**.

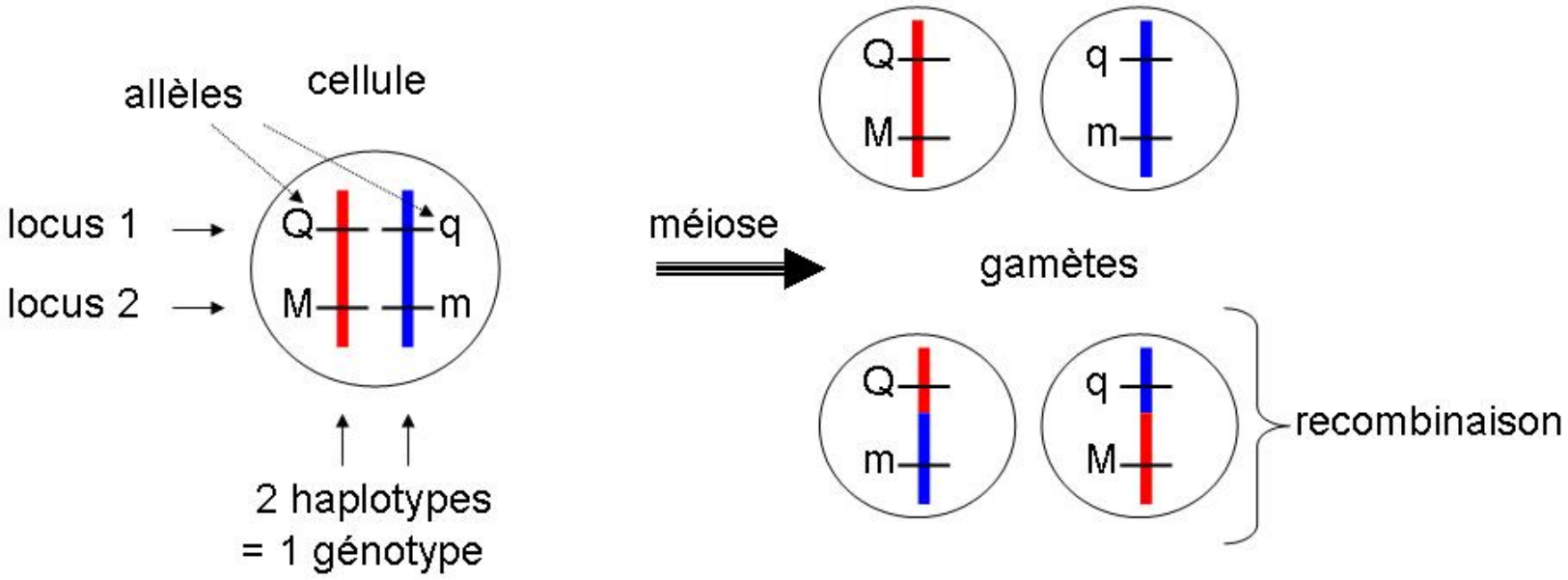
Plan de l'exposé

1. Définitions et modèles
   - Modélisation du génome
   - Cartographie de QTL
2. Cartographie par déséquilibre de liaison
   - Principe
   - Etat de l'art
   - Approche fréquentiste
3. Cartographie de QTL à l'aide de marqueurs flanquants : HAPim
4. Loi des fréquences d'haplotypes sous un modèle à deux loci
5. Bilan et perspectives

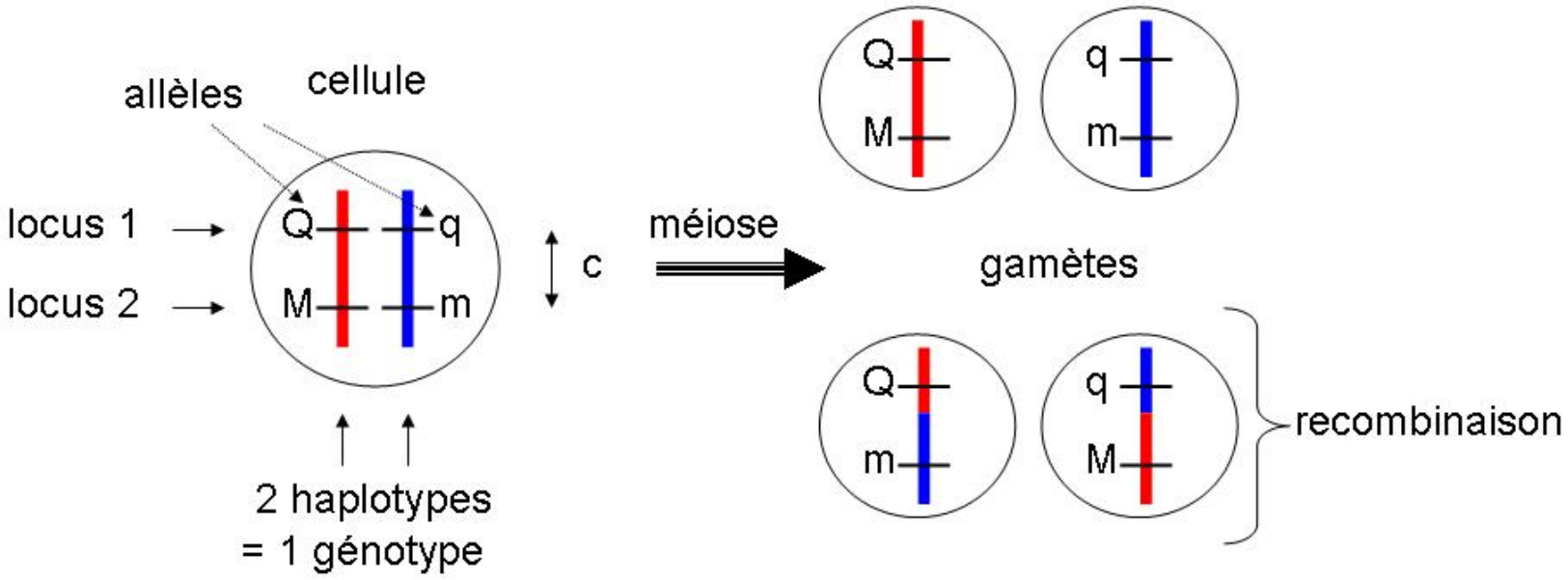
Définitions



Définitions



Définitions



c **taux de recombinaison** entre les loci

Gène à caractère quantitatif (QTL)

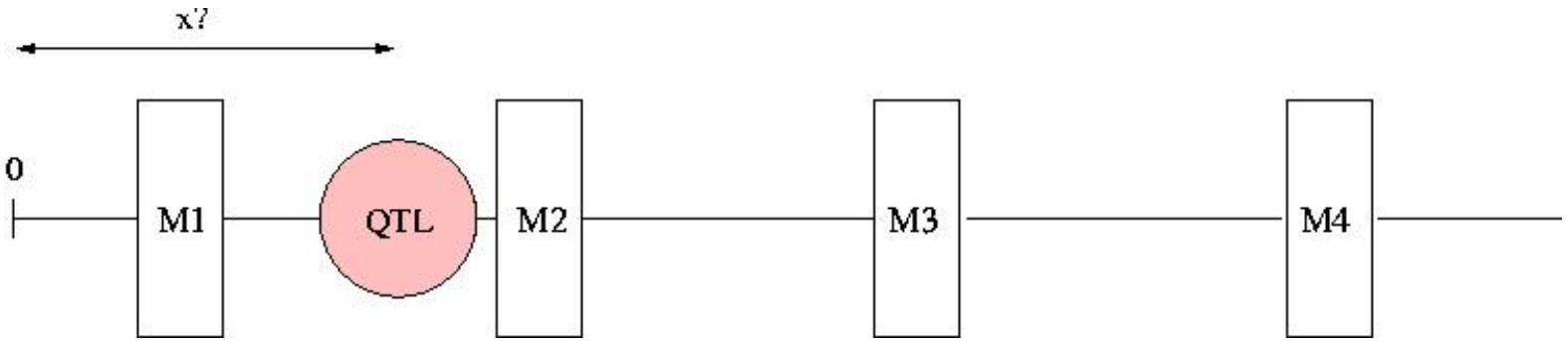
- Modèle :

$$Z_n \rightsquigarrow \psi_{g_n}$$

  - Z_n **phénotype** de l'individu n pour un caractère quantitatif.
  - g_n **génotype** de l'individu n au QTL.
- Hyp : deux allèles possibles au QTL (Q et q).

$$\psi_{Q/Q} \neq \psi_{Q/q} \neq \psi_{q/q}$$

Cartographie génétique



- Un caractère quantitatif.
- Un QTL de position inconnue x dans un segment de chromosome donné.
- **Carte génétique** de ce segment : plusieurs loci de position connue (**marqueurs**).
- Objectif : estimer x.

Type de données

- Echantillon de N_s individus de la même population.
- Pour chaque individu n :
  - Phénotype z_n.
  - Paire d'haplotypes pour les marqueurs de la carte : (h_n^1/h_n^2).
- $\mathcal{D} = \{(z_n, h_n^1/h_n^2), n = 1, \ldots, N_s\}$.

Comment estimer x à partir de $\mathcal{D}$?

1 Définitions et modèles
- Modélisation du génome
- Cartographie de QTL

2 Cartographie par déséquilibre de liaison
- Principe
- Etat de l'art
- Approche fréquentiste

3 Cartographie de QTL à l'aide de marqueurs flanquants : HAPim

4 Loi des fréquences d'haplotypes sous un modèle à deux loci

5 Bilan et perspectives

Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

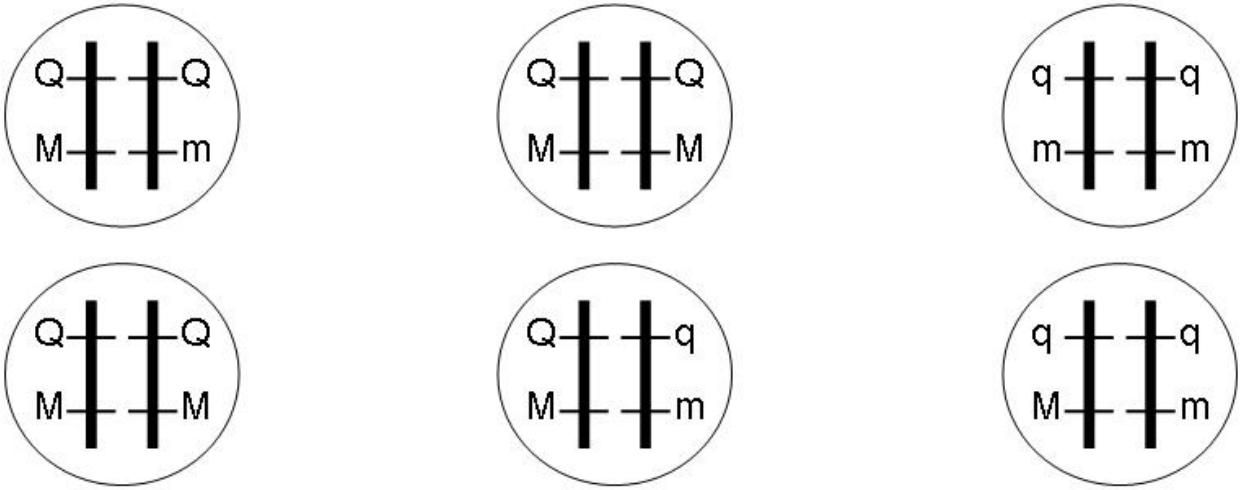
Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.}\Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.}\Pi_{.,i_2}$$

$$D_{Q,M} = \Pi_{Q,M} - \Pi_{Q,.}\Pi_{.,M}$$



Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

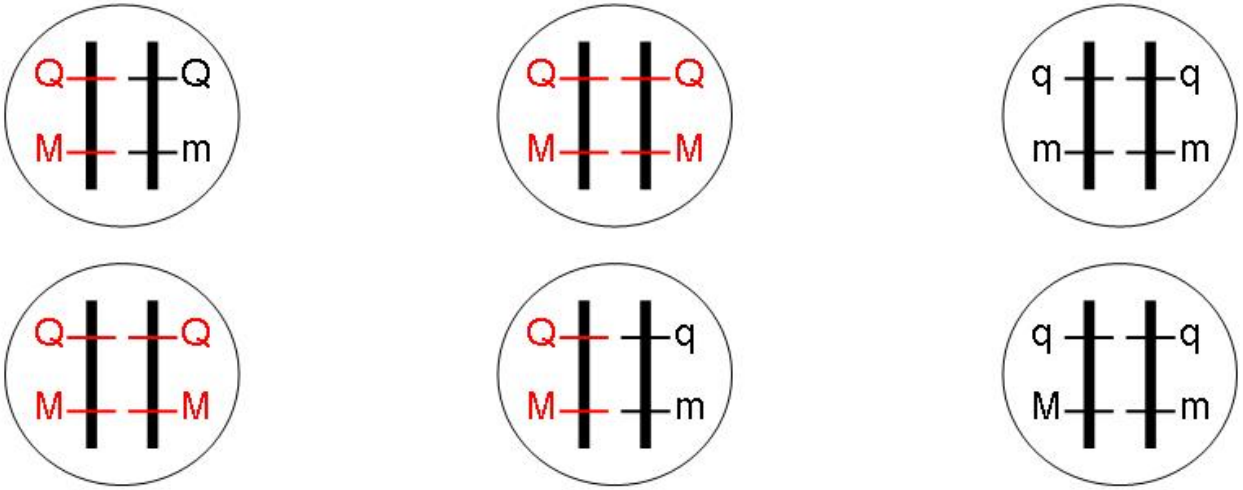
Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

$$D_{Q,M} = 6/12 - \Pi_{Q,.} \Pi_{.,M}$$



Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

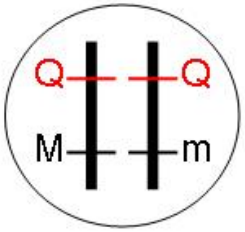
Equilibre de liaison :

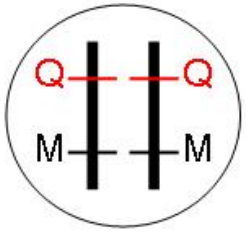
$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

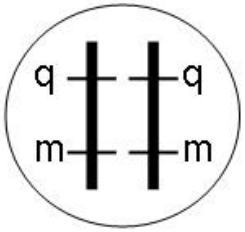
Déséquilibre de liaison (LD) :

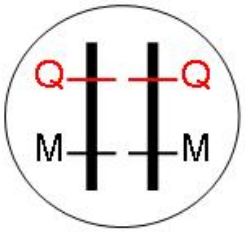
$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

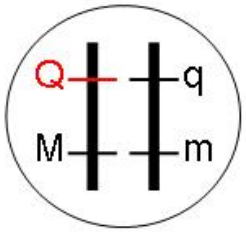
$$D_{Q,M} = 6/12 - 7/12 \, \Pi_{.,M}$$

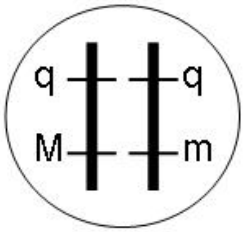












Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

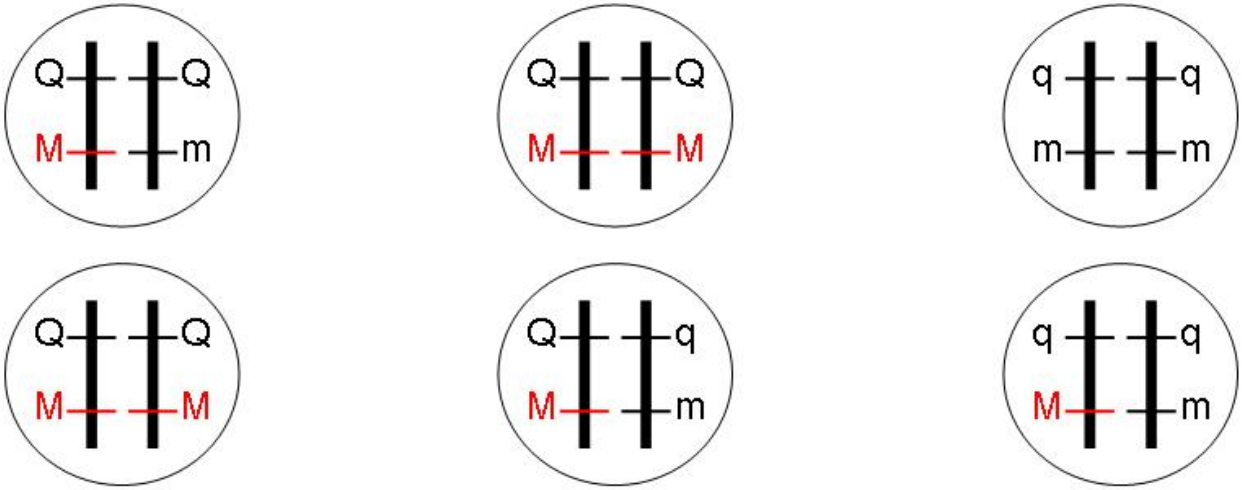
Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

$D_{Q,M} =$ 6/12 − 7/12 7/12



Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

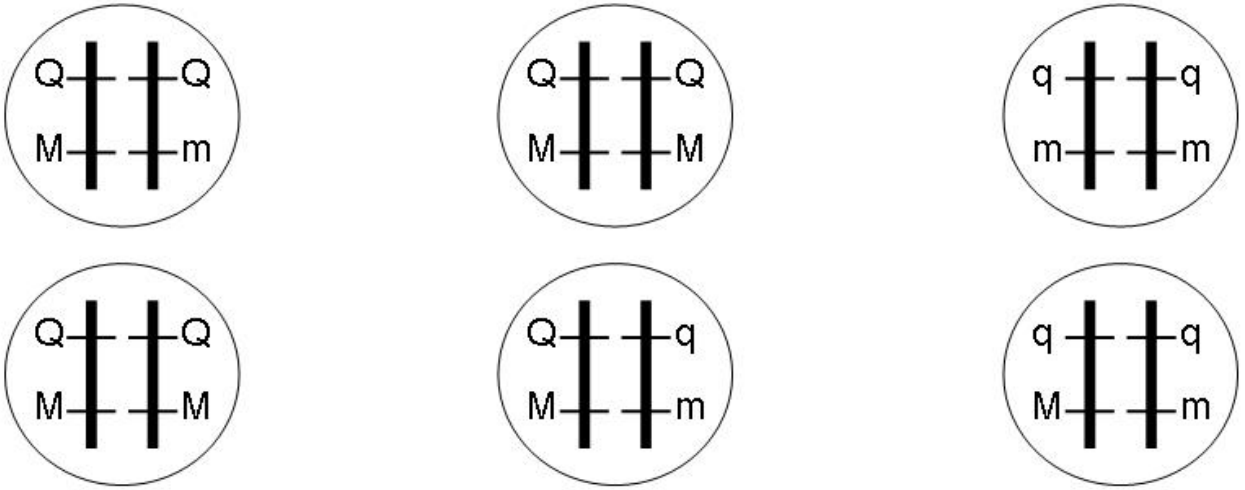
Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

33/124 = 6/12 − 7/12 7/12



Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

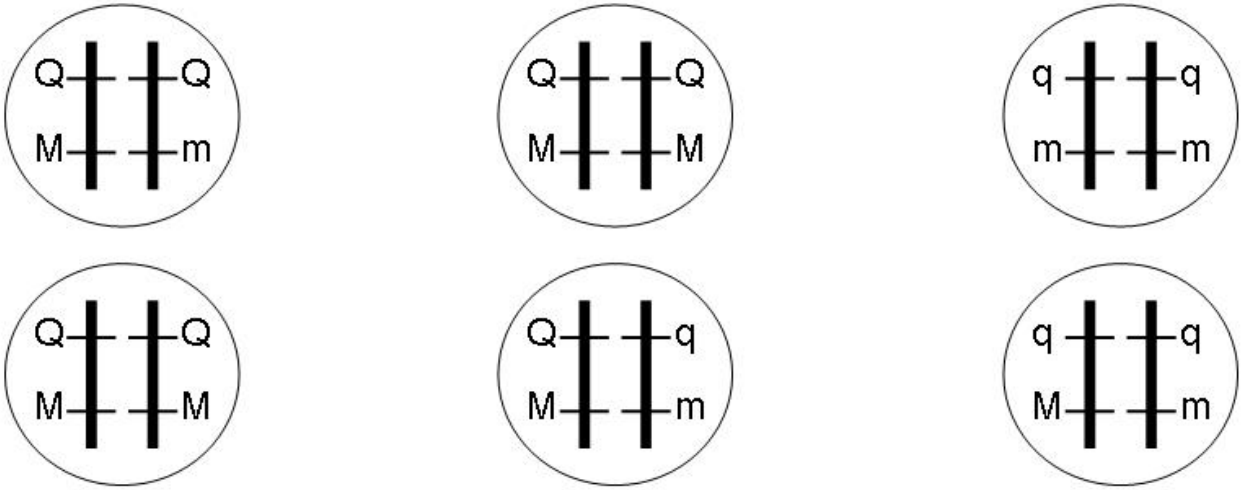
Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

1) Deux loci proches sur la carte sont en LD



Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

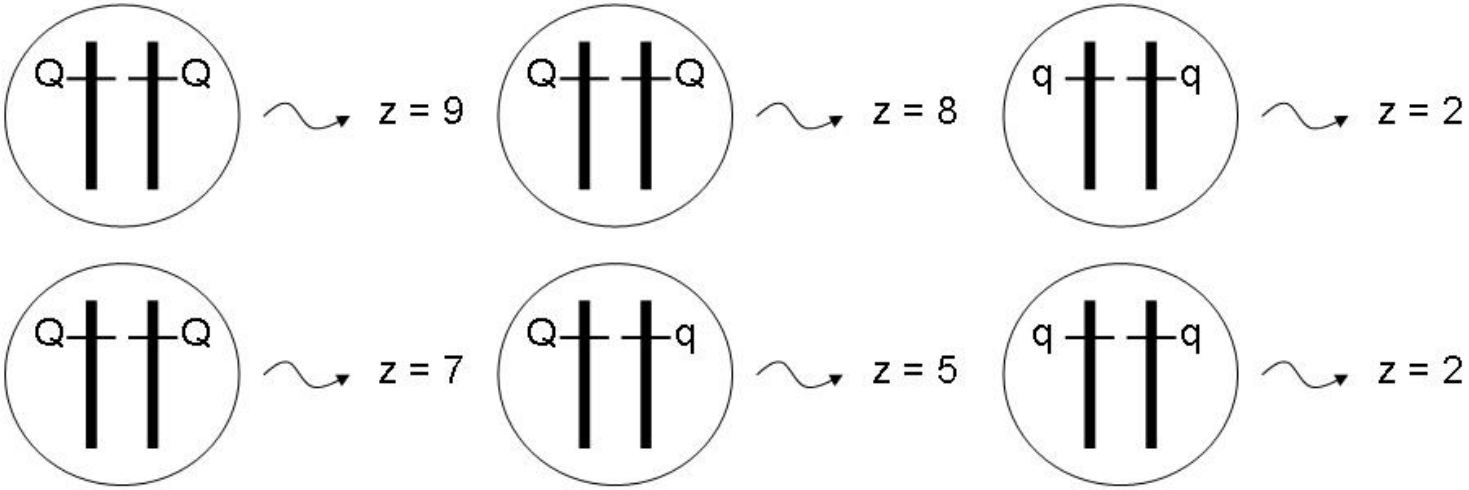
Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

2) Trouver un locus en LD avec un QTL



Déséquilibre de liaison et cartographie

A l'échelle d'une population, pour 2 loci :

Π_{i_1,i_2} fréquence de l'haplotype (i_1, i_2)

$\Pi_{i_1,.} = \sum_{i_2} \Pi_{i_1,i_2}$ fréquence de l'allèle i_1

$\Pi_{.,i_2} = \sum_{i_1} \Pi_{i_1,i_2}$ fréquence de l'allèle i_2

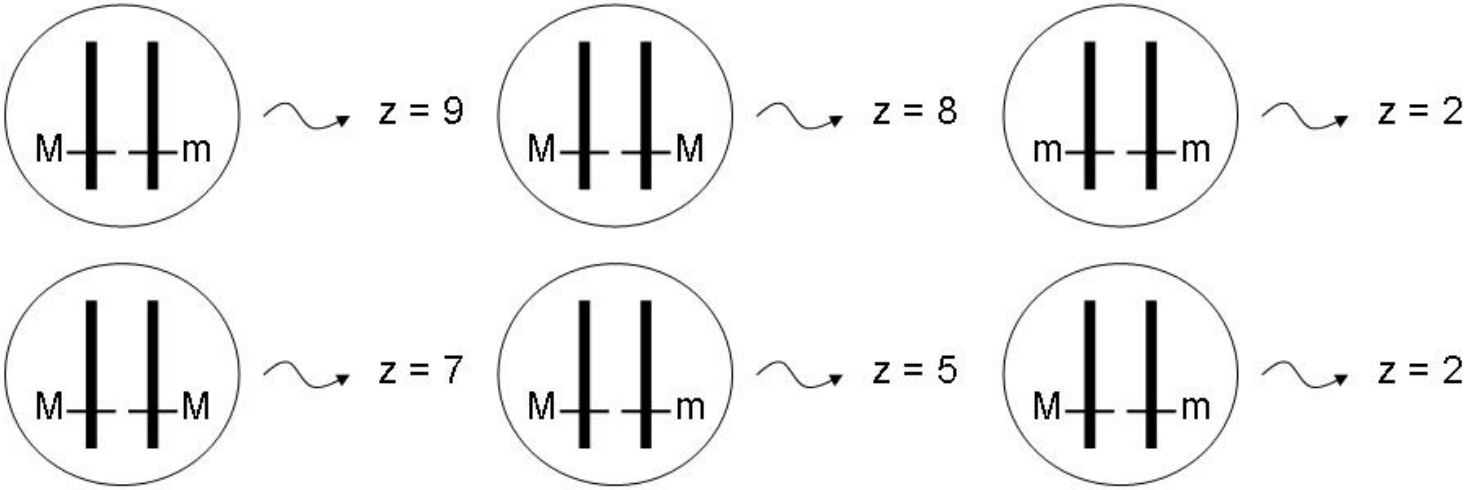
Equilibre de liaison :

$$\Pi_{i_1,i_2} = \Pi_{i_1,.} \Pi_{.,i_2}$$

Déséquilibre de liaison (LD) :

$$D_{i_1,i_2} = \Pi_{i_1,i_2} - \Pi_{i_1,.} \Pi_{.,i_2}$$

2) Trouver un locus en LD avec un QTL



Méthodes d'association (1/2)

(Boerwinkle *et al*, 1986, 1987)

- Modèle linéaire pour chaque marqueur :

$$Z_{k,j} = m_k + \epsilon_{k,j},\ k = 1, \ldots, K,\ j = 1, \ldots, N_k \quad (M)$$

  - $Z_{k,j}$ phénotype de l'individu j de génotype k.
  - m_k effet du génotype k.
  - $\epsilon_{k,j}$ bruit gaussien centré.
- Test dans (M) de H_0 : "$m_1 = \ldots = m_K$"
- Estimer x par la position du test le plus significatif.

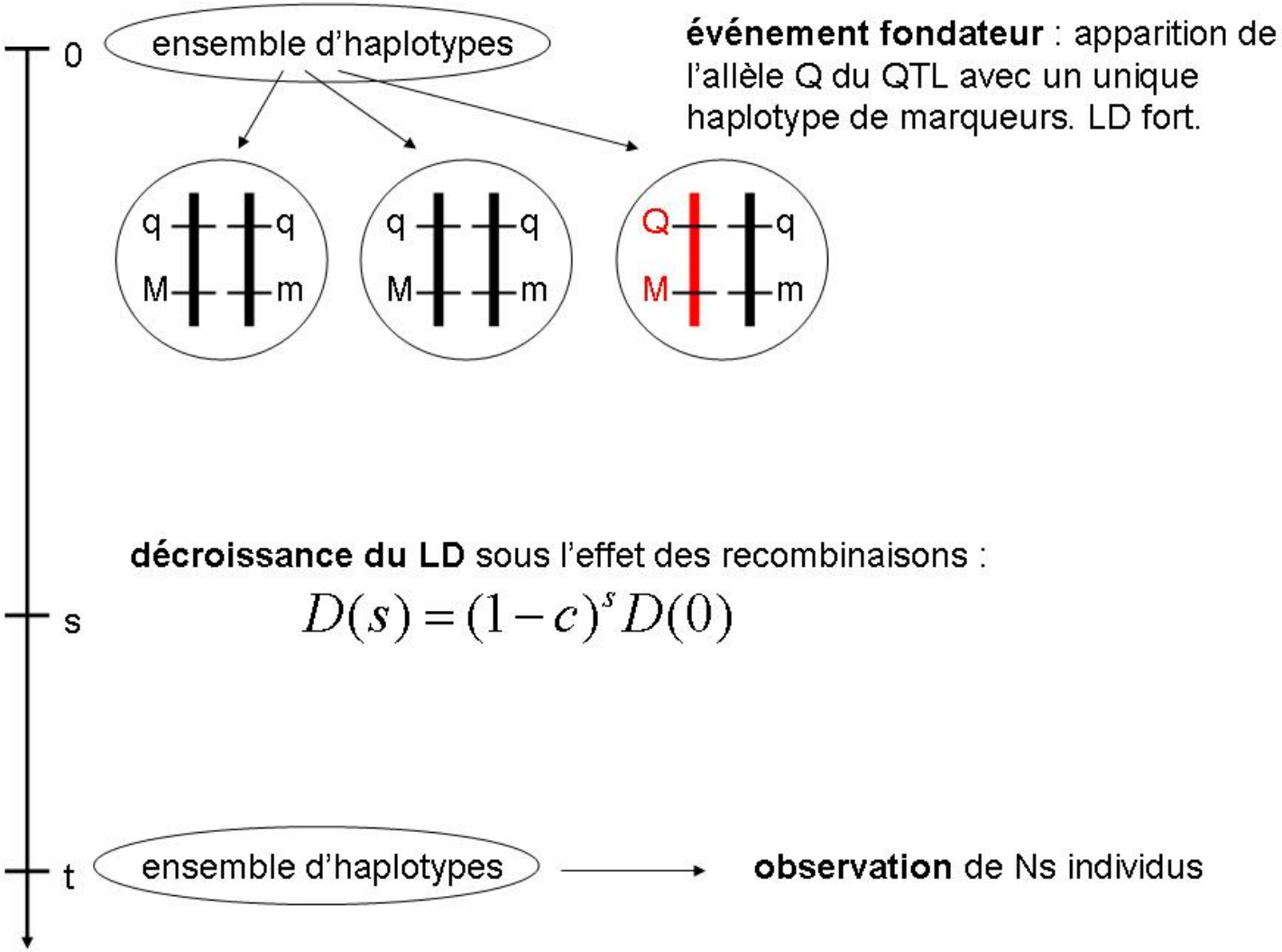
Méthodes d'association (2/2)

Quelques problèmes :

- Toutes les positions ne sont pas testées.
- Tests multiples pour la détection.
- Nombre de paramètres augmente très vite pour les modèles à plusieurs loci

$\longrightarrow$ Utilisation de **modèles d'évolution de la population** pour expliquer la structure des données observées.

Hypothèse de l'événement fondateur
(Hästbacka *et al*, 1992)



Utilisation de modèles d'évolution de la population

- Calcul de la vraisemblance

$$\mathcal{L}(x) = \mathbb{P}(\mathcal{D} \mid x) = \mathbb{E}[\mathbb{P}(\mathcal{D} \mid x, \mathcal{I})] = \int \mathbb{P}(\mathcal{D} \mid x, \mathcal{I})\mathbb{P}(\mathcal{I} \mid x)\, d\mathcal{I}$$

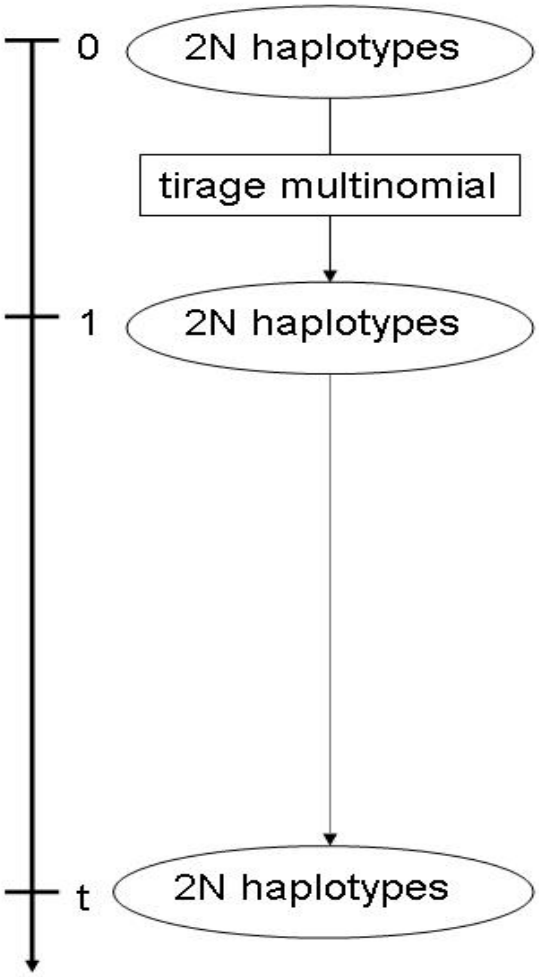
  $\mathcal{I} \rightarrow$ histoire de la population.
- $\mathbb{P}(\mathcal{D} \mid x, \mathcal{I})$ calculable, $\mathbb{P}(\mathcal{I} \mid x)\ d\mathcal{I}$ non.
- Méthodes : EM, MCMC, importance sampling.
- $\mathcal{I}$ dans un espace très large $\rightarrow$ **problème numérique**.

Différentes approches

- **Identité par descendance (IBD)** :
  - $\mathcal{I}$: étendue des segments IBD entre individus de l'échantillon.
  - McPeak et Strahs (1999), Morris *et al* (2000), Meuwissen et Goddart (2000,2001)*, Grapes *et al* (2005)*, Perez-Encizo (2003)*.
- **Coalescence** :
  - $\mathcal{I}$: généalogie des individus de l'échantillon.
  - Stephens et Donelly (2000), Morris *et al* (2002), Zöllner et Pritchard (2005)*.
- **Fréquentiste** :
  - $\mathcal{I}$: fréquences d'haplotypes à l'instant t.
  - Kaplan *et al* (1995), Terwilliger(1995), Xiong et Guo (1997), Abdallah *et al* (2004)*.

Modèle de Wright-Fisher

Modèle général



$X(0)$ vecteur des effectifs d'haplotypes

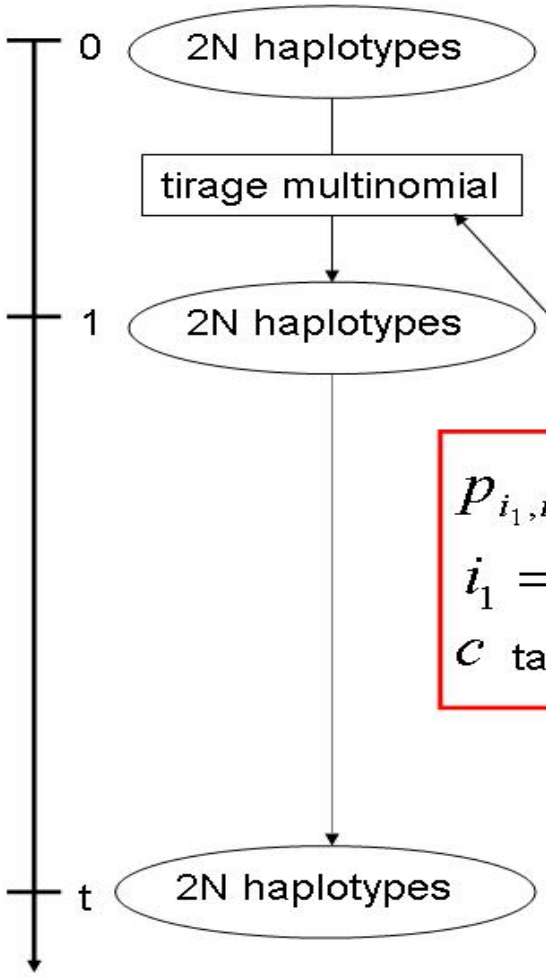
$\Pi(0) = \dfrac{X(0)}{2N}$ vecteur des fréquences d'haplotypes

$X(1), \Pi(1)$

$X(t), \Pi(t)$

Modèle de Wright-Fisher

Modèle à 2 loci



$$X(0) = (X_{1,1}(0), X_{1,2}(0), \ldots, X_{I_1,I_2}(0))$$

$$\Pi(0) = \frac{X(0)}{2N}$$

$$X(1), \Pi(1)$$

$$p_{i_1,i_2} = (1-c)\Pi_{i_1,i_2}(0) + c\Pi_{i_1,.}(0)\Pi_{.,i_2}(0)$$

$$i_1 = 1,\ldots,I_1 \quad i_2 = 1,\ldots,I_2$$

c taux de recombinaison entre les loci

$$X(t), \Pi(t)$$

Approche fréquentiste (1/2)

- Vraisemblance :

$$\mathcal{L}(x) = \mathbb{E}[\mathbb{P}(\mathcal{D} \mid x, \Pi(t))] = \int \mathbb{P}(\mathcal{D} \mid x, \Pi(t))\mathbb{P}(\Pi(t) \mid x)\, d\Pi(t)$$

  $\Pi(t)$ vecteur des fréquences d'haplotypes (QTL+marqueurs)
- Loi de $\Pi(t)$: modèle de Wright-Fisher + événement fondateur.
- Approximation d'ordre 1 quand $N \to \infty$ (Hill et Weir, 1995) :

$$\mathcal{L}(x) \approx \mathbb{P}(\mathcal{D} \mid x, \mathbb{E}[\Pi(t)])$$

Approche fréquentiste (2/2)

méthode	marqueurs	vraisemblance	caractère
Kaplan *et al* (1995)	1	Monte Carlo	maladie
Terwilliger (1995)	1*	ordre 1	maladie
Xiong et Guo (1997)	1	ordre 2	maladie rare
Xiong et Guo (1997)	2	ordre 1	maladie rare
Abdallah *et al* (2004)	1*	ordre 1	QTL
partie 3	2	ordre 1	QTL
partie 4	1	diffusion	maladie

* : et composite.

1. Définitions et modèles
   - Modélisation du génome
   - Cartographie de QTL
2. Cartographie par déséquilibre de liaison
   - Principe
   - Etat de l'art
   - Approche fréquentiste
3. Cartographie de QTL à l'aide de marqueurs flanquants : HAPim
4. Loi des fréquences d'haplotypes sous un modèle à deux loci
5. Bilan et perspectives

Modèles

- Modèle de mélange :

$$Z_n \rightsquigarrow \begin{cases} \mathcal{N}(\mu + a, \sigma^2) & \text{avec probabilité } \mathbb{P}(Q/Q \mid h_n^1/h_n^2) \\ \mathcal{N}(\mu - a, \sigma^2) & \text{avec probabilité } \mathbb{P}(q/q \mid h_n^1/h_n^2) \\ \mathcal{N}(\mu + d, \sigma^2) & \text{avec probabilité } \mathbb{P}(Q/q \mid h_n^1/h_n^2) \end{cases}$$

  μ effet moyen, a effet additif, d effet de dominance, σ^2 variance (Falconer et Mackay, 1996).

- Equilibre d'Hardy-Weinberg :

$$\mathbb{P}(Q/Q \mid h_n^1/h_n^2) = \mathbb{P}(Q \mid h_n^1)\ \mathbb{P}(Q \mid h_n^2)$$

- $\mathbb{P}(\mathbf{Q} \mid \mathbf{h}) = \frac{\mathbf{\Pi_{Q,h}(t)}}{\mathbf{\Pi_h(t)}}$.

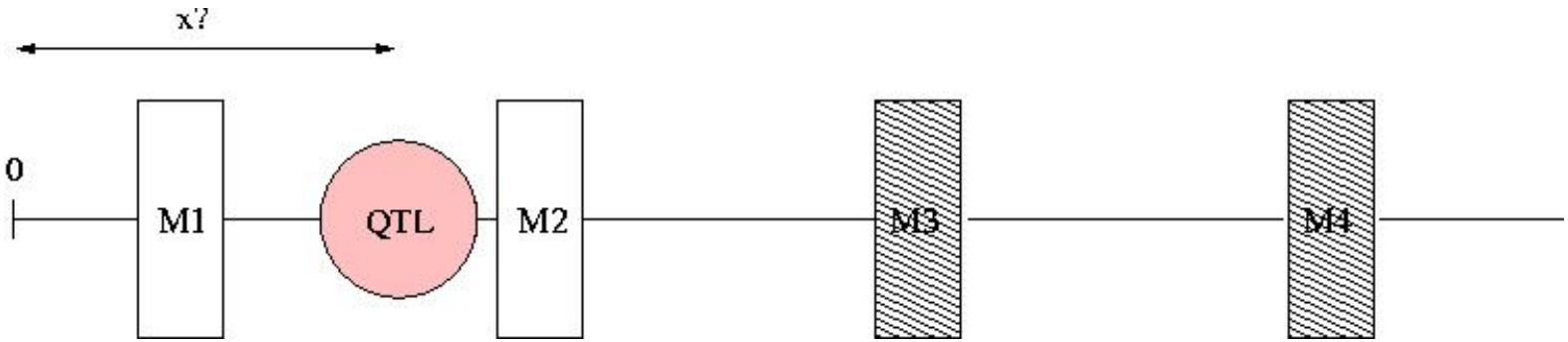
Vraisemblance

- Sachant les fréquences d'haplotypes à l'instant t :

$$\begin{aligned}\mathbb{P}(\mathcal{D} \mid x, \Pi) &= \prod_{n=1}^{N_s} \Big[\phi_{\mu+a,\sigma^2}(z_n) \frac{\Pi_{Q,h_n^1}}{\Pi_{h_n^1}} \frac{\Pi_{Q,h_n^2}}{\Pi_{h_n^2}} \\ &+ \phi_{\mu-a,\sigma^2}(z_n) \frac{\Pi_{q,h_n^1}}{\Pi_{h_n^1}} \frac{\Pi_{q,h_n^2}}{\Pi_{h_n^2}} \\ &+ \phi_{\mu+d,\sigma^2}(z_n) \Big(\frac{\Pi_{q,h_n^1}}{\Pi_{h_n^1}} \frac{\Pi_{Q,h_n^2}}{\Pi_{h_n^2}} + \frac{\Pi_{Q,h_n^1}}{\Pi_{h_n^1}} \frac{\Pi_{q,h_n^2}}{\Pi_{h_n^2}} \Big) \Big]\end{aligned}$$

- Ordre 1 : $\mathbb{P}(\mathcal{D} \mid x, \mathbb{E}[\Pi])$.

Utilisation des marqueurs flanquants



- A chaque position x, utilisation des 2 marqueurs flanquants.
- Un haplotype h = une paire (i_1, i_2)
  i_1 allèle du marqueur 1, i_2 allèle du marqueur 2.
- $\Pi_{Q,h} = \Pi_{i_1,Q,i_2}$.

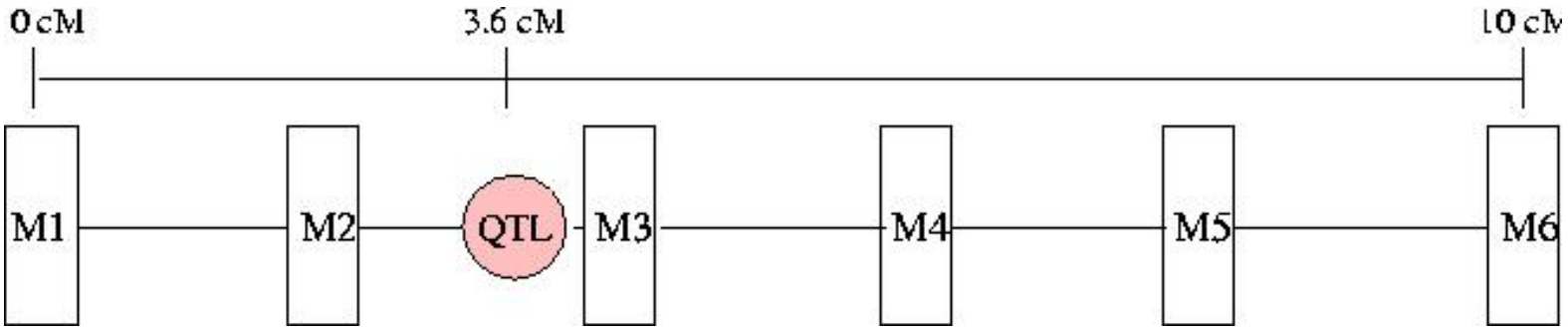
Démarche

1. Expression de $\mathbb{E}[\Pi_{i_1,Q,i_2}(t)]$:
   - Modèle de Wright-Fisher à 3 loci.
   - Hyp 1 : fréquences des allèles aux marqueurs constantes.
   - Hyp 2 : équilibre de liaison entre marqueurs à $t = 0$.
2. Calcul direct de la vraisemblance en fonction de :
   - μ, a, d, σ^2.
   - x, t, fréquences alléliques.
3. Maximisation de la vraisemblance (simplexe librairie NAG).

Résultats de simulation (1/2)

N	distance inter-marqueurs	EQM	
		T2	HAPim
200	2 cM	2.93	**2.77**
400	2 cM	1.81	**1.44**
200	0.25 cM	0.71	**0.46**
400	0.25 cM	0.49	**0.30**

Erreurs quadratiques moyennes (EQMs) en cM^2 des estimations de la position x du QTL. Comparaison de HAPim avec la méthode T2 de Abdallah *et al* (2004). $t = 100$, $a = d = \sigma = 1$, 5 allèles par marqueur. 1 cM = 1% de recombinaison.



Résultats de simulation (1/2)

distance inter-marqueurs	EVAM			
	HAPim	IBD(2)	IBD(4)	IBD(6)
1 cM	1.30 (66%)	1.32 (65.4%)	1.25 (67.2%)	**1.20** (**68.9%**)
0.5 cM	0.72 (64%)	0.70 (62.9%)	**0.63** (**66.9%**)	0.70 (61.3%)
0.25 cM	0.39 (60%)	0.38 (58.3%)	**0.34** (**61.2%**)	0.40 (52.5%)

Erreurs en valeur absolue moyennes (EVAMs) en cM des estimations de la position x du QTL (et pourcentage d'estimations à moins d'un intervalle de la bonne position). Comparaison de HAPim avec la méthode IBD de Grapes *et al* (2005).
$t = 100$, $N = 100$, $a = \sigma = 1$, $d = 0$, 2 allèles par marqueur.

1. Définitions et modèles
   - Modélisation du génome
   - Cartographie de QTL
2. Cartographie par déséquilibre de liaison
   - Principe
   - Etat de l'art
   - Approche fréquentiste
3. Cartographie de QTL à l'aide de marqueurs flanquants : HAPim
4. Loi des fréquences d'haplotypes sous un modèle à deux loci
5. Bilan et perspectives

Principe

- Motivation : calcul de la densité de $\Pi(t)$ ou de quantités liées à cette densité comme $\mathcal{L}(x) = \mathbb{E}[\mathbb{P}(\mathcal{D} \mid x, \Pi(t))]$.
- Démarche :
  1. Approximation du processus de Wright-Fisher $\Pi(t)$ par le processus de diffusion $Y(\tau)$ où $\tau = \frac{t}{2N}$ et $\rho = 2Nc$. (Ethier and Nagylaki, 1989).
  2. Pour $Y(\tau)$, équations de Kolmogorov pour la densité (forward) ou la vraisemblance (backward).
  3. Résolution numérique de ces équations (2 loci bi-alléliques).

Equation "forward" de Kolmogorov

- $Y(\tau) = (Y_{1,1}(\tau), Y_{1,2}(\tau), Y_{2,1}(\tau))$.
- $Y(\tau) \in \mathbb{F}_3 = \{y \in [0,1]^3, y_1 + y_2 + y_3 \leq 1\}$.
- $b(y)$ et $a^2(y)$ coefficients infinitésimaux de $Y(\tau)$.
- $f(y^0, y, \tau)$ densité de transition de y^0 à y en un temps τ.

Théorème

Pour tout $y^0, y \in \mathring{\mathbb{F}}_3$ et $\tau > 0$, la densité f existe et vérifie :

$$\frac{\partial f}{\partial \tau} = L^* f = -\sum_j \frac{\partial (b_j(y) f)}{\partial y_j} + \frac{1}{2} \sum_{j_1, j_2} \frac{\partial^2 (a^2_{j_1, j_2}(y) f)}{\partial y_{j_1} \partial y_{j_2}}$$

$j = (1,1), (1,2), (2,1)$

Résolution par les différences finies

- N_τ intervalles de temps, $\Delta_\tau = \tau / N_\tau$.
- N_y intervalles de fréquences, $\Delta_y = 1/N_y$.
- $n = 0, \ldots, N_\tau$, $k, l, m = 0, \ldots, N_y$,

$$\tilde{f}^{(n)}(k, l, m) \approx f(y^0, (k\Delta_y, l\Delta_y, m\Delta_y), n\Delta_\tau)$$

- $\tilde{f}^{(0)}$ nul sauf autour de y^0.
- $n = 1, \ldots, N_\tau$, $k, l, m = 1, \ldots, (N_y - 1)$

$$\frac{\tilde{f}^{(n)}(k, l, m) - \tilde{f}^{(n-1)}(k, l, m)}{\Delta_\tau} = \tilde{L}^* \tilde{f}^{(n-1)}(k, l, m)$$

Conditions aux bords

- Bords : $y_{1,1} = 0$, $y_{1,2} = 0$, $y_{2,1} = 0$ et $y_{1,1} + y_{1,2} + y_{2,1} = 1$.
- Besoin de conditions supplémentaires.
- Problèmes :
  - Conditions standards ($f = 0$, $\nabla f = 0$) ne s'appliquent pas.
  - Absorbants ($y_{1,1} + y_{1,2} = 0$) ou non absorbants ($y_{1,2} + y_{2,1} = 0$).
- Approximation : annulations des dérivées secondes.

$$\frac{\tilde{f}^{(n)}(0) - 2\tilde{f}^{(n)}(1) + \tilde{f}^{(n)}(2)}{\Delta_y^2} = 0$$

Discussion

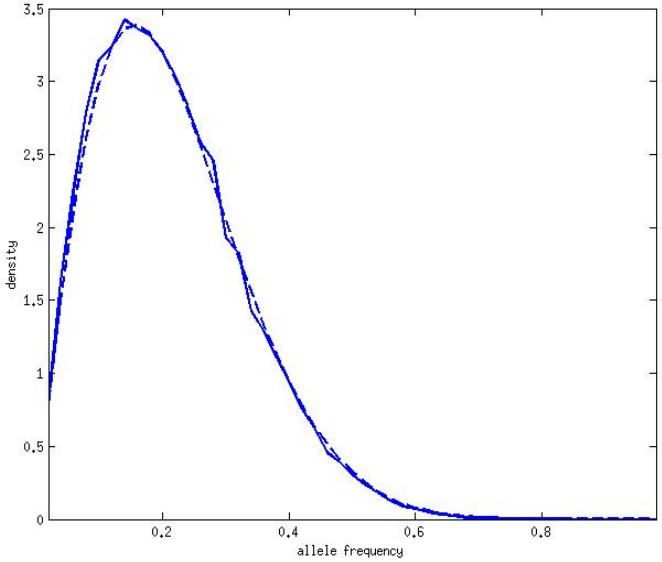
- Avantages de l'approche par la diffusion :
  - Démarche classique en génétique.
  - $f(y^0, y, \tau)$ pour tout $y \in \mathring{\mathbb{F}}_3$.
  - Plus rapide que simulations.
  - Valable en régime transitoire.
  - Mutations, sélection possibles.
- Tests de précision :
  - $c = 0$: modèle à un locus $\rightarrow$ densité connue (Griffiths, 1979).
  - $c \neq 0$: simulations du modèle de Wright-Fisher.

Résultats pour $c = 0$

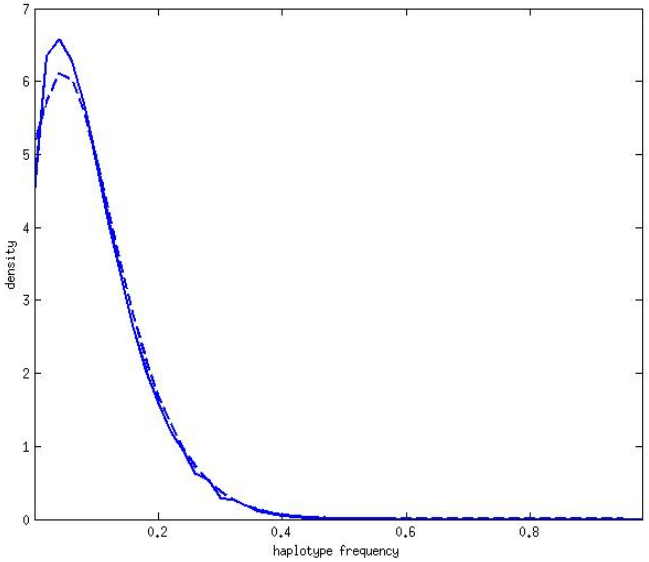
	τ		
y^0	0.1	0.5	1
$(0.15, 0.05, 0.25)$	9.41e-04	1.28e-04	2.73e-06
$(0.10, 0.01, 0.30)$	6.80e-02	1.70e-03	3.20e-05

Distance d'Hellinger entre la densité exacte des fréquences d'haplotypes et notre approximation pour un modèle de diffusion à deux loci liés ($\rho = 0$).

Résultats pour $c \neq 0$



(a) $Y_{1,.}(\tau) = Y_{1,1}(\tau) + Y_{1,2}(\tau)$



(b) $Y_{2,1}(\tau)$

FIG.: Densités unidimensionnelles pour le modèle à 2 loci obtenues par notre méthode (lignes pointillées) ou par simulations de Wright-Fisher pour $N = 200$ (lignes pleines). $\tau = 0.1$, $\rho = 10$, $y^0 = (0.15, 0.05, 0.25)$.

1. Définitions et modèles
   - Modélisation du génome
   - Cartographie de QTL
2. Cartographie par déséquilibre de liaison
   - Principe
   - Etat de l'art
   - Approche fréquentiste
3. Cartographie de QTL à l'aide de marqueurs flanquants : HAPim
4. Loi des fréquences d'haplotypes sous un modèle à deux loci
5. Bilan et perspectives

Contributions

- Participation à un article de revue sur le déséquilibre de liaison (**partie 2**). Cierco *et al* (2004), *Research Signpost*.
- Développement d'une méthode de cartographie de QTL utilisant des marqueurs flanquants (**partie 3**).
  Boitard *et al* (2006), *BMC Genomics*.
- Développement d'une méthode numérique de résolution des équations de Kolmogorov pour un modèle de diffusion à 2 loci bi-alléliques.
  - équation "forward" : densité de transition (**partie 4**).
    Boitard et Loisel, en relecture à *Theor. Popul. Biol.*
  - équation "backward" : vraisemblance pour le taux de recombinaison.
- Loi asymptotique du Transmission Desequilibrium Test (TDT) dans un modèle de population structurée.

Perspectives

- Calcul de la vraisemblance du taux de recombinaison : comparer notre algorithme avec des simulations de "l'ancestral recombination graph".
- Extension des méthodes :
  - HAPim : hypothèse d'équilibre, haplotypes plus longs, sélection, méthodes LDL ...
  - Densité des fréquences d'haplotypes : améliorer conditions aux bords, modèles avec mutation ou sélection.
- Utilisation de la densité des fréquences d'haplotypes pour d'autres problèmes de génétique des populations
  Ex : effet d'entraînement par un locus sous sélection (hitchhiking).