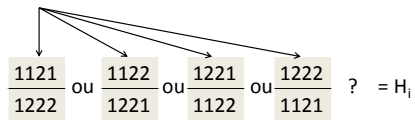


## Analyse de liaison

Etablissement des phases  
Calcul des probabilités de transmission

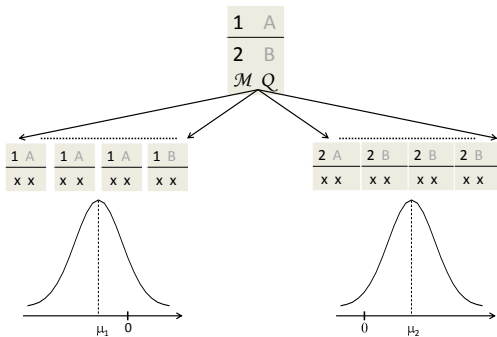
### Données de base :

(11, 12, 22, 12) =  $M_i$  les phénotypes marqueur du parent i

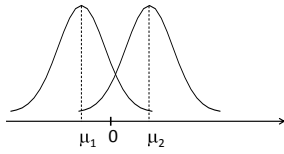


« Phasage » : retrouver les gamètes qui ont créé l'individu

### Principe de l'analyse de liaison (pourquoi faut-il phaser les parents ?)



### Principe de l'analyse de liaison



$$\begin{aligned} \mu_1 &= \frac{1}{2} \mu_A + \frac{1}{2} \mu_B \\ \mu_2 &= \frac{1}{2} \mu_B + \frac{1}{2} \mu_A \end{aligned}$$

Si  $\mu_1 \neq \mu_2 \Rightarrow \mathcal{M}$  est lié à Q

---

---

---

---

---

---

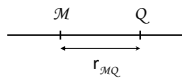
---

---

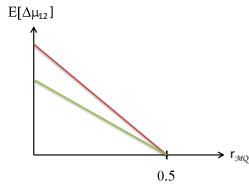
---

---

L'effet du QTL en Q sur les différences de moyenne entre classes de marqueurs  $\mathcal{M}$  diminue si le taux de recombinaison  $r_{\mathcal{M}Q}$  augmente



$$\begin{aligned} E[\mu_1] &= (1-r_{\mathcal{M}Q}) \mu_A + r_{\mathcal{M}Q} \mu_B \\ \text{et} \\ E[\mu_2] &= (1-r_{\mathcal{M}Q}) \mu_B + r_{\mathcal{M}Q} \mu_A \\ \Rightarrow E[\Delta\mu_{12}] &= (1-2r_{\mathcal{M}Q}) \Delta\mu_{AB} \end{aligned}$$



La puissance du test dépend beaucoup du taux de recombinaison marqueur - QTL

---

---

---

---

---

---

---

---

---

---

### Cartographie d'intervalle

1	A	1
2	B	2
$\mathcal{M}$	Q	$\mathcal{N}$

$n_{11}$	$1 \times 1$	$n_{12}$	$1 \times 2$	$n_{21}$	$2 \times 1$	$n_{22}$	$2 \times 2$
$f_{11}$	1 A 1	1 A 2	1 B 2	2 A 1	2 B 1	$f_{22}$	2 A 2
$1-f_{11}$	1 B 1					$1-f_{22}$	2 B 2

$$\begin{aligned} \mu_{11} &= f_{11} \mu_A + (1-f_{11}) \mu_B & \mu_{22} &= f_{22} \mu_B + (1-f_{22}) \mu_A \\ \mu_{11} - \mu_{22} &= (f_{11} - 1 + f_{22}) \mu_A + (1 - f_{11} - f_{22}) \mu_B \\ \Rightarrow \Delta\mu_{11,22} &= (f_{11} + f_{22} - 1) \Delta\mu_{AB} \end{aligned}$$


---

---

---

---

---

---

---

---

---

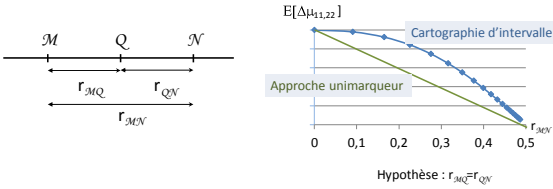
---

### Cartographie d'intervalle

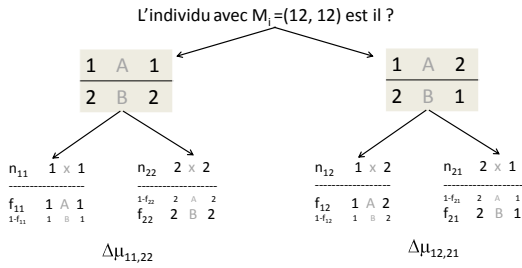
$$\Delta\mu_{11,22} = (f_{11} + f_{22} - 1) \Delta\mu_{AB}$$

$E[f_{11}] = \text{prob}(\text{l'individu a reçu 1A1 / il a reçu 11}) = \text{prob}(1A1)/\text{prob}(11) = 1/2 (1 - r_{MQ}) (1 - r_{QN}) / 1/2 (1 - r)$ 
 $E[f_{22}] = \text{prob}(2B2)/\text{prob}(22) = (1 - r_{MQ}) (1 - r_{QN}) / (1 - r) = E[f_{11}]$

$$E[\Delta\mu_{11,22}] = (2(1 - r_{MQ})(1 - r_{QN}) / (1 - r) - 1) \Delta\mu_{AB}$$



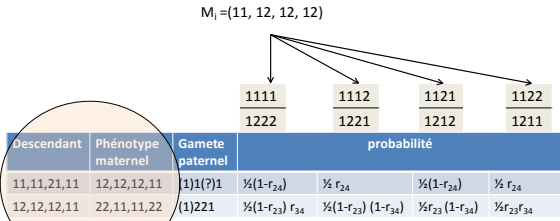
### Pour la cartographie d'intervalle Il faut phaser les parents



### Méthodes disponibles dans QTLMap

- Approche exhaustive par énumération
- Approche de type Windig et Meuwissen
- Approche Favier et al

### Approche exhaustive par énumération (1/3)



$Prob(\text{observation} \mid H_i=1111/1222) = [\frac{1}{2}(1-r_{24})] \cdot [\frac{1}{2}(1-r_{23})r_{34}] \dots$   
 $Prob(\text{observation} \mid H_i=1112/1221) = [\frac{1}{2}r_{24}] \cdot [\frac{1}{2}(1-r_{23})(1-r_{34})] \dots$

---

---

---

---

---

---

---

---

---

---

---

### Approche exhaustive par énumération (2/3)

$$Prob(H_i=1111/1222 \mid \text{observation}) = \frac{Prob(\text{observation} \mid H_i=1111/1222) \cdot Prob(H_i=1111/1222)}{Prob(\text{observation})}$$

En général :  $Prob(H_i \mid \Omega) = \frac{Prob(\Omega \mid H_i) \cdot Prob(H_i)}{\sum_j Prob(\Omega \mid H_j) \cdot Prob(H_j)}$

A priori sur les phases... en pratique  $(1/2)^{\text{nb\_hetero}}$

Pour chaque phase possible (maximum :  $2^{**\text{nb}} \text{ marqueur}$ )  
 On calcule les proba des phénotypes de chaque candidat :  $Prob(\Omega_k \mid H_i)$   
 En listant pour chaque phénotype le ou les allèles que le père a pu transmettre  
 Donc les occurrences, ou non, de recombinaisons entre marqueurs  
 Puis on multiplie ces proba pour obtenir les  $Prob(\Omega \mid H_i)$  ( $= \prod_k Prob(\Omega_k \mid H_i)$ )  
 Et on en déduit les  $Prob(H_i \mid \Omega)$

---

---

---

---

---

---

---

---

---

---

---

### Approche exhaustive par énumération (3/3)

**Intérêt**  
Solutions exactes  
Référence historique

**Inconvénient**  
Hypothèse d'équilibre de liaison  
Temps de calcul prohibitif si  $M > 12$   
Besoins de stockage importants

QTLMap : option --haplotype= 1, 2 ou 3

---

---

---

---

---

---

---

---

---

---

---

### Approche de type Windig et Meuwissen (1/2)

Méthode approximative par exploration successive des marqueurs

$$M_i = (11, 12, 22, 12, 12, 11, 12, \dots)$$

$$\text{Etape 1 : } H_i(1) = \frac{1}{1}$$

$$\text{Etape 2 : } H_i(1,2) = \frac{11}{12}$$

$$\text{Etape 3 : } H_i(1,2,3) = \frac{112}{122}$$

$$\text{Etape 4 : si } g_{1121} + g_{1222} > g_{1122} + g_{1221} \Rightarrow H_i(1,2,3,4) = \frac{1121}{1222}$$

$$\text{si } g_{1121} + g_{1222} < g_{1122} + g_{1221} \Rightarrow H_i(1,2,3,4) = \frac{1122}{1221}$$

.....

---

---

---

---

---

---

---

---

---

---

---

---

### Approche de type Windig et Meuwissen (2/2)

Méthode approximative par exploration successive des marqueurs

**Intérêt**  
Très rapide  
Adapté aux cartes denses

**Inconvénient**  
Hypothèse d'équilibre de liaison  
Incohérences possibles

QTLMap : option --haplotype = 4

---

---

---

---

---

---

---

---

---

---

---

---

### Approche Favier et al (1/5)

Méthode exacte et rapide

Parent  $i$   
 $M_i = (11, 12, 22, 12, 12) \Rightarrow H_i = (11211 \mid 12222)$   
 ou  $(12211 \mid 11222)$   
 ou  $(11221 \mid 12212)$  etc.

Descendant  $ik$       Evènements de transmission de  $i$  à  $k$   
 $\Omega_{ik} = (11, 11, 12, 12, 22) \Rightarrow T_{ik} = (0, 1, 0, 0, 2)$

0 = Inconnu  
 1 = 1<sup>er</sup> allele lu  
 2 = 2<sup>ème</sup> allele lu

---

---

---

---

---

---

---

---

---

---

---

---

### Approche Favier et al (2/5)

Méthode exacte et rapide

Descendant  $ik$  Evènements de transmission de  $i$  à  $k$   
 $\Omega_{ik}=(11,11,12,12,22) \Rightarrow T_{ik}=(0,1,0,0,2)$  0 = Inconnu  
1 = 1<sup>er</sup> allele lu  
2 = 2<sup>eme</sup> allele lu

$$\text{Prob}(H_i | \Omega) = \frac{\text{Prob}(\Omega | H_i) \cdot \text{Prob}(H_i)}{\sum_j \text{Prob}(\Omega | H_j) \cdot \text{Prob}(H_j)} = \frac{\text{Prob}(T | H_i) \cdot \text{Prob}(H_i)}{\sum_j \text{Prob}(T | H_j) \cdot \text{Prob}(H_j)}$$

avec  $\text{Prob}(T | H_i) = \prod_k \text{Prob}(T_{ik} | H_i)$  et  $\text{Prob}(T_{ik} | H_i) = \prod_l \text{Prob}(T_{ikl} | H_i, T_{ik,l})$

---

---

---

---

---

---

---

---

---

---

---

---

### Approche Favier et al (3/5)

Méthode exacte et rapide

On cherche le  $H_i$  qui maximise  $\text{Prob}(H_i | \Omega)$

La phase  $H_i$  est décrite par le vecteur  $h_i$   
 avec  $h_{ij} = -1$  si le 1<sup>er</sup> allele lu au locus  $l$  vient du grand père  
 et  $h_{ij} = 1$  s'il vient de la grand mère  
 $M_i = (11, 12, 22, 12, 12)$  et  $H_i = (11221 | 12212) \Rightarrow h_i = (1, -1, 1, 1, -1)$  par exemple

---

---

---

---

---

---

---

---

---

---

---

---

### Approche Favier et al (4/5)

Méthode exacte et rapide

« Maximiser  $\text{Prob}(H_i | \Omega)$  » est reformulé en « maximiser  $V_i = h_i' W_i h_i$  »

La matrice  $W_i$  est une fonction des observations et des taux de recombinaisons

$$W_{ik} = (N_{ik}^+ - N_{ik}^-) [ \text{Log}(1-r_{ik}) - \text{Log}(r_{ik}) ]$$

$N_{ik}^+$  : nombre de descendants avec  
 $T_{ik} \neq 0 \quad T_{k,i1}=0 \quad \dots \quad T_{k,i2}=0 \quad T_i = T_{ik}$   
 $N_{ik}^-$  : nombre de descendants avec  
 $T_{ik} \neq 0 \quad T_{k,i1}=0 \quad \dots \quad T_{k,i2}=0 \quad T_i \neq T_{ik}$

Un problème d'optimisation abordable par des algorithmes spécifiques (Toolbar, MIA Toulouse)

---

---

---

---

---

---

---

---

---

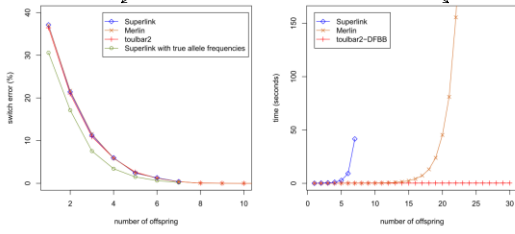
---

---

---

### Approche Favier et al (5/5)

Méthode exacte et rapide



QTLMap : option --haplotype = 5

---

---

---

---

---

---

---

---

### Analyse de liaison

Etablissement des phases

Calcul des probabilités de transmission

---

---

---

---

---

---

---

---

### Cas simple : un seul marqueur

1) Soit un père  $i$  de phénotype marqueur  $M_i = \{12\}$

Cas 1 : les grands parents sont génotypés  
 La phase est la répartition des allèles du père en terme d'origine grand parentale  
 $H_i = \{12\}$  allèle venant du grand père    allèle venant de la grand-mère

Cas 2 : pas d'information sur les grands parents  
 La « phase »  $H_i$  est arbitraire  
 $H_i = \{12\}$  premier allèle lu    deuxième allèle lu

---

---

---

---

---

---

---

---

### Cas simple : un seul marqueur

2) On cherche quel allèle a été transmis du père aux descendants

Le père i est  $H_i = (12)$

Un descendant ik, de phénotype  $\Omega_{ik}$  quelle transmission ( $T_{ik}$ ) ?

Mère inconnue

$\Omega_{ik}$	$T_{ik}$
11	1
12	0
22	2

Mère connue

$\Omega_{ik}$	$M_{ij}$	11	12	22
11	1	1	1	♣
12	2	2	0	1
22	♣	2	2	2

---

---

---

---

---

---

---

---

---

---

### Cas simple : un seul marqueur

On suppose le père A/B en un QTL proche du marqueur

1	A
2	B
M	Q

Par définition A est l'allèle du chromosome 1 (du grand père ou premier lu)

3) Connaissant  $H_i$  et  $T_{ik}$  on calcule la probabilité que le père ait transmis A (ou B)

$T_{ik}$	Prob( $Q_{ik}=A / T_{ik}$ )	Prob( $Q_{ik}=B / T_{ik}$ )
1	$1-r_{MQ}$	$r_{MQ}$
2	$r_{MQ}$	$1-r_{MQ}$
0	0.5	0.5

---

---

---

---

---

---

---

---

---

---

### Utilisation des marqueurs flanquants

Soit un père i de phénotype marqueur  $M_i = (12,12)$

On suppose sa phase  $H_i = (12 / 12)$  connue (origine grand parentale ou ordre de lecture)

1) On cherche quels allèles ont été transmis du père aux descendants

Un descendant ik, de phénotype  $\Omega_{ik}$  quelle transmission ( $T_{ik}$ ) ?

Mère inconnue

$\Omega_{ik}$	$T_{ik}$
11,11	11
11,12	10
11,22	12
12,11	01
12,12	00
12,22	02 ....

Mère connue

$\Omega_{ik}$	$M_{ij}$	11,11	11,12	11,22	12,11	12,12	12,22
11,11	11	11	11	♣	11	11	♣
11,12	12	10	10	11	12	10	11
11,22	♣	12	12	♣	12	12	12
12,11	21	21	♣	01	01	♣	♣
12,12	22	20	21	02	00	01	01
12,22	♣	22	22	♣	02	02 ....	02 ....

---

---

---

---

---

---

---

---

---

---



## Utilisation des marqueurs flanquants

On suppose le père A/B en un QTL entre les 2 marqueurs

1	A	1
2	B	2
$\mathcal{M}$	Q	$\mathcal{N}$

Par définition A est l'allèle du chromosome 1 (du grand père ou premier lu)

2) Connaissant  $H_i$  et  $T_{ik}$  on calcule la probabilité que le père ait transmis A (ou B)

$T_{ik}$	Prob( $Q_{ik}=A / T_{ik}$ )	Prob( $Q_{ik}=B / T_{ik}$ )
11	$(1-r_{MQ})(1-r_{QN})/(1-r_{MN})$	$r_{MQ}r_{QN}/(1-r_{MN})$
12	$(1-r_{MQ})r_{QN}/r_{MN}$	$r_{MQ}(1-r_{QN})/r_{MN}$
10	$(1-r_{MQ})$	$r_{MQ}$
21	$r_{MQ}(1-r_{QN})/r_{MN}$	$(1-r_{MQ})r_{QN}/r_{MN}$
22	$r_{MQ}r_{QN}/(1-r_{MN})$	$(1-r_{MQ})(1-r_{QN})/(1-r_{MN})$
20	$r_{MQ}$	$(1-r_{MQ})$
01	$(1-r_{QN})$	$r_{QN}$
02	$r_{QN}$	$(1-r_{QN})$
00	0.5	0.5

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

## Utilisation des marqueurs flanquants

$T_{ik}$	Prob( $Q_{ik}=A / T_{ik}$ )	Prob( $Q_{ik}=B / T_{ik}$ )
11	$(1-r_{MQ})(1-r_{QN})/(1-r_{MN})$	$r_{MQ}r_{QN}/(1-r_{MN})$
12	$(1-r_{MQ})r_{QN}/r_{MN}$	$r_{MQ}(1-r_{QN})/r_{MN}$
10	$(1-r_{MQ})$	$r_{MQ}$
21	$r_{MQ}(1-r_{QN})/r_{MN}$	$(1-r_{MQ})r_{QN}/r_{MN}$
22	$r_{MQ}r_{QN}/(1-r_{MN})$	$(1-r_{MQ})(1-r_{QN})/(1-r_{MN})$
20	$r_{MQ}$	$(1-r_{MQ})$
01	$(1-r_{QN})$	$r_{QN}$
02	$r_{QN}$	$(1-r_{QN})$
00	0.5	0.5

L'information sur la transmission n'est pas toujours de bonne qualité

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

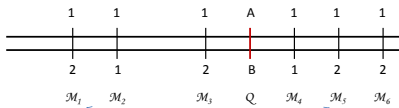
---

---

---

---

## Utilisation des marqueurs flanquants informatifs



$\Omega_{ik}=(11,11,12,12,11,22) \Rightarrow T_{ik}=(100012)$

Pour le descendant ik, les marqueurs flanquants informatifs (MFI) Q sont  $\mathcal{M}_1$  et  $\mathcal{M}_5$

$Prob(Q_{ik}=A / T_{ik}) = (1-r_{M1Q})(1-r_{QM5})/(1-r_{M1M5})$

Les MFI varient d'un descendant à l'autre

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

Généralisation aux transmissions des deux parents



$\Omega_k$	$T_k$
11,11	12,11
11,12	12,00
11,22	12,22
12,11	00,11
12,12	00,00
12,22	02,22
22,11	22,11
22,12	22,00
22,22	22,22

---

---

---

---

---

---

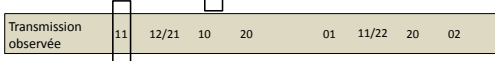
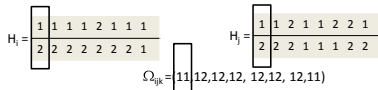
---

---

---

---

Utilisation de Zones Flanquantes Informatives




---

---

---

---

---

---

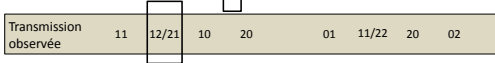
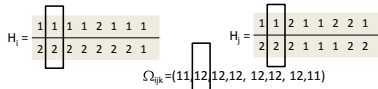
---

---

---

---

Utilisation de Zones Flanquantes Informatives




---

---

---

---

---

---

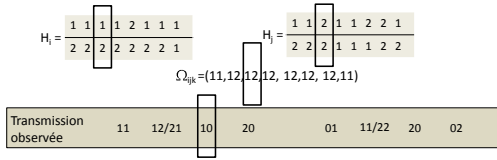
---

---

---

---

Utilisation de Zones Flanquantes Informatives




---

---

---

---

---

---

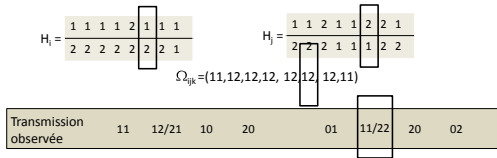
---

---

---

---

Utilisation de Zones Flanquantes Informatives




---

---

---

---

---

---

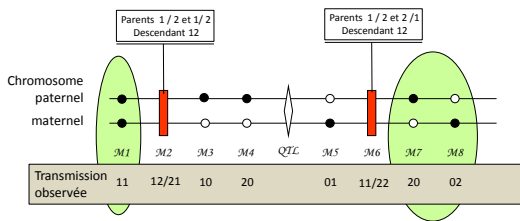
---

---

---

---

Utilisation de Zones Flanquantes Informatives



- Origine parentale connue
- Origine parentale inconnue

---

---

---

---

---

---

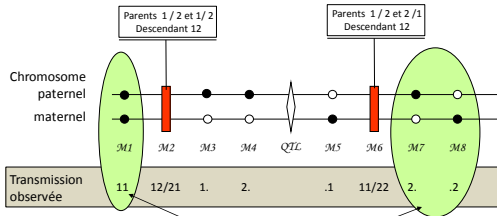
---

---

---

---

### Utilisation de Zones Flanquantes Informatives



#### Zone Flanquantes Informatives

Le calcul des proba de transmission  
N'utilise que les marqueurs compris  
entre les deux ZIF

- Origine parentale connue
- Origine parentale inconnue

Elsen et al, GSE, 2009

---

---

---

---

---

---

---

---

---

---