

# Cartographie des gènes et des caractères

Maxime Bonhomme

May 10, 2022

# Plan

- 1 Cartographie des gènes**
  - Quoi, pourquoi
  - Comment ?
  
- 2 Cartographie génétique des caractères quantitatifs**
  - Cartographie de Quantitative Trait Loci (QTL)
  - Cartographie par génétique d'association

# Les cartes: s'orienter dans le génome en positionnant les marqueurs

## Types

- **Cartes physiques** : distance réelle (Kb, Mb), à partir de fragments d'ADN. Résolution habituellement élevée.
- **Cartes génétiques** : s'appuie sur la recombinaison durant la méiose. Distance "statistique".

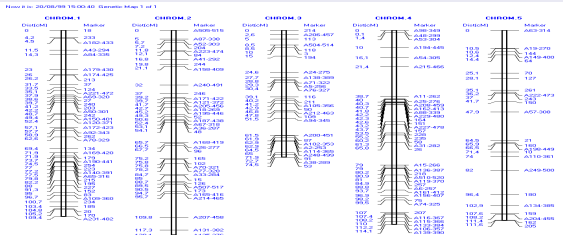
## Carte génétique

Représentation d'un génome positionnant un ensemble de repères (marqueurs) dont on connaît les positions sur des groupes de liaison (chromosomes idéalement).

Quoi, pourquoi

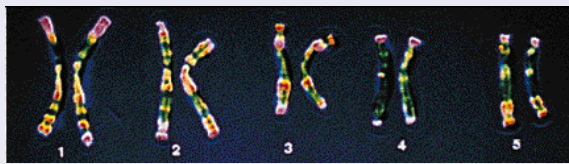
# Exemple

## Carte génétique



Groupes de liaison génétique

## Génome



Chromosomes

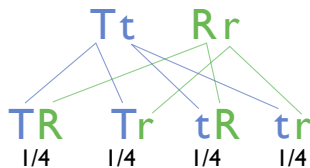
# Pourquoi

- Identifier les régions du génome influençant un caractère d'intérêt (maladie ou caractère quantitatif plus complexe)
- Positionner et identifier un gène (clonage positionnel)
- Comparer les génomes (étude de la synténie, évolution, transfert d'information)
- Faciliter la construction de cartes physiques, assemblage
- Étudier la méiose

# Les bases: lois de Mendel

## Loi de ségrégation indépendante

L'assortiment de plusieurs gènes dans une cellule sexuelle se fait de façon indépendante entre les différents gènes. Taille  $Tt$  et forme  $Rr$  (ridé).



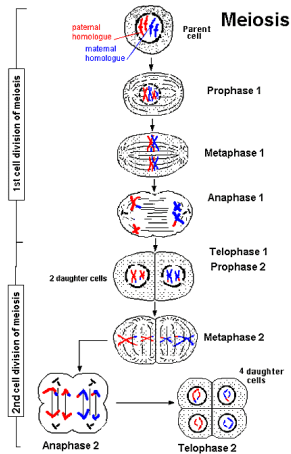
Quoi, pourquoi

# Le principe historique de la cartographie

## Liaison génétique (Bateson 1905)

Pour certaines paires de gènes, la fréquence des combinaisons parentales dans les gamètes est supérieure à ce que l'on attend. On parle de **liaison génétique**.

Expliqué par Morgan (1911) par l'appartenance à un même chromosome et un éventuel chiasma durant la méiose (crossing-over).







# Information génétique

- **Loci, gènes, marqueurs** :  $\mathcal{A}, \mathcal{B}$ . Emplacement sur un chromosome
- **Polymorphisme** : présente au moins deux formes différentes (allèles  $Aa$ ).
- **Génotype** : séquence des paires d'allèles portés par les chromosomes homologues ( $\frac{A}{a} \frac{B}{b}$  par exemple).
- **Homozygote** : paire d'allèles identiques ( $\frac{A}{A}$  ou  $\frac{a}{a}$ ), sinon hétérozygote ( $\frac{A}{a}$ ).
- **Haplotype** : séquence des allèles portés par chacun des chromosomes ( $\frac{Ab}{aB}$  par exemple).
- **Phase**: information des deux haplotypes à partir du génotype à  $\mathcal{A}$  et  $\mathcal{B}$ .

Comment ?

# Cartographie génétique

## Comment ?

- 1 Accumulation d'observations du génotype sur un ensemble de marqueurs et sur un bon nombre de méioses (beaucoup d'individus et/ou de générations d'individus)
  - Parents bien caractérisés (phase haplotypique).
  - Observation sur la descendance.
- 2 Estimer les taux de recombinaison entre marqueurs.
- 3 Calculer les distances génétiques entre marqueurs à partir de ces taux.
- 4 Ordonner les marqueurs à partir des distances.



Comment ?

## Estimer le taux de recombinaison

Taux de recombinaison  $\leq \frac{1}{2}$

Le taux de recombinaison  $\rho_{AB}$  entre les deux marqueurs  $A$  et  $B$  est la proportion de *recombinants*.

### Exemple

Entre 3 gènes  $\mathcal{Y}$  (yellow),  $\mathcal{W}$  (white),  $\mathcal{M}$  (miniature) de la drosophile, on observe  $\rho_{\mathcal{Y},\mathcal{W}} = 1.3\%$ ,  $\rho_{\mathcal{W},\mathcal{M}} = 32.6\%$  et  $\rho_{\mathcal{Y},\mathcal{M}} = 33.8\%$ . On peut penser que les marqueurs sont dans l'ordre  $\mathcal{Y} - \mathcal{W} - \mathcal{M}$

Du fait des doubles crossing-overs, pour un ordre  $\mathcal{Y} - \mathcal{W} - \mathcal{M}$  :

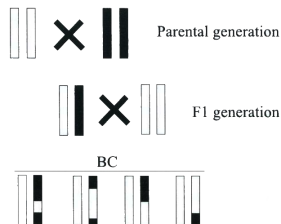
$$\rho_{\mathcal{Y},\mathcal{M}} < \rho_{\mathcal{Y},\mathcal{W}} + \rho_{\mathcal{W},\mathcal{M}} \quad (\text{non additif})$$

Comment ?

# Estimer $\rho$ entre 2 marqueurs: exemple avec le back-cross

## Un individu peut avoir

- deux marqueurs homozygotes ( $AA, BB$ ) ou hétérozygotes ( $Aa, Bb$ ) : **non recombinant (NR)**.
- un hétérozygote, un homozygote ( $Aa, BB$  ou  $AA, Bb$ ) : **recombinant (R)**.



## Taux de recombinaison

$$\rho = \frac{R}{R + NR}$$

Comment ?

# Distance génétique

## Définition

La distance génétique  $d_{AB}$  entre deux marqueurs  $A$  et  $B$  est le nombre moyen de *crossing-overs* entre les deux marqueurs par méiose.

## Propriétés

- Additif
- 1cM (centiMorgan) correspond à un crossover sur un haplotype pour 100 méioses.

## Fonction de Haldane - sans interférence (1919)

$$\rho = \frac{1}{2}(1 - e^{-2d}) \quad d = -\frac{1}{2} \log(1 - 2\rho)$$

Pour de faibles distances/taux de recombinaison,  $d \approx \rho$ .

Comment ?

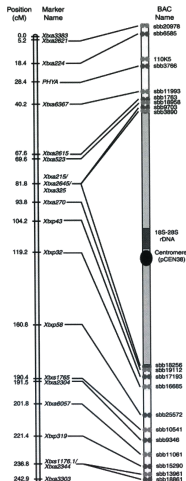
# Construction de la carte

## Cartographeur

Pour chaque groupe de liaison (partie de chromosome), déterminer l'ordre des marqueurs et les distances (taux de recombinaisons) qui séparent deux marqueurs adjacents

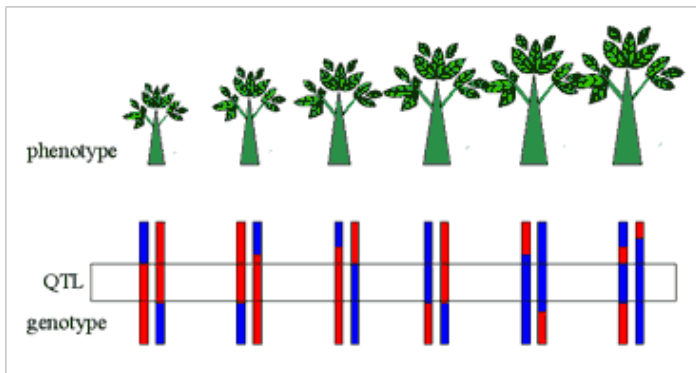
## Carte saturée

autant de groupes que de chromosomes, tous les marqueurs de la carte sont liés à un groupe.



## Que cherche-t-on à faire ?

⇒ lier un caractère binaire ou quantitatif (phénotype) à un endroit du génome (génotype au gène ou au QTL) **dans la descendance d'un croisement**, où les individus ont le même degré d'apparentement.





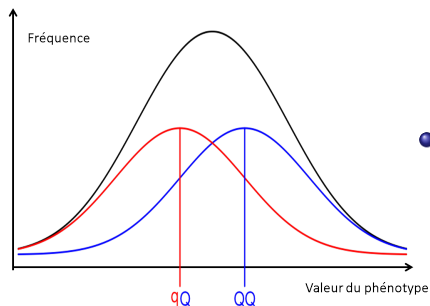
## Comment s'y prend-t-on ?

Nous allons nous restreindre

- 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**
- les marqueurs servent à connaître le génotype au QTL (par exemple: **qq** ou **Qq**)

## Cartographie de Quantitative Trait Loci (QTL)

## La distribution du phénotype (courbe noire, Gaussienne)



- 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

## paramètre génétique du back-cross

effet de substitution allélique

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

OU

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



## Que cherche-t-on à faire ?

⇒ lier un caractère binaire ou quantitatif (phénotype) à un endroit du génome (génotype au gène ou au QTL) **dans une population naturelle** où les individus ont des degrés d'apparentement très divers. Exploiter les événements de recombinaison historiques.

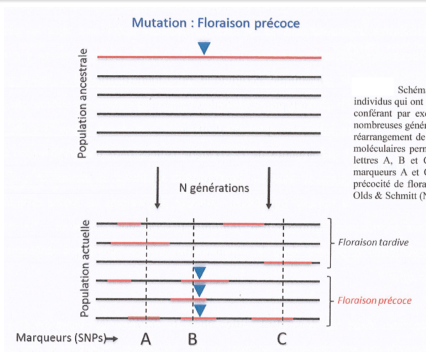
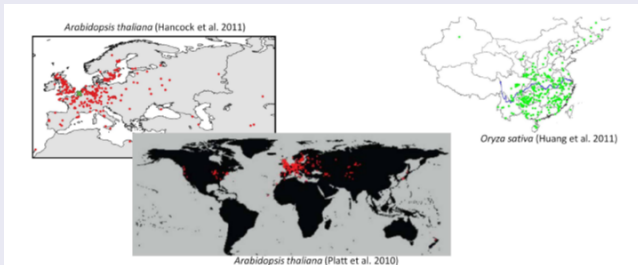


Schéma du principe du GWA mapping. Une population ancestrale contient plusieurs individus qui ont soit un génome rouge, soit un génome noir. Une nouvelle mutation (triangle bleu) conférant par exemple la précocité de floraison est localisée sur un génome rouge. Après de nombreuses générations de recombinaison, la population actuelle contient des génomes qui sont un réarrangement de petites régions génomiques dérivées de la population ancestrale. Les marqueurs moléculaires permettant de génotyper les individus de la population actuelle sont figurés par les lettres A, B et C, ainsi que les lignes verticales pointillées en gris clair. Contrairement aux marqueurs A et C, le marqueur B est en déséquilibre de liaison avec la mutation induisant la précocité de floraison et est donc associé à variation de la date de floraison. Adapté de Mitchell-Olds & Schmitt (Nature 2006).

# Genome-Wide Association Study (GWAS)

## Matériel et données nécessaires

- échantillon d'individus dans l'aire de distribution: forte richesse allélique !
  - populations structurées génétiquement du fait de l'échantillonnage large.
  - apparemment plus fort entre individus d'une même zone géographique.



# Genome-Wide Association Study (GWAS)

## Matériel et données nécessaires

- phénotyper chaque individu pour le caractère étudié (le caractère doit être suffisamment héritable)
- connaître le génotype de chaque individus à des milliers (millions) de marqueurs Single Nucleotide Polymorphism (SNPs) identifiés par séquence à haut-débit des différents individus.

Position sur le chromosome →

lignée	SNP 1	SNP 2	.....
HM4000	G	C	A
HM4001	G	G	C
HM4002	G	G	G
HM4003	G	G	G
HM4004	G	G	G
HM4005	G	G	G
HM4006	G	G	G
HM4007	G	G	G
HM4008	G	G	G
HM4009	G	G	G
HM4010	G	G	G
HM4011	G	G	G
HM4012	C	G	G
HM4013	G	G	A
HM4014	G	G	G
HM4015	C	G	G
HM4016	G	G	G
HM4019	G	G	G
HM4020	G	G	G
HM4021	G	G	G

# Genome-Wide Association Study (GWAS)

## Analyse statistique

- principe: le même que pour l'analyse QTL: comparaison de moyennes aux différents génotypes à un SNP. **MAIS** on doit tenir compte du biais d'association dû à la structure et à l'apparentement.
  - modèle linéaire incluant l'effet de génotype au SNP et l'effet de la population d'origine (pop1, pop2, ...).
  - modèle linéaire mixte incluant en plus une matrice d'apparentement qui permet de donner une valeur génétique variable d'un individu à l'autre.

# Genome-Wide Association Study (GWAS)

## Analyse statistique

- on effectue le test pour chaque SNP, et on récupère la p-valeur.
- on définit un seuil de significativité de la p-valeur pour rejeter  $H_0$  (en tenant compte des tests multiples, par exemple  $10^{-6}$  à  $10^{-8}$ ), pour dire qu'il y a (ou pas) une association significative entre le phénotype et le marqueur SNP.
- on peut visualiser les résultats des tests à l'échelle du génome avec un **Manhattan plot**.



# Genome-Wide Association Study (GWAS)

## Manhattan plot

- Consiste à représenter l'intensité d'association à chaque SNP le long du génome en utilisant  $-\log_{10}(p - \text{value})$ .
- Exemple: GWAS de la résistance quantitative de *Medicago truncatula* au parasite racinaire *Aphanomyces euteiches*.

