

La cartographie génétique















Thomas Faraut, Simon de Givry et Thomas Schiex

Laboratoire de génétique cellulaire
Laboratoire MIAT
INRA Toulouse

17 octobre 2013 / M2R informatique

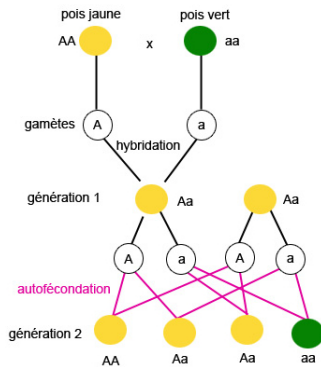
Les lois de Mendel : le concept de gène

Les lois de Mendel

Graine		Fleur	Cosse		Tige	
Forme	Cotylédons	Couleur	Forme	Couleur	Emplacement	Taille
						
Gris & lisse	Jaune	Blanc	Plein	Jaune	Cosse axiale Fleur tout du long	Long (~3m)
						
Blanc & Ridé	Vert	Violet	Étroit	Vert	Cosse terminales Fleurs en haut	Court (~30 cm)
1	2	3	4	5	6	7

- Les botanistes se sont longtemps interrogés sur les principes qui régissent la formation des hybrides végétaux.
- Gregor Mendel, en choisissant soigneusement les plantes et les caractères, parvient à établir des règles.

Les expériences de Mendel



Croisements effectués par Mendel

Phénotype parental	F_1	F_2	Rapport dans la F_2
1. Graines rondes × ridées	Toutes rondes	5474 rondes; 1850 ridées	2.96:1
2. Graines jaunes × vertes	Toutes jaunes	6022 jaunes; 2001 vertes	3.01:1
3. Pétales violets × blancs	Tous violets	705 violets; 224 blancs	3.15:1
4. Cosses gonflées × étroites	Toutes gonflées	882 gonflées; 299 étroites	2.95:1
5. Cosses vertes × jaunes	Toutes vertes	408 axiales; 152 terminales	2.82:1
6. Fleurs axiales × longues	Toutes axiales	681 axiales; 207 terminales	3.14:1
7. Tiges longues × courtes	Toutes longues	787 longues; 277 courtes	2.84:1

- La descendance ne se présente pas comme une combinaison intermédiaire des types parentaux
- Les relations de dominance dissimulent une distribution simple des allèles
- Elle présente les types parentaux selon des rapports simples (3:1 et 1:2:1) qui ne dépendent pas du sens de croisement

Croisements effectués par Mendel

Phénotype parental	F_1	F_2	Rapport dans la F_2
1. Graines rondes × ridées	Toutes rondes	5474 rondes; 1850 ridées	2.96:1
2. Graines jaunes × vertes	Toutes jaunes	6022 jaunes; 2001 vertes	3.01:1
3. Pétales violets × blancs	Tous violets	705 violets; 224 blancs	3.15:1
4. Cosses gonflées × étroites	Toutes gonflées	882 gonflées; 299 étroites	2.95:1
5. Cosses vertes × jaunes	Toutes vertes	408 axiales; 152 terminales	2.82:1
6. Fleurs axiales × longues	Toutes axiales	681 axiales; 207 terminales	3.14:1
7. Tiges longues × courtes	Toutes longues	787 longues; 277 courtes	2.84:1

- La descendance ne se présente pas comme une combinaison intermédiaire des types parentaux
- Les relations de dominance dissimulent une distribution simple des allèles
- Elle présente les types parentaux selon des rapports simples (3:1 et 1:2:1) qui ne dépendent pas du sens de croisement

Les lois de Mendel

- Mendel postule l'existence de 2 copies (allèles) par gène.
- Si l'on désigne par A le caractère dominant et a le récessif, la distribution des descendants s'exprime simplement.

Représentation symbolique

/	A	a
A	A/A	A/a
a	a/A	a/a

Les deux allèles d'un gène ségrègent lors de la formation des gamètes : une moitié des gamètes porte l'un des allèles et la moitié restante l'autre.

Les lois de Mendel

- Mendel postule l'existence de 2 copies (allèles) par gène.
- Si l'on désigne par A le caractère dominant et a le récessif, la distribution des descendants s'exprime simplement.

Représentation symbolique

/	A	a
A	A/A	A/a
a	a/A	a/a

Les deux allèles d'un gène ségrègent lors de la formation des gamètes : une moitié des gamètes porte l'un des allèles et la moitié restante l'autre.

Les lois de Mendel

- Mendel postule l'existence de 2 copies (allèles) par gène.
- Si l'on désigne par A le caractère dominant et a le récessif, la distribution des descendants s'exprime simplement.

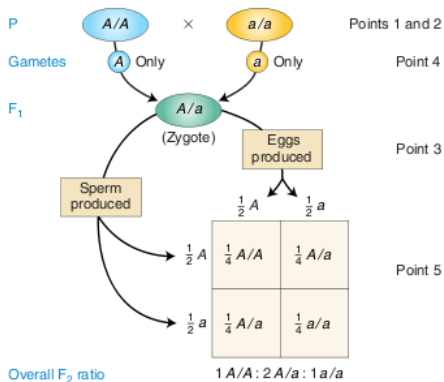
Représentation symbolique

/	A	a
A	A	A
a	A	a

Les deux allèles d'un gène ségrègent lors de la formation des gamètes : une moitié des gamètes porte l'un des allèles et la moitié restante l'autre.

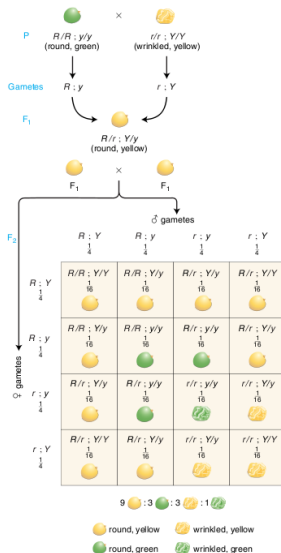
Les lois de Mendel

Interprétation mendélienne 1:2:1 du rapport 3:1



Les lois de Mendel

La ségrégation de plusieurs caractères



Mendel propose un formalisme facilitant le dialogue entre expérience et théorie et énonce les lois suivantes

- Ségrégation équitable des deux allèles d'un gène lors de la formation des gamètes
- Transmission indépendante des différents caractères

Théorie chromosomique de l'hérédité

Les chromosomes

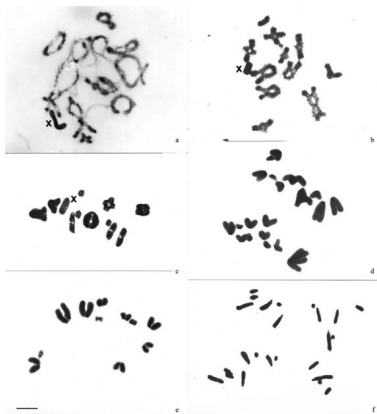
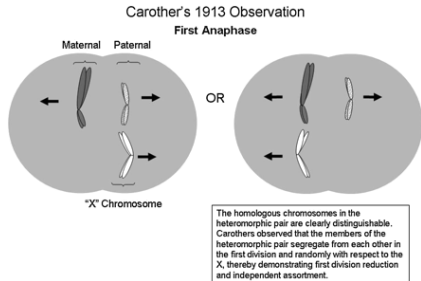


Figure 1 - Meiotic cells of *S. flavofasciata* (a,d) and *S. pallens* (b,c,e,f) after orcein staining: (a) pachytene, (b) diplotene, (c) metaphase I, (d) anaphase I, (e) metaphase II, (f) anaphase II. Bar = 10 μ m.

- Walter Sutton (1877-1916) et les chromosomes de sauterelles

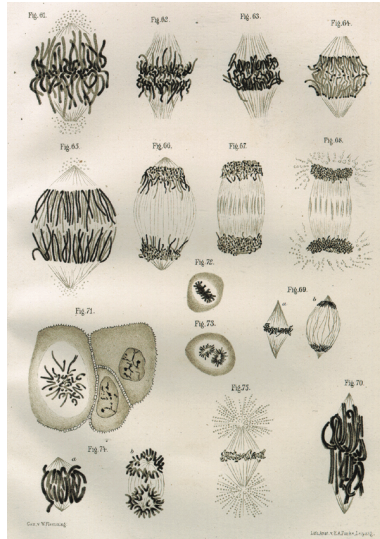
La méiose



Estella Eleanor Carothers (1883-1957)

- Le comportement des chromosomes à la méiose est semblable à la ségrégation indépendante des gènes énoncée par Mendel.

La méiose



Walther Flemming (1843-1905)

Expérience de Thomas Morgan

On considère chez la mouche drosophile les caractères couleur des yeux (pr pourpre, pr^+ rouge) et longueur des ailes (vg vestigial, vg^+ normal) et le croisement test (ou croisement en retour) suivant

$$P \quad pr^+/pr^+ \cdot vg^+/vg^+ \text{ ♀} \times pr/pr \cdot vg/vg \text{ ♂}$$

$$F_1 \quad pr^+/pr \cdot vg^+/vg$$

$$\text{Backcross} \quad pr^+/pr \cdot vg^+/vg \text{ ♀} \times pr/pr \cdot vg/vg \text{ ♂}$$

On s'intéresse à la constitution génétique des descendants que l'on va obtenir grâce à l'observation de leur phénotype.

Expérience de Thomas Morgan

On considère chez la mouche drosophile les caractères couleur des yeux (pr pourpre, pr^+ rouge) et longueur des ailes (vg vestigial, vg^+ normal) et le croisement test (ou croisement en retour) suivant

$$P \quad pr^+/pr^+ \cdot vg^+/vg^+ \text{ ♀} \times pr/pr \cdot vg/vg \text{ ♂}$$

$$F_1 \quad pr^+/pr \cdot vg^+/vg$$

$$\text{Backcross} \quad pr^+/pr \cdot vg^+/vg \text{ ♀} \times pr/pr \cdot vg/vg \text{ ♂}$$

On s'intéresse à la constitution génétique des descendants que l'on va obtenir grâce à l'observation de leur phénotype.

Experience de Thomas Hunt Morgan

Effectifs observés pour le croisement

$$pr^+ / pr \cdot vg^+ / vg \times pr / pr \cdot vg / vg$$

$pr^+ \cdot vg^+$	1339
$pr \cdot vg$	1195
$pr^+ \cdot vg$	151
$pr \cdot vg^+$	154
	<hr/>
	2839

Nous sommes loin des rapports 1:1:1:1 ou 1:1:0:0.
Que se passe-t-il ?

Experience de Thomas Hunt Morgan

Effectifs observés pour le croisement

$$pr^+ / pr \cdot vg^+ / vg \times pr / pr \cdot vg / vg$$

$pr^+ \cdot vg^+$	1339
$pr \cdot vg$	1195
$pr^+ \cdot vg$	151
$pr \cdot vg^+$	154
	<hr/>
	2839

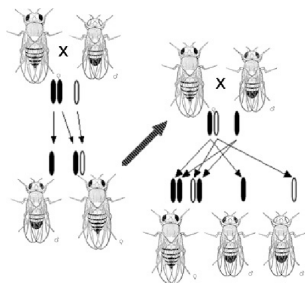
Nous sommes loin des rapports 1:1:1:1 ou 1:1:0:0.
Que se passe-t-il ?

Exercice

Nous allons, dans cet exercice, tenter de reproduire les analyses d'Alfred H. Sturtevant réalisées en 1913, sur la mouche *Drosophile*. Nous rappelons d'abord, dans la prochaine diapositive, la particularité de la liaison à un chromosome sexuel.

La liaison au sexe

Coségrégation d'un caractère avec un chromosome sexuel : la liaison à un chromosome sexuel



- Thomas Hunt Morgan ¹ (1866-1945) et la drosophile

La cartographie des chromosomes

La recombinaison

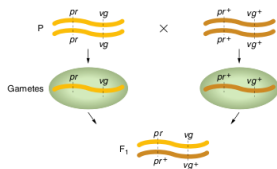
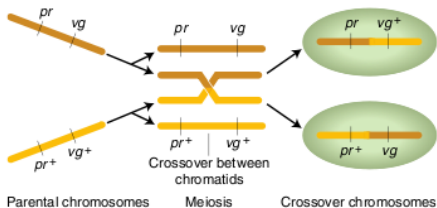


FIG. 64. Scheme to illustrate a method of crossing over of the chromosomes.

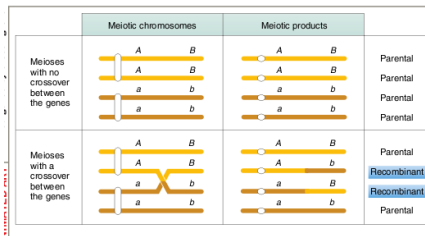
- Bien qu'étant sur le même chromosome, les configurations d'allèles (ici *pr*⁺ *vg*⁺ et *pr* *vg*) peuvent être recombinaison dans la descendance et donc donner ici des combinaisons *pr*⁺ *vg* et *pr* *vg*⁺. Ce phénomène s'appelle la recombinaison génétique.

La recombinaison

- La recombinaison résulte de l'échange de portions de chromatides entre deux bivalents homologues.

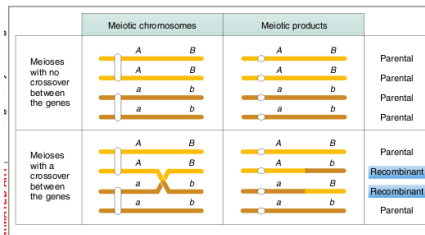


La liaison génétique



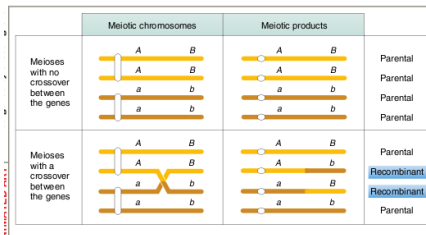
- La liaison reflète la synténie
- L'intensité de la liaison fournit une mesure de distance entre gènes

La liaison génétique



- La liaison reflète la synténie
- L'intensité de la liaison fournit une mesure de distance entre gènes

La liaison génétique



- La liaison reflète la synténie
- L'intensité de la liaison fournit une mesure de distance entre gènes

Estimer la fréquence de recombinaison

Calculons la fréquence de recombinaison sur l'exemple précédent des mouches. Pour simplifier les notions nous renommons les allèles :

$$A = pr^+, a = pr, B = vg^+, b = vg$$

n désigne l'effectif d'une classe et l'on pose

$$n = n_{AB} + n_{Ab} + n_{aB} + n_{ab}$$

La fréquence de recombinaison s'exprime simplement comme le nombre de descendants recombinants sur le nombre total :

$$\theta = \frac{n_{Ab} + n_{aB}}{n} = \frac{151 + 154}{2839} = 10,7\%$$

Croisement dihybride

Il n'est pas toujours possible d'estimer la fréquence de recombinaison en *comptant* les recombinants dans la descendance. Considérons l'exemple suivant d'un croisement dit dihybride

$$Ab/aB \times Ab/aB$$

Les gamètes recombinants sont AB et ab mais il n'est pas possible de les identifier avec certitude à partir des phénotypes :

- (dominance) Le phénotype AB peut résulter des 3 génotypes suivants Ab/aB , AB/aB ou AB/ab dont seuls les 2 derniers sont recombinants.
- (codominance) Le phénotype $ABab$ peut résulter des 2 génotypes suivants Ab/aB et AB/ab ou seul le dernier est constitué de gamètes recombinés.

On doit donc avoir recours à des méthodes d'estimation tels que le maximum de vraisemblance pour estimer les fréquences de recombinaison.

La fonction cartographique

On note I_{AB} l'intervalle qui sépare deux gènes A et B

- Le fréquence de recombinaison θ_{AB} fournit une mesure de la distance qui sépare A et B
- Cette mesure n'est cependant pas additive
- Le nombre moyen d'événements de recombinaison, par gamète, qui se produisent sur I_{AB} est utilisé comme mesure de la distance génétique
- La fonction cartographique permet de transformer la fréquence de recombinaison θ_{AB} en nombre moyen d'événements de recombinaison sur I_{AB}

La distance génétique

En première approximation, le nombre de crossing-overs peut se modéliser par un processus de Poisson. Si m désigne le nombre moyen de crossing-overs (chiasma) sur un segment, la loi du nombre de chiasma est donnée par

$$P(X = k) = \frac{m^k e^{-m}}{k!}$$

La formule de Mather

$$\theta = \frac{1}{2}(1 - P(X = 0)),$$

dont la démonstration est proposée en exercice [ici](#), permet d'obtenir la relation entre m et fréquence de recombinaison θ (la fonction cartographique):

$$\theta = \frac{1}{2}(1 - e^{-m})$$

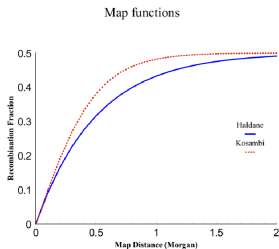
On remarque que l'on a en particulier $0 \leq \theta \leq \frac{1}{2}$.

La distance génétique

L'unité cartographique, appelée Morgan, est définie par $d = m/2$. C'est le nombre moyen d'événements de recombinaison par gamète.

La fonction cartographique, dite de Haldane, s'écrit

$$\theta = \frac{1}{2}(1 - e^{-2d}) \iff d = -\frac{1}{2} \ln(1 - 2\theta)$$



Pour de petites distances, la distance génétique coïncide avec la fréquence de recombinaison ($d \sim \theta$, sur un segment de longueur $1cM$ se produit en moyenne 1% de recombinants).

Estimation de θ

Si l'on peut compter les recombinants, n_r parmi n , la fréquence de recombinaison s'estime simplement par

$$\hat{\theta} = \frac{n_r}{n}$$

La probabilité d'observer n_r recombinants parmi n sur un intervalle où la proportion attendue de recombinant est θ s'écrit

$$L(\theta; n_r, n) = \theta^{n_r} (1 - \theta)^{n - n_r}$$

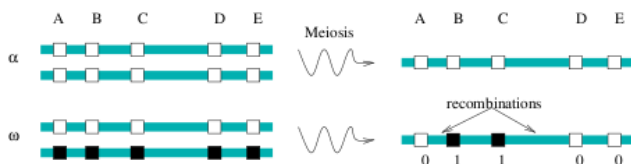
On l'appelle la vraisemblance des données. La valeur de θ qui maximise cette vraisemblance s'appelle l'estimation du maximum de vraisemblance

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}(L(\theta; n_r, n))$$

Pour un problème d'estimation moins trivial voir l'exercice.

Construction d'une carte génétique

On considère un backcross entre un individu α homozygote à tous les n loci et un individu ω hétérozygote à tous ces loci.



On peut alors coder par 0 et 1 les deux allèles de l'individu hétérozygote à chaque locus.

Les données

Soit K individus typés pour n marqueurs

	M_1	M_2	M_3	\dots	M_{n-1}	M_n
I_1	1	1	0	\dots	0	0
I_2	0	1	1	\dots	1	1
\vdots	\vdots	\vdots	\vdots		\vdots	\vdots
I_K	1	1	1	\dots	0	0

X_i^j allèle transmis par le parent hétérozygote à l'individu i au marqueur j . On note n_r^{st} la distance de Hamming

$$n_r^{st} = \sum_{i=1}^K |X_i^s - X_i^t|$$

L'estimation de la fréquence de recombinaison entre les marqueurs M_s et M_t s'écrit

$$\hat{\theta}_{st} = \frac{n_r^{st}}{K}$$

La vraisemblance des données

Un ordre π des marqueurs M_1, \dots, M_n est caractérisé par une permutation σ telle que²

$$\pi = (M_{\sigma^{-1}(1)}, M_{\sigma^{-1}(2)}, \dots, M_{\sigma^{-1}(n)})$$

pour simplifier les notations on note $j_k = \sigma^{-1}(k)$, donc

$$\pi = (M_{j_1}, M_{j_2}, \dots, M_{j_n})$$

et également simplement θ_{j_k} la fréquence $\theta_{j_k j_k + 1}$ et $n_r^{j_k}$ la distance de Hamming $n_r^{j_k j_k + 1}$.

La vraisemblance des observations pour l'ordre π s'écrit alors:

$$L(\pi, \theta; X) = \prod_{k=1}^n \theta_{j_k}^{n_r^{j_k}} (1 - \theta_{j_k})^{n - n_r^{j_k}}$$

2. Si M_k est en première position on a bien $\sigma(k) = 1$

La vraisemblance des données

Le logarithme de cette vraisemblance s'écrit

$$\log L(\pi, \theta; X) = \sum_{k=1}^n n_r^{j_k} \log(\theta_{j_k}) + (n - n_r^{j_k}) \log(1 - \theta_{j_k})$$

ou encore

$$-\log L(\pi, \theta; X) = \sum_{j=1}^n d_j$$

avec

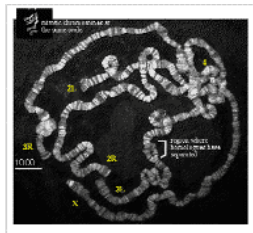
$$d_{j_k} = n_r^{j_k} \log(\theta_{j_k}) + (n - n_r^{j_k}) \log(1 - \theta_{j_k})$$

les contributions élémentaires positives de la vraisemblance.

Exercice : proposer une méthode pour trouver l'ordre qui maximise la vraisemblance des données.

Les cartes chromosomiques

Les premières cartes chromosomiques apparaissent.



The polytene chromosomes of *D. melanogaster*

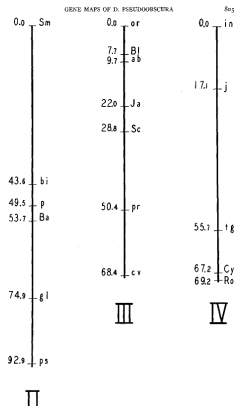
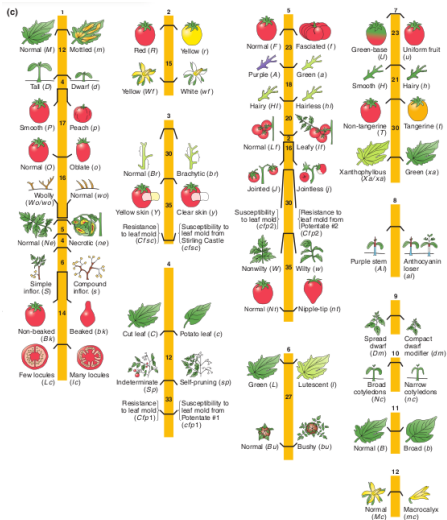


FIGURE 1.—Genetic maps of the three autosomes of *Drosophila pseudoobscura*. II—the second chromosome, III—the third chromosome, and IV—the fourth chromosome.

Les cartes génétiques

Une carte génétique du génome de la tomate



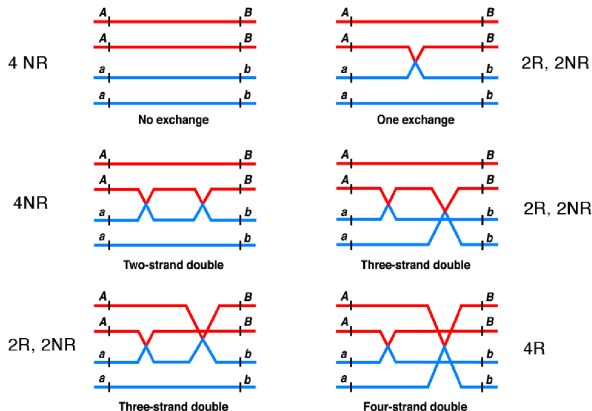
- On nomme gènes les unités élémentaires de l'hérédité
- Les gènes sont portés par les chromosomes
- La connaissance du patrimoine génétique des parents permet de prédire la distribution du patrimoine génétique de la descendance
- L'étude la descendance nous renseigne sur la composition du patrimoine génétique des parents
- En particulier, l'estimation de la recombinaison permet de cartographier les gènes sur les chromosomes

Lexique

- allèle : le variant d'un gène
- génotype : consitution génétique d'un individu. Pour un gène autosomique, le génotype est consitué des deux allèles que portent l'individu
- morgan : unité cartographique (1cM correspond à 1% de recombinants)
- phénotype : ensemble des caractères observables d'un individu
- ploïdie : caractérise le nombre d'exemplaire de chromosomes d'une cellule (un gamète est haploïde alors qu'une cellule somatique est généralement diploïde)
- synténie : littéralement sur le même fil. Deux gènes sont dīts synténiques s'ils sont sur le même chromosome.

Formule de Mather

Différentes configurations avec 0, 1 et 2 événements de recombinaison sur l'intervalle AB



Formule de Mather

Pour un segment chromosomique délimité par deux gènes A et B , on note X le nombre de crossing-overs sur le segment considéré et θ la fréquence de recombinaison pour les gènes A et B .

On fait les hypothèses supplémentaires suivantes :

- (i) Il n'y a pas de crossing over entre chromatides soeurs³
- (ii) Il n'y pas de phénomène d'interférence de chromatides (hypothèse NIC).⁴

3. Ceux de même couleur dans le schéma de la page précédente

4. Le choix d'une des deux chromatides soeurs pour un crossing-over n'influence pas le choix de la chromatide pour un crossing-over voisin. Il y a indépendance du choix de la chromatide.

1. Montrer que s'il se produit N ($N > 0$) crossing-overs sur l'intervalle AB , la probabilité qu'une chromatide donnée soit impliquée dans i crossing-overs est

$$\binom{N}{i} \left(\frac{1}{2}\right)^i \left(1 - \frac{1}{2}\right)^{N-i}$$

2. En déduire qu'en présence de N crossing-overs la proportion de gamètes recombinés est donnée par la formule suivante

$$P(\text{Recomb} | X = N) = \sum_{k=0}^{(N-1)/2} \binom{N}{2k+1} \left(\frac{1}{2}\right)^{2k+1} \left(1 - \frac{1}{2}\right)^{N-(2k+1)}$$

3. En déduire la formule de Mather (*Indication* : $\sum_{i=0}^n \binom{n}{i} = 2^n$ et $\binom{N}{i-1} + \binom{N}{i} = \binom{N+1}{i}$)