

Cours 2. 'Evolution des populations et des espèces'

Diversité intraspécifique : origine et description (PG : 5.1, 5.2)

1. Origine de la variabilité génétique
2. Diversité phénotypique : qualitative vs. quantitative
3. Typologie des polymorphismes

Quelle prise pour l'évolution ? (PG : 5.2, 5.3)

1. Eléments de génétique des populations
2. Dérive génétique et structure spatiale de la variation génétique
3. (Introduction à la théorie de la coalescence)
4. Introduction à la théorie neutraliste de l'évolution moléculaire
5. Mécanismes de l'évolution darwinienne

Conséquences (PG : 3.1, 3.3, 5.2)

1. La sélection en action
2. L'adaptation : un concept 'délicat'
3. Notion d'espèce et spéciation
4. (Phylogéographie et phylogénie)

Diversité intraspécifique : origine et description

Support et circulation de l'information génétique

⇒ Gène – locus

- Continuité des caractères à travers les générations avec lois de transmission
- mutation => changement éventuel de l'état du caractère
- allèles / génotype / polymorphisme => états homozygote et hétérozygote

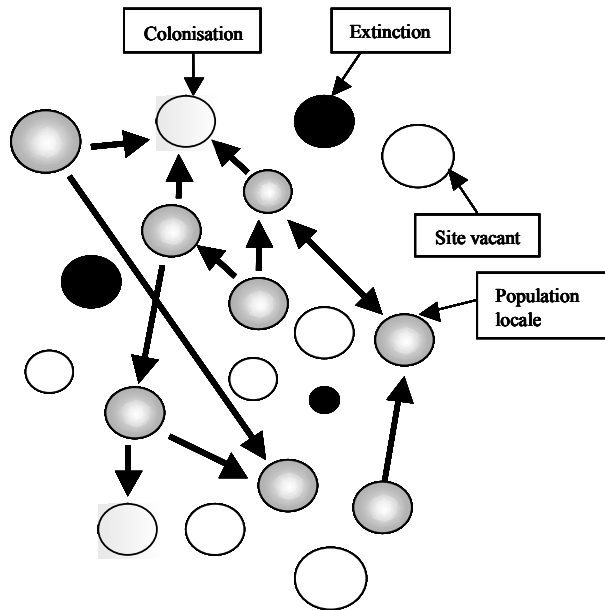
⇒ Populations

- communautés reproductrices partageant un même pool génique (T. Dobzhansky)
- Panmixie : union aléatoire des gamètes mais :
 - . Effet de voisinage / consanguinité de position
 - . Métapopulation
 - . polytypisme

⇒ Structure génétique

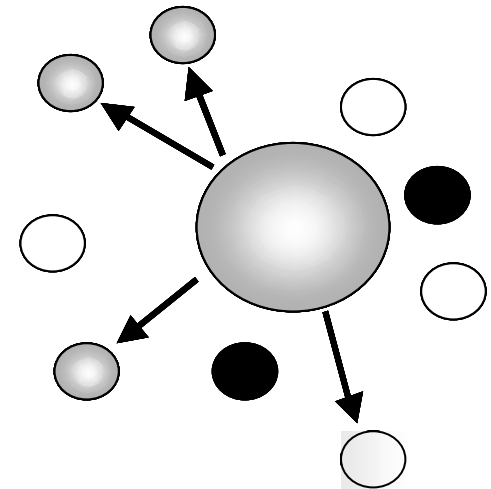
Métapopulation = réseau de populations locales (ou dèmes) interconnectées mais génétiquement différenciées

2 exemples :



Métapopulation 'classique'

=> Importance des extinctions-recolonisations locales

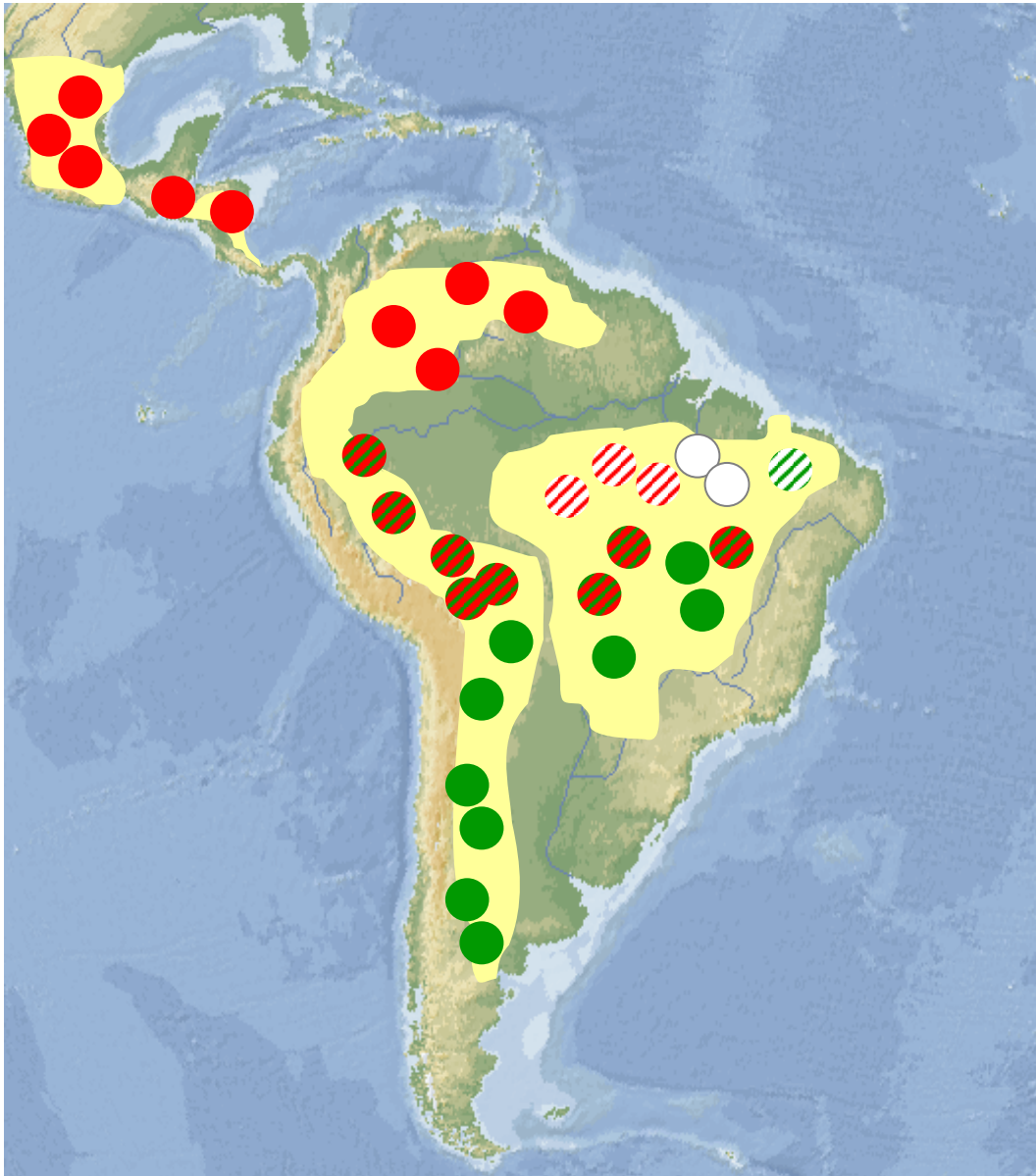


Modèles "île-continent" vs. "source-puit"

=> I/C : Analogie avec la théorie de la biogéographie insulaire

=> S/P : notion de qualité d'habitat (excédent démographique)

Changement d'échelle : notion d'espèce polytypique



-  Aire de répartition géographique
-  Population monomorphe
-  Population polymorphe



Vers la notion de race géographique
(= sous-espèce)

Description de la variation et définition(s) du polymorphisme

- ⇒ **Notion de biodiversité**
- ⇒ **polymorphisme : intérêt et limites**

Comment définir le polymorphisme (génétique) ?

« Cohabitation au sein d'une même population de 2 ou + catégories d'individus séparables par des caractères tranchés obéissant à un déterminisme mendélien (1 ou quelques gènes) »

« Présence simultanée en 1 même lieu de 2 ou + formes discontinues (morphes) d'une même espèce telle que la plus rare ne puisse être maintenue par les seules mutations récurrentes »

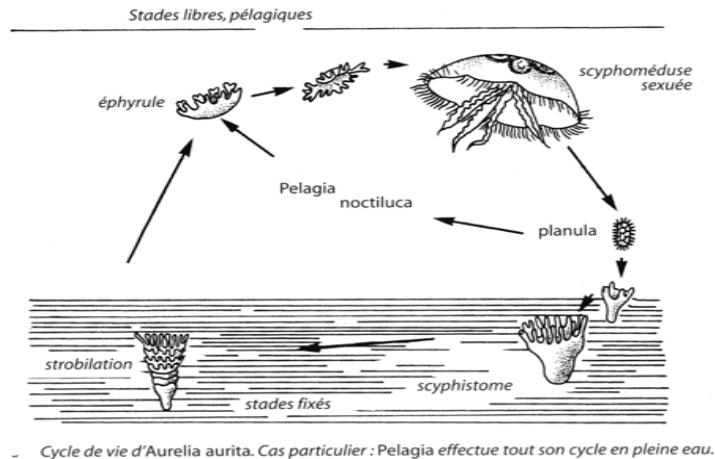
« gène présent dans un groupe sous la forme d'au moins 2 allèles, le plus rare étant présent à une fréquence supérieure à 0.01 (cf variants génétiques rares). »

Oui, mais....

variation héréditaire ('*nature*') vs. variation non héréditaire ('*nurture*')

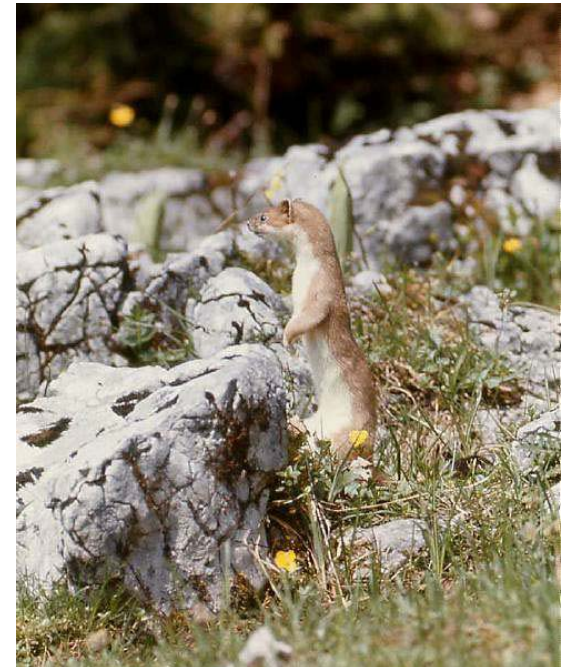
=> Notion de plasticité phénotypique

âge - cycle : méduses/polypes ; larve/imago



saisons : pelage, polyphénisme

=> 1 génotype produit plusieurs phénotypes



Condition sociale

Ex : castes termites



soldat



ouvrier



reproducteur

Conditions environnementales

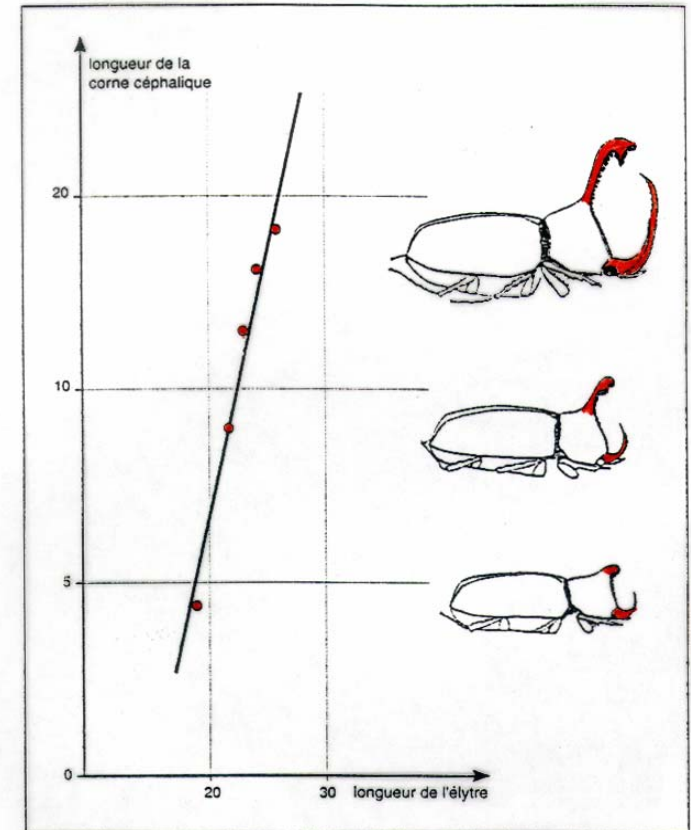
Exs : morphologie, couleur



Sagittaire

Allométrie : forme et physiologie

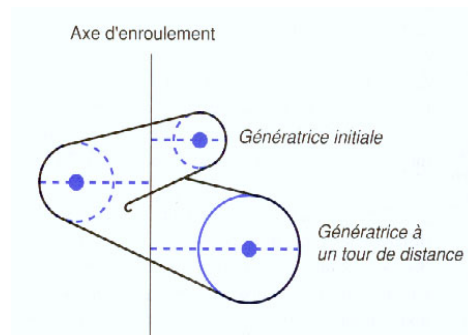
Du sexe....



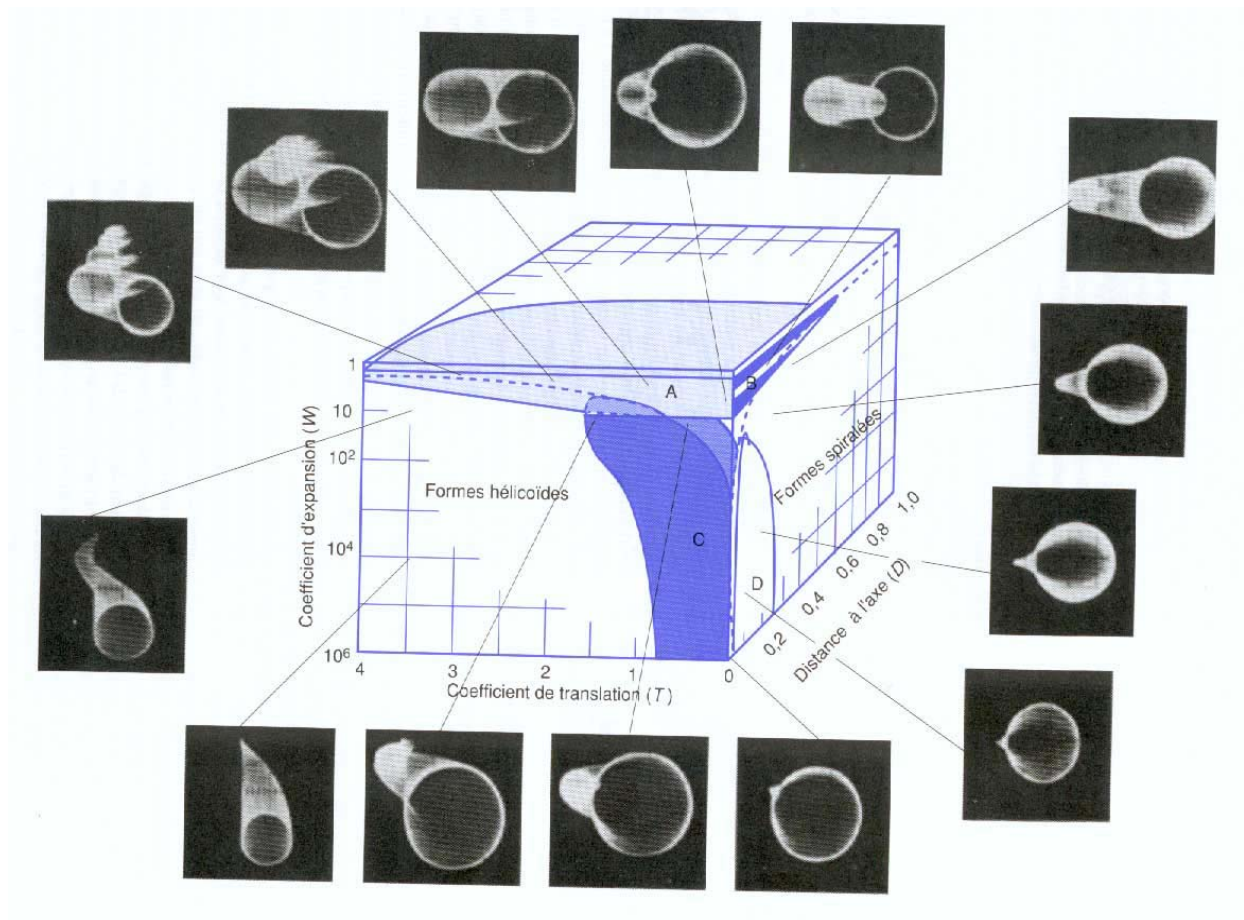
... Et des contraintes moins drôles

Haldane : « l'anatomie comparée reflète l'histoire de la lutte pour accroître son rapport surface/volume »

**Espace des formes définies
par le modèle de Raup :**



T = coef. de translation
W = coef. d'expansion
D = distance axe-ouverture



polymorphisme « génétique » vs. polymorphisme « statistique » : Deux génétiques des populations ?

Génétique 'qualitative'

- nature d'1 caractère (allèle A)
- variation discontinue : classes phénotypiques (AA, Aa, aa)
- gène détecté par son effet propre
- analyse de descendance par croisements individuels
- dénombrements et calcul de proportions

Génétique 'quantitative'

- degré d'1 caractère
- variation continue : gamme de phénotypes
- déterminisme polygénique : effets individuels des gènes trop faibles pour être détectés
- analyse de populations où tous les croisements sont possibles
- estimation de paramètres stats (moyenne, variance...)

Variation phénotypique : pas si simple de s'en sortir....

Aujourd'hui :

Canalisation : phénotype 'constant' $\forall$ les perturbations

$\Rightarrow$ **environnementale** : $V_P = V_A + V_{NA} + V_E$ ($\searrow V_E$)

$\Rightarrow$ **génétique** : $V_P = V_A + V_{NA} + V_E$ ($\searrow V_A$)

$\Rightarrow$ **dans les 2 cas : processus sélectionnés**

Stabilité de développement : phénotype constant malgré des irrégularités fortuites du développement

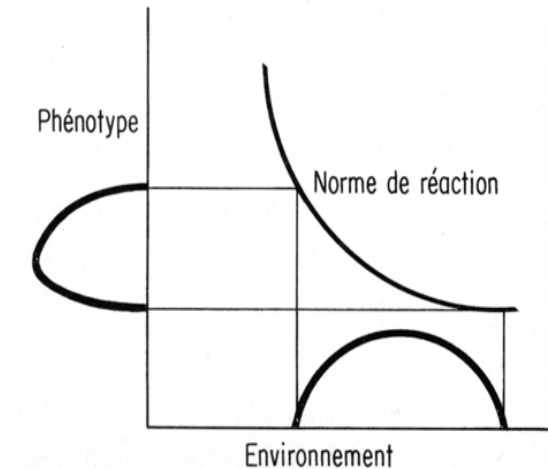
$\Rightarrow V_P = V_{inter} + V_{intra}(\leftrightarrow FA) (\searrow V_{intra})$

Plasticité phénotypique



$\Rightarrow$ **'contrainte'** (réponses continues vs. discontinues ?)

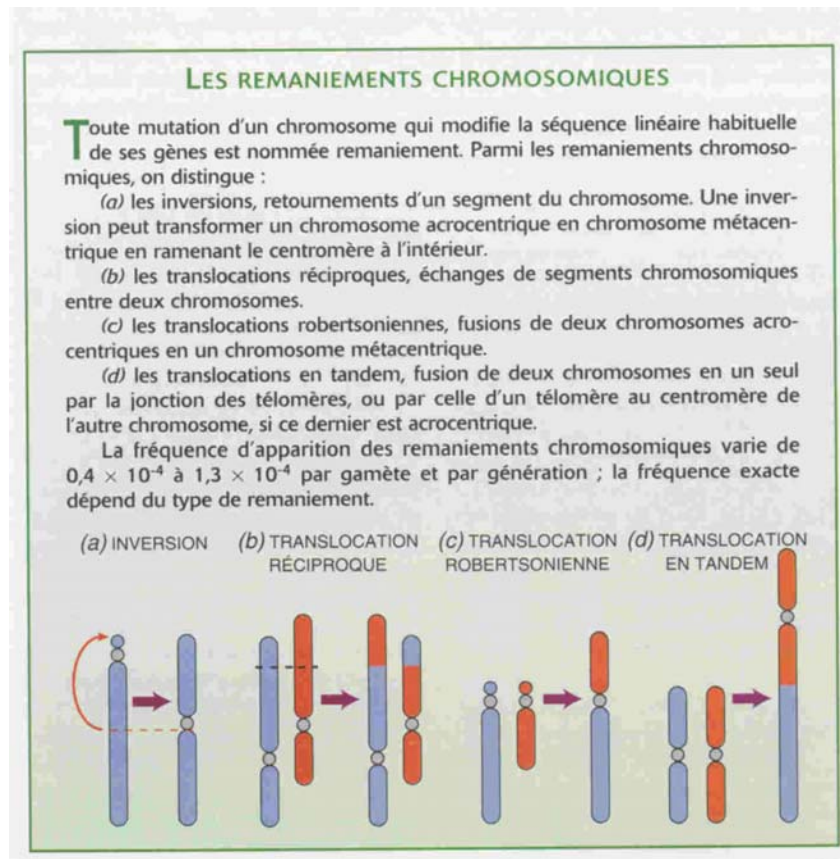
$\Rightarrow$ **adaptative** : $V_P = V_A + V_{NA} + V_{\mu E} + V_{ME}$
($\nearrow V_{ME}$)



Typologie des polymorphismes

1. Polymorphismes chromosomiques

- remaniements chromosomiques / mutations chromosomiques (gain, perte, réarrangement de segments)
- changement du nombre de chromosomes sans changement de structure : hétéroploïdie
=> aneuploïdie et euploïdie (aneuploïdie vs. polyploïdie) / mutations génomiques
- marquage chromosomique



Diversité phénotypique

- Polychromatisme



Cepaea nemoralis

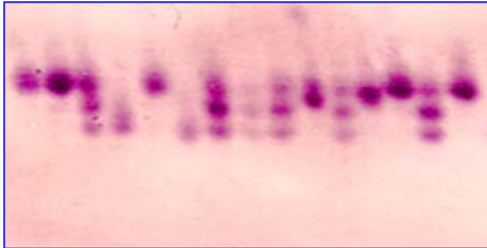


Biston betularia

- Polymorphisme morphologique

Diversité moléculaire

- Polymorphisme protéinique



**1 bond en avant pour la génétique des 70's...
... mais des limites 'incontournables'**

- Polymorphismes moléculaires : aujourd'hui, une solution technique à chaque problème

- RFLP, RAPD, AFLP
- microsatellites
- SNP

Éléments de génétique des populations

Hypothèse nulle de l'évolution (principe de Hardy-Weinberg) :
en l'absence de mutation dans une population d'effectif infini,
les fréquences alléliques restent indéfiniment constantes au
fil des générations.

1. Structure génétique des populations

➤ Structure génotypique en un locus

$$fr(AA) = \frac{n_{AA}}{N}$$

Fr(AA) fréquence des individus ayant le génotype AA

n_{AA} nombre d'individus ayant ce génotype

N nombre total d'individus de la population

➤ Structure allélique

Cas d'un locus diallélique A,a : 3 génotypes : AA, Aa, aa

- **Fréquence de l'allèle A :**

$$p = \frac{\text{nombre de gènes } A}{\text{nombre total de gènes}} = \frac{2n_{AA} + n_{Aa}}{2N} = fr(AA) + \frac{1}{2} fr(Aa)$$

- **Fréquence de l'allèle a :**

$$q = 1 - p = \frac{\text{nbre de gènes } a}{\text{nbre total de gènes}} = \frac{2n_{aa} + n_{Aa}}{2N} = fr(aa) + \frac{1}{2} fr(Aa)$$

- **Variance de la distribution des estimations (cf. loi binomiale) :**

$$\text{Var}(p) = \text{Var}(q) = pq/2N$$

Cas d'un locus multiallélique : A₁, A₂, A_i,...A_n

Fréquences alléliques : p₁, p₂, p_i,...p_n

$$p_i = \frac{2n_{A_i} + \sum_{j \neq i} n_{A_i A_j}}{2N} = f_{A_i A_i} + \frac{1}{2} \sum f_{A_i A_j}$$

Remarque :

Fréquences génotypiques $\Rightarrow$ fréquences alléliques

$$\text{Fr (AA)} = 0,30$$

$$\text{Fr (Aa)} = 0,50$$

$$\text{Fr (aa)} = 0,20$$

$$\Rightarrow \begin{aligned} \text{Fr (A)} &= 0,30 + \frac{1}{2} * 0,50 = 0,55 \\ \text{Fr (a)} &= 1 - 0,55 = 0,45 \end{aligned}$$

Fréquences alléliques $\nRightarrow$ fréquences génotypiques

$$\text{Fr (A)} = 0,50$$

$$\text{Fr (a)} = 0,50$$

$$\Rightarrow \begin{array}{lll} \text{Fr (AA)} = 0,50 & \text{ou} & \text{Fr (AA)} = 0 \\ \text{Fr (Aa)} = 0 & & \text{Fr (Aa)} = 1 \\ \text{Fr (aa)} = 0,50 & & \text{Fr (aa)} = 0 \end{array} \quad \text{ou ???}$$

2. Le cas idéal : le modèle de Hardy-Weinberg

G.H. Hardy & W. Weinberg (1908)

Hypothèses du modèle :

H1 : hypothèse de la **panmixie**

H2 : la population a une **taille infinie**

H3 : la fréquence des gènes n'est pas modifiée d'une génération à la suivante par **mutation, sélection ou migration**

Soit une population diploïde présentant en un locus 2 allèles **A** (p) et **a** (q) avec $p + q = 1$

⇒ **Fréquences des gamètes A = p et a = q**

Sous **H1** et **H2**, la génération diploïde suivante est :

		Mâles	
		<i>A</i>	<i>a</i>
<i>Allèle</i> <i>fréquence</i>		<i>p</i>	<i>q</i>
Femelles	<i>A</i> <i>p</i>	<i>AA</i> <i>p</i> ²	<i>Aa</i> <i>pq</i>
	<i>a</i> <i>q</i>	<i>Aa</i> <i>pq</i>	<i>aa</i> <i>q</i> ²

Carré de Punnet : les différentes aires sont proportionnelles aux fréquences génotypiques.

La structure génotypique de Hardy-Weinberg de la nouvelle génération produite est donc :



P^2 individus AA
 $2pq$ individus Aa
 q^2 individus aa

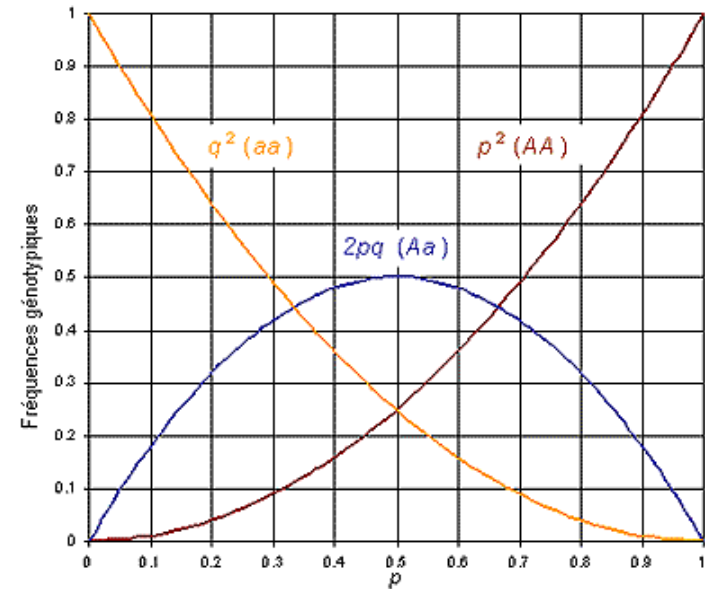
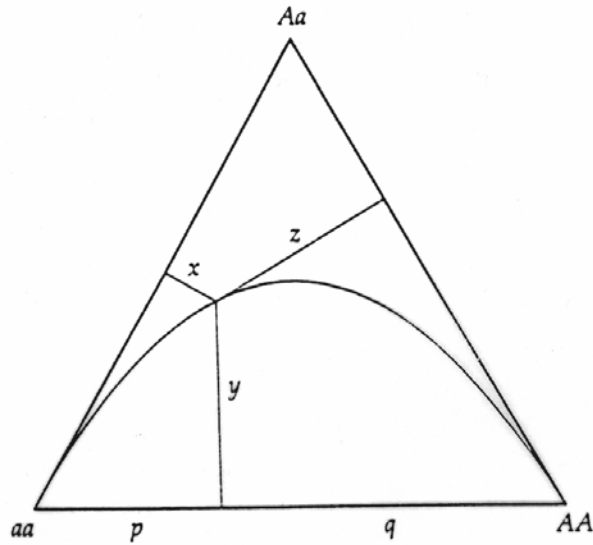
Selon **H3**, les gamètes que produira cette nouvelle génération contiendront :

⇒ Le gène A en fréquence $p' = p^2 + \frac{1}{2} (2pq) = p(p + q) = p$

⇒ Le gène a en fréquence $q' = \frac{1}{2} (2pq) + q^2 = q(q + p) = q$



**Les fréquences des gènes et des génotypes
ne varient pas d'une génération à l'autre**



Panmixie : relation fréquences alléliques – fréquences génotypiques dans un système de coordonnées triangulaires
=> 1 pop panmictique = 1 point d'équation
 $H = 1 - p^2 - q^2$

La panmixie confère à la population une structure génotypique caractérisée par $2pq$ hétérozygotes. L'un des moyens utilisés pour juger des écarts à la panmixie, sous l'influence d'un régime de reproduction particulier ou d'une pression évolutive, est fondé sur ce taux d'hétérozygotes attendus.

=> Tests de conformité à l'équilibre

Enoncé de la loi de Hardy Weinberg :

Dans une population isolée d'effectif illimité, non soumise à la sélection, et dans laquelle il n'y a pas de mutation, les fréquences alléliques restent constantes.

Si les accouplements sont panmictiques, les fréquences génotypiques se déduisent directement des fréquences alléliques selon la relation ci-dessous ; elles restent également constantes.

$$\begin{array}{ccc} AA & Aa & aa \\ p^2 & 2pq & q^2 \end{array} \quad (\Sigma = 1)$$

Remarque

Si plus de 2 allèles au locus considéré ($A_1, A_2, \dots, A_j, \dots, A_n$) en fréquence $p_1, p_2, \dots, p_j, \dots, p_n$,

↳ il existe $n(n + 1)/2$ génotypes différents et la structure de HW devient :

- Homozygotes $A_i A_i$ en fréquence p_i^2
- Hétérozygotes $A_i A_j$ en fréquence $2p_i p_j$
-

Et si les conditions ne sont pas respectées...

=> situation non panmictique

Exemple :

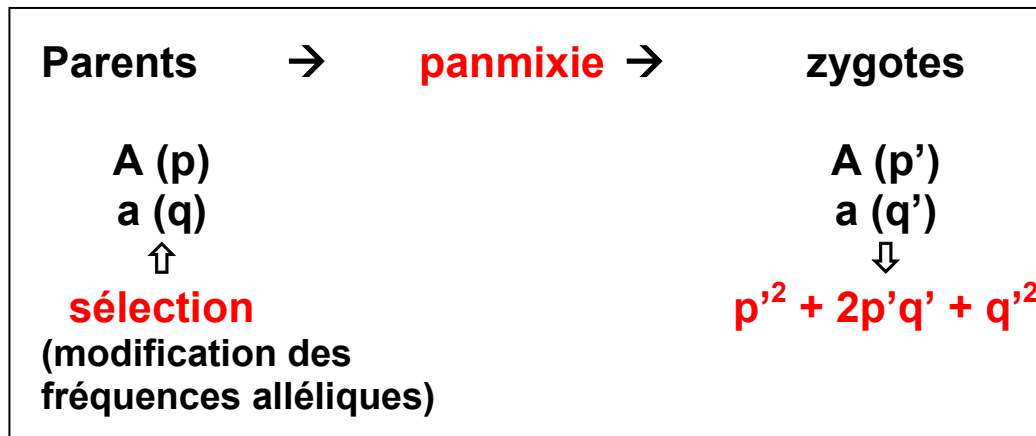
Population de bactéries (haploïde, multiplication végétative) : 2 allèles A1 et A2 à un locus.

Hypothèses :

- **absence de mutation de A1 $\Leftrightarrow$ A2**
- **absence de sélection**
- **absence de migration**
- **taille infinie de la population**

↪ pourtant, stabilité des fréquences alléliques

- Si panmixie : fréquences alléliques $\Rightarrow$ fréquences génotypiques



Conclusion : structure de HW décrit relativement bien la structure génétique des populations, même si les hypothèses ne sont pratiquement jamais vérifiées.

3. Influence du régime de reproduction

Régime de reproduction = modalité d'union des gamètes pour 1 gène donné pour former la génération suivante.

3.1. La Panmixie

H1 = association au hasard des gamètes ou panmixie.

↪ union au hasard des gamètes ou des couples relativement aux gènes portés par les conjoints au locus considéré.

⇒ pas de corrélation entre leurs génotypes

Le régime de reproduction dépend de la structure génotypique de la population comprenant $2pq$ hétérozygotes :

- Si taux d'hétérozygotes $< 2pq \Rightarrow$ régimes fermés ;
- Si taux d'hétérozygotes $> 2pq \Rightarrow$ régimes ouverts.

3.2. Les régimes fermés

3.2.1. L'autogamie

Reproduction par autofécondation :

- homozygotes **AA** $\Rightarrow$ **AA**
- homozygotes **aa** $\Rightarrow$ **aa**
- hétérozygotes **Aa** $\Rightarrow$ disjonction à chaque génération suivant la formule mendelienne :

AA	Aa	aa
1/4	1/2	1/4

proportion d'hétérozygotes divisée par 2 à chaque génération

↪ population tend rapidement vers l'homozygotie totale avec seulement des individus de génotypes **AA** et **aa**.

A la génération **g**, la proportion d'hétérozygotes est donc :

$$H_g = \frac{H_{g-1}}{2} = \frac{H_0}{2^g}$$

Qu'en est-il de la structure allélique ?

Un locus, 3 génotypes **AA** (D_g), **Aa** (H_g), **aa** (R_g) à la génération **g**. Les fréquences alléliques sont :

$$\mathbf{A} \quad p_g = D_g + \frac{1}{2}H_g = \left(D_{g-1} + \frac{1}{4}H_{g-1}\right) + \frac{1}{2}\left(\frac{1}{2}H_{g-1}\right)$$

$$p_g = \left(D_{g-1} + \frac{1}{2}H_{g-1}\right) = p_{g-1}$$

$$\mathbf{a} \quad q_g = (1 - p_g = 1 - p_{g-1}) = q_{g-1}$$



L'autogamie = constance des fréquences alléliques

≠ modification des fréquences génotypiques

↳ structure génotypique : $p\mathbf{AA} + q\mathbf{aa}$

En autogamie, la proportion des hétérozygotes dans le génome de chaque individu est divisée par 2 à chaque génération. On tend donc vers des individus homozygotes à tous les locus, ce qu'on appelle des **lignées pures**.

3.2.2. La consanguinité

unions entre apparentés = individus ayant un ou plusieurs ancêtres communs

individu issu d'apparentés = individu consanguin

Dans une population : unions consanguines $\Rightarrow$ déficit d'hétérozygotes ($< 2pq$)

Mesures de la consanguinité

2 mesures :

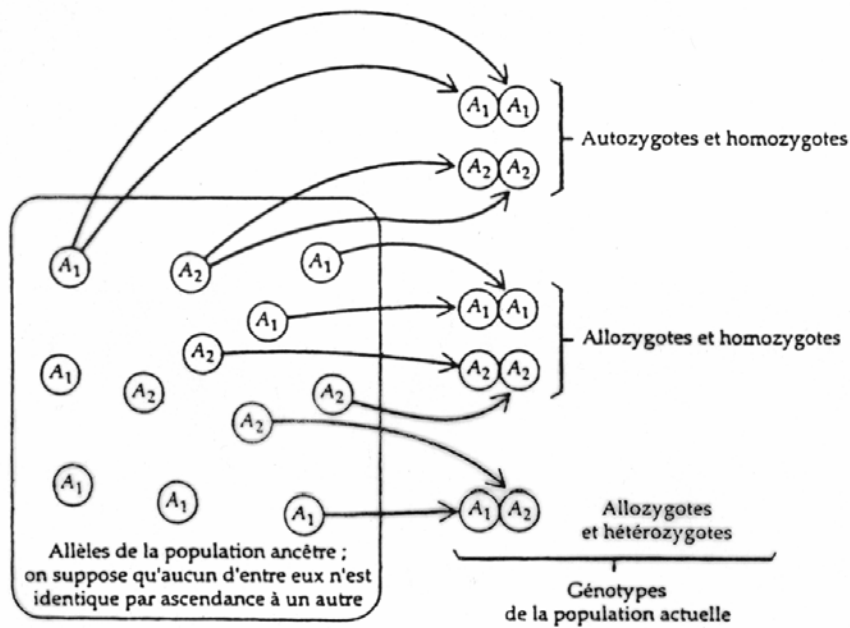
a. Coefficient de parenté

- de 2 individus (Malécot 1948) (ϕ_{PM}) : proba que 2 gènes homologues tirés au hasard chez I et chez J soient identiques
- proba que 2 allèles tirés au hasard d'1 population soient identiques

b. Coefficient de consanguinité d'un individu (f_i) : probabilité que 2 gènes homologues de l'individu I soient identiques

Remarque 1 : le coefficient de consanguinité de I est égal au coefficient de parenté de ses deux parents ($f_I = \phi_{PM}$).

Remarque 2



2 allèles (gènes homologues)
 $\Rightarrow$ identiques (par ascendance) = 2
 copies sans mutation d'un même gène
 ancêtre = gènes **autozygotes**
 $\Rightarrow$ non répliqués d'un seul gène
 ancestral = gènes **allozygotes**.

Un individu allozygote peut être soit
 homozygote soit hétérozygote.

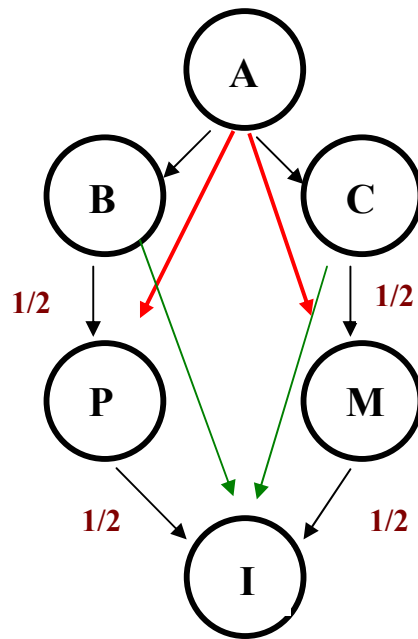
a. Approche probabiliste

PROBABILITE QUE LES DEUX GENES HOMOLOGUES DE L'INDIVIDU I SOIENT IDENTIQUES



généalogie des individus $\Rightarrow$ estimation des paramètres

Pour un individu I de parents issus d'un ancêtre commun A :



$n1 + n2$ maillons entre P, A, M

$n1, n2$: maillons de
B à I ou de C à I

⇒ **probabilité** pour qu'un gène tiré au hasard chez **P** et un gène tiré au hasard chez **M** proviennent de **A** est $(1/2)^{n_1 + n_2}$:

$$f_i = \phi_{PM} = (1/2)^{n_1+n_2} (1/2 + 1/2f_A)$$
$$f_i = (1/2)^{n+1} (1 + f_A) \quad (\text{avec } n = n_1 + n_2)$$

Exemple de coefficients de parenté :

Père-fils	$(1/2)^2 = 1/4$
Pleins frères	$(1/2)^3 + (1/2)^3 = 1/4$
Demi-frères	$(1/2)^3 = 1/8$
Cousins germains	$(1/2)^4 = 1/16$

b. En termes d'hétérozygotie (=> structure de Wright)

Population → structure génotypique quelconque :

génotypes	AA	Aa	aa
fréquences	D	H _o	R

Si **consanguinité** = seul facteur à influencer la structure génétique de la population, elle peut être exprimée par **l'indice de fixation F** (Wright, 1931) c-à-d écart entre fréquence d'hétérozygotes observés **H_o** et **2pq** :

$$F = (2pq - H_o) / 2pq \text{ soit } F = 1 - H_o/2pq$$

$$\Leftrightarrow H_o = 2pq(1 - F)$$

↳ description de la structure génétique de la population.

D, H et R connus ⇒ estimation de p et F (ou p et F connus ⇒ estimation de D, H et R)

$$AA \quad D = p^2 + Fpq = (1 - F)p^2 + Fp$$

$$Aa \quad H_o = (1 - F) (2pq)$$

$$aa \quad R = q^2 + Fpq = (1 - F)q^2 + Fq$$

En l'absence de panmixie : fréquences alléliques $\Rightarrow$ fréquences génotypiques.



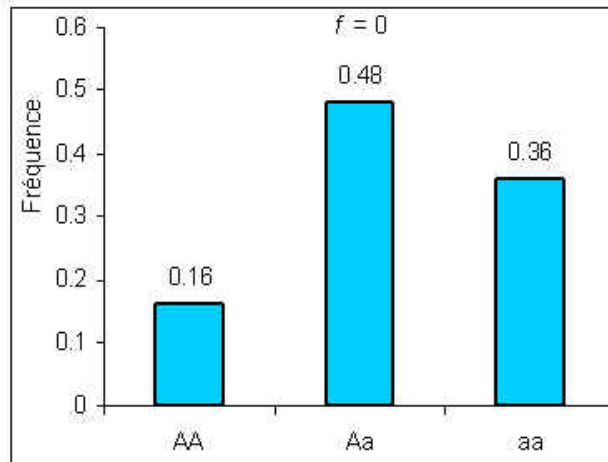
structure génotypique = $\sim$ fréquences alléliques
(p et q) + $\sim$ 'régime' d'association des allèles (F) :

Ainsi, si :

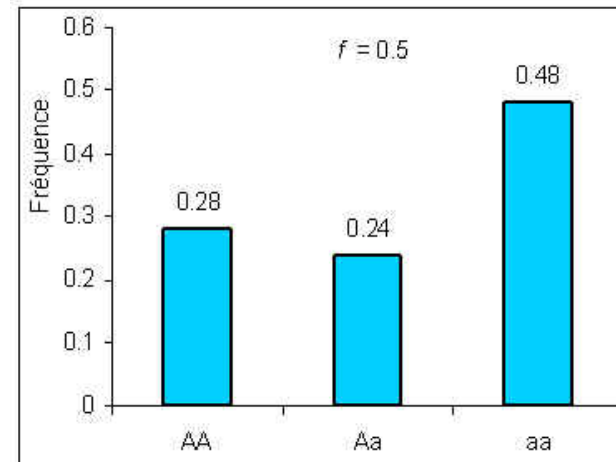
- $F > 0$: Hobs $<$ Hatt ; déficit en hétérozygotes ;
- $F < 0$: Hobs $>$ Hatt ; excès d'hétérozygotes ;
- $F = 0$: Hobs = Hatt ; population à l'équilibre.

Remarques :

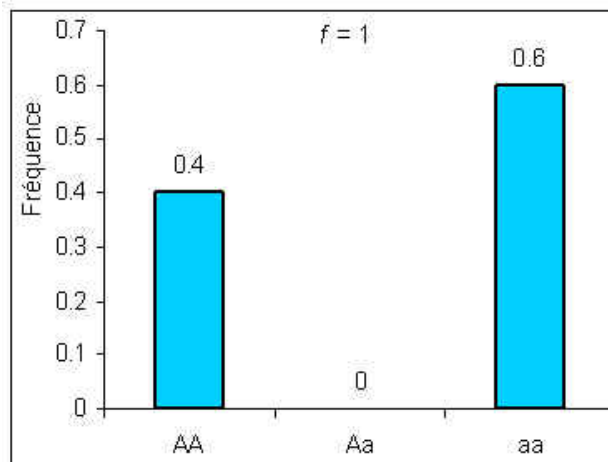
- Wright appelle F 'coefficient de consanguinité de la population', ce qui est abusif car, (i) la consanguinité est rarement le seul facteur provoquant un déséquilibre de la population par rapport à la structure de HW, (ii) un apparentement généralisé (populations finies) ne provoque pas d'écart à HW.
- F = coefficient de corrélation entre les gamètes qui s'unissent pour former le zygote ($0 < F < 1$).

A

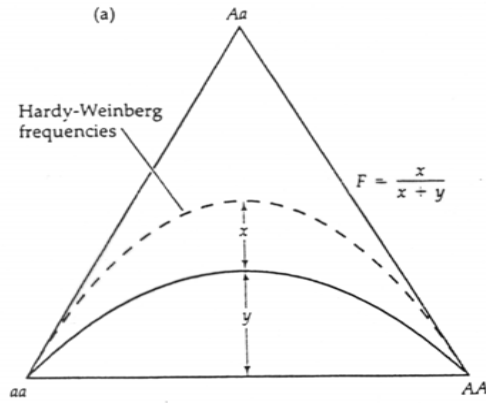
Absence de consanguinité, les fréquences génotypiques sont celles attendues sous l'équilibre de Hardy Weinberg

B

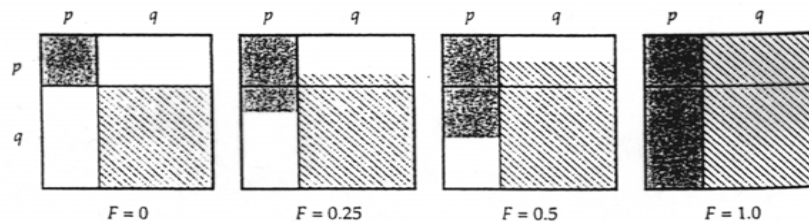
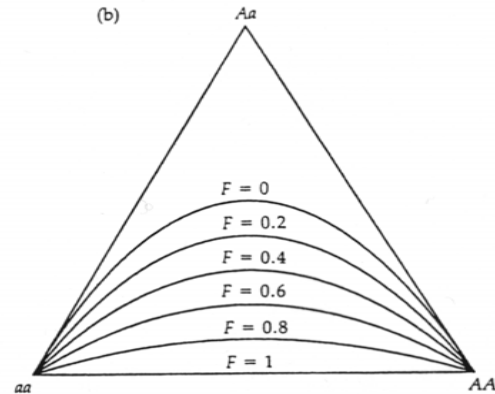
Avec un coefficient de consanguinité de f égal à 0.5, les fréquences génotypiques diffèrent des fréquences de HW => forte réduction du nombre d'hétérozygotes

C

Dans une population pratiquant l'autofécondation ($f=1$), on n'observe que des génotypes homozygotes de fréquences égales aux allèles qui les constituent.



Schémas des fréquences génotypiques dans des populations caractérisées par $\neq$ degrés de consanguinité



f(AA) : noir
f(Aa) : blanc
f(aa) : hachuré

3.2.3. L'homogamie

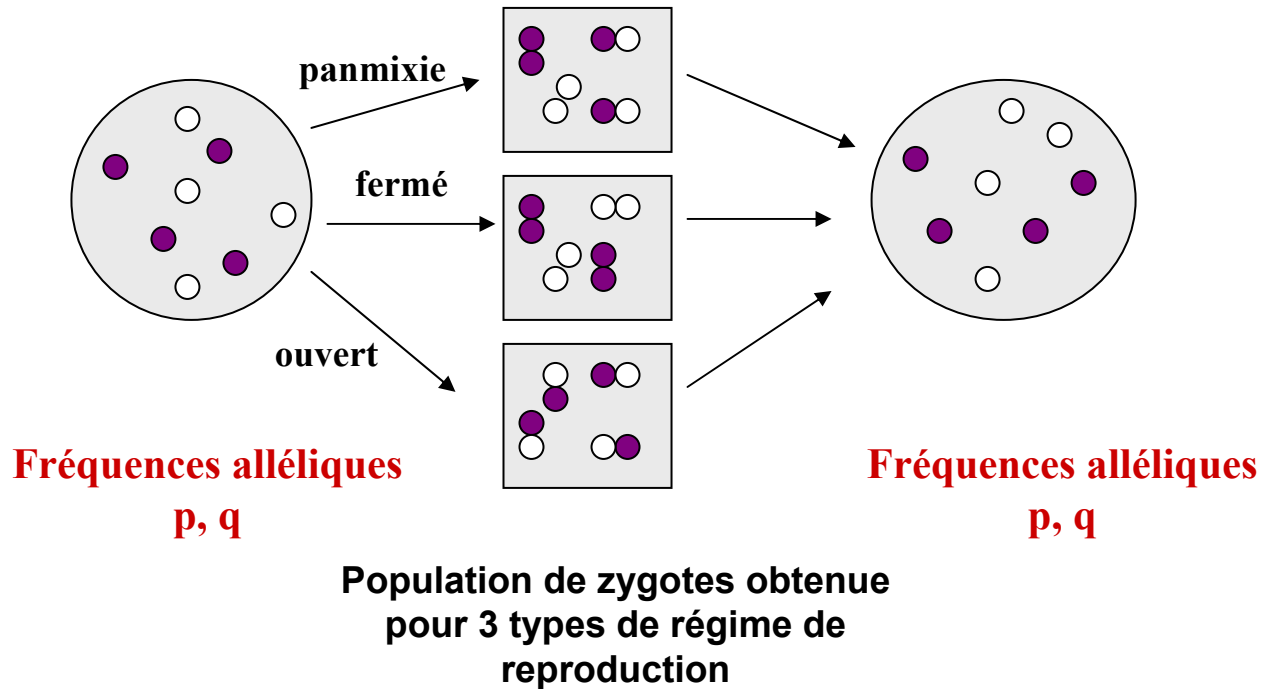
unions entre individus **phénotypiquement semblables** $\Rightarrow$ diminution du taux d'hétérozygotes à chaque génération.

3.3. Un régime ouvert : l'hétérogamie

unions entre individus **phénotypiquement dissemblables** (homogamie négative) $\Rightarrow$ taux d'hétérozygotes supérieur à $2pq$.

Population de gamètes
(génération $g - 1$).

Population de gamètes
(génération g).



Le régime de reproduction détermine l'association des gènes homologues dans les zygotes, mais il est sans effet sur les fréquences alléliques.

4. Influence des pressions évolutives

pression évolutive = action qui s'exerce sur la population en modifiant la structure allélique [**hasard**, sélection, **mutation**, **migration**].



4.1. L'anémie falciforme : l'avantage d'être hétérozygote

La drépanocytose (locus autosomal à 2 allèles) :

- Hb^S/Hb^S meurent avant âge adulte
- Hb^A/Hb^S viables (hématies normales et déformées)

$f(Hb^S)$ faible sauf populations africaines où $f(Hb^A/Hb^S) = 1 - 2\%$ [$f(Hb^S) = 10-15\%$]

Rmq : fréquence plus élevée dans régions touchées par le paludisme.

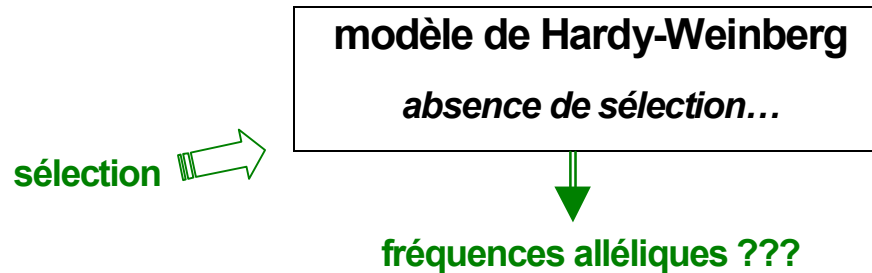
Comment expliquer une fréquence si élevée pour un allèle létal récessif ?

Hb^A/Hb^A + paludisme < **Survie** < Hb^A/Hb^S + paludisme

car parasitémie plus faible chez sujets atteints de drépanocytose (cycle érythrocytaire de *plasmodium* sp.).

⇒ **Cet avantage conféré aux hétérozygotes explique le maintien de l'allèle létal dans les populations.**

➤ Un modèle simple pour mesurer l'effet de la sélection



Population à l'équilibre :

AA	Aa	aa
p_0^2	$2 p_0 q_0$	q_0^2

Soit **W = valeur sélective** (*fitness*, coefficient de sélection absolu) caractéristique de chaque génotype (w_1, w_2, w_3), elle même décomposée en 2 autres paramètres :

- ✓ La **viabilité** (v) : la probabilité pour un zygote ayant ce génotype d'arriver à l'âge reproducteur ;
- ✓ La **fertilité** (f) : nombre moyen de zygotes qu'il laisse à la génération suivante.



$w = v.f$

✓ Si Δ viabilité et fertilité selon génotypes $\Rightarrow$ fréquences génotypiques :

AA	Aa	Aa
$w_1 \cdot p_0^2$	$w_2 \cdot 2 p_0 q_0$	$w_3 \cdot q_0^2$

La proportion des gamètes A produits est :

$$p' = \frac{w_1 p^2 + \frac{1}{2} \cdot w_2 \cdot 2pq}{\bar{w}}$$

où $\bar{w} = w_1 \cdot p^2 + 2w_2 \cdot pq + w_3 \cdot q^2$: valeur sélective moyenne de la population
(quantité moyenne de zygotes de la génération suivante)



variation de la fréquence de l'allèle A d'une génération à l'autre :

$$\Delta p = p' - p = \frac{w_1 p^2 + \frac{1}{2} \cdot w_2 \cdot 2pq}{\bar{w}} - p$$



$$\Delta p = pq \frac{(w_1 - w_2)p + (w_2 - w_3)q}{\bar{w}}$$

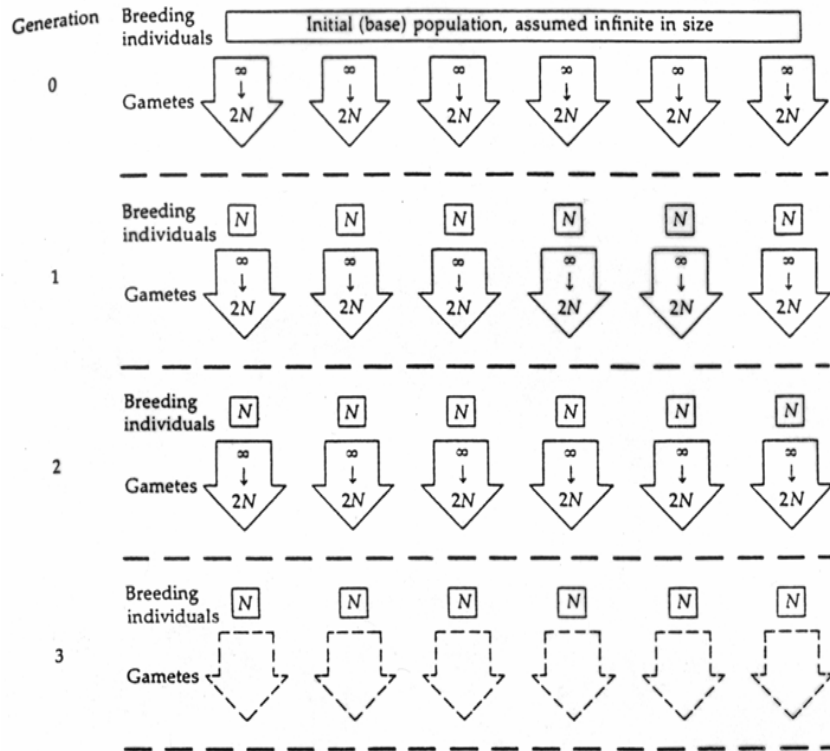
$W \Rightarrow \Delta p \text{ (+ ou -)} \Rightarrow \text{sens de l'évolution}$

- $w_1 > w_2 > w_3$: AA se reproduit davantage que Aa lui-même meilleur que aa. Δp est positif, l'allèle A favorable par rapport à a, augmente jusqu'à atteindre 1 (situation identique si $w_1 = w_2 > w_3$ ou si $w_1 > w_2 = w_3$).
- $w_1 \leq w_2 \leq w_3$; cas symétrique avec Δp négatif (fixation de l'allèle a) ;
- $w_1 > w_2 < w_3$ ou $w_1 < w_2 > w_3$; dans ces cas plus complexes, il peut exister un équilibre où les 2 allèles se maintiennent.



Ce modèle a l'avantage d'être simple mais les paramètres qu'il utilise - fréquences alléliques chez les zygotes ou les gamètes, valeurs sélectives - sont très difficiles à estimer dans les populations. Il faut des études démographiques précises pour estimer les viabilités et fertilités des différents génotypes....

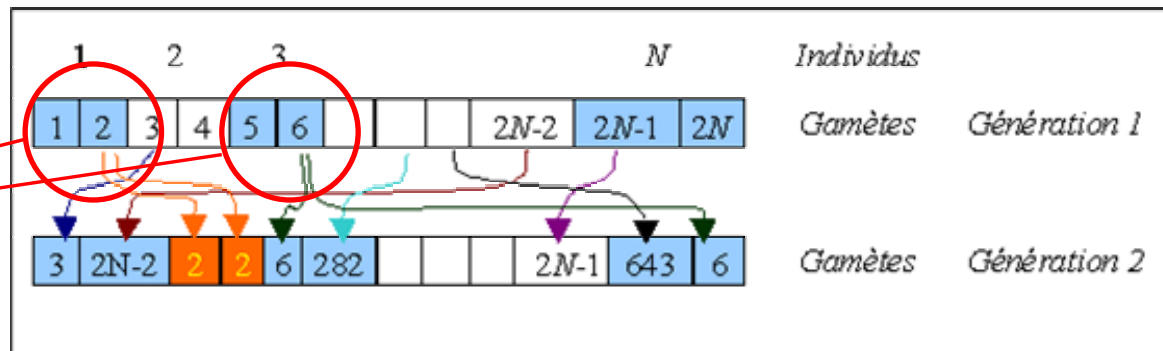
Modèle de dérive de Wright-Fisher



1. Organisme diploïde
2. Reproduction sexuée
3. Générations non recouvrantes
4. Sous-pops indépendantes de taille N constante
5. Accouplements au hasard
6. Pas de mutation
7. Pas de sélection

=> sous-populations 'idéales'

Tirage avec répétitions



Dérive : définition

- Tri aléatoire des individus sans relation avec la fitness
- Pop d'effectif fini => fluctuations fortuites des fréquences alléliques d'1 génération à la suivante (échantillonnage binomial / tirage d'un N réduit de gamètes) :
- Si p_0 = fréquence de A à $t=0$, nombre x_1 d'allèles A à la génération suivante = variable aléatoire de loi B ($2N$, p_0) :

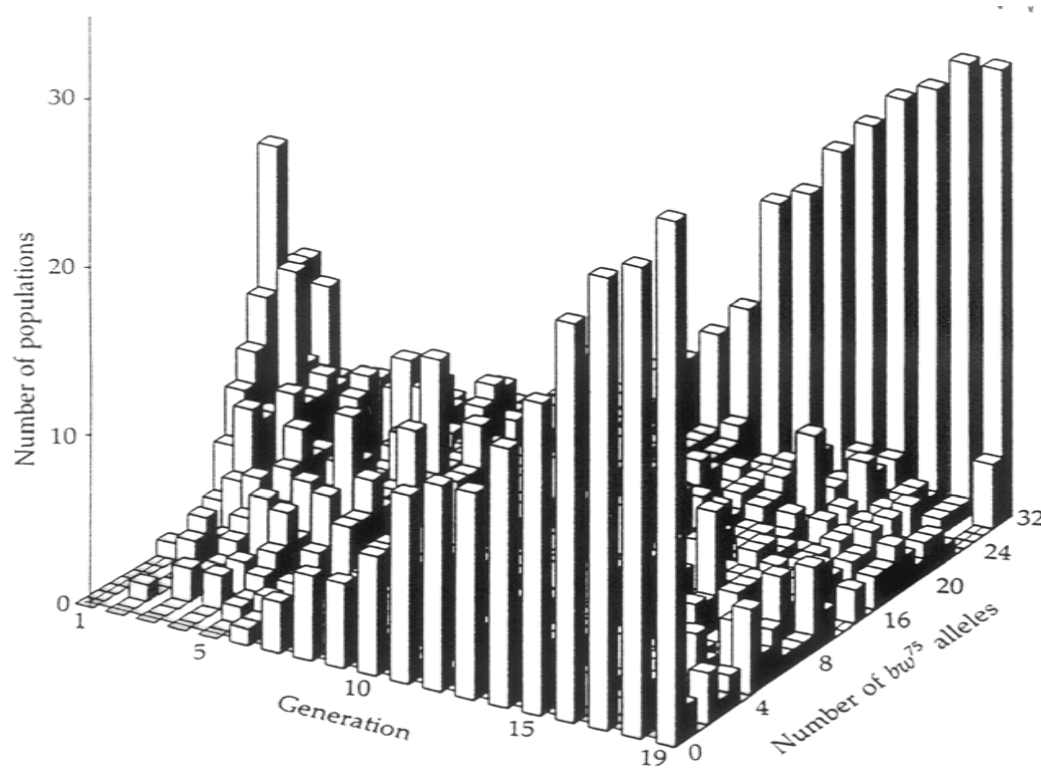
$$E(x_1) = 2N p_0$$

$$Var(x_1) = 2N p_0 (1 - p_0)$$

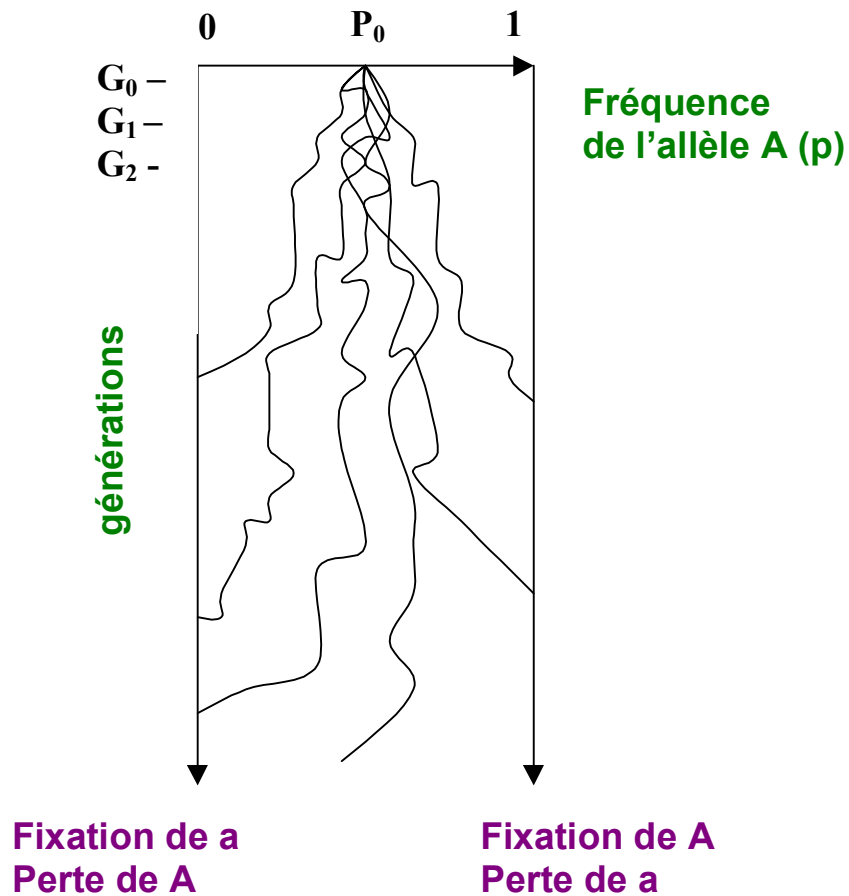
en proportion : $p_1 = x_1/2N$ (d'espérance p_0) de variance : $p_0(1-p_0)/2N$ (variance et hétérozygotie décroissent au même taux).

Dérive dans 107 pops de drosophiles de 16 hz pour le caractère 'œil marron'

⇒ 1 génération = 8 mâles + 8 femelles tirés au hasard



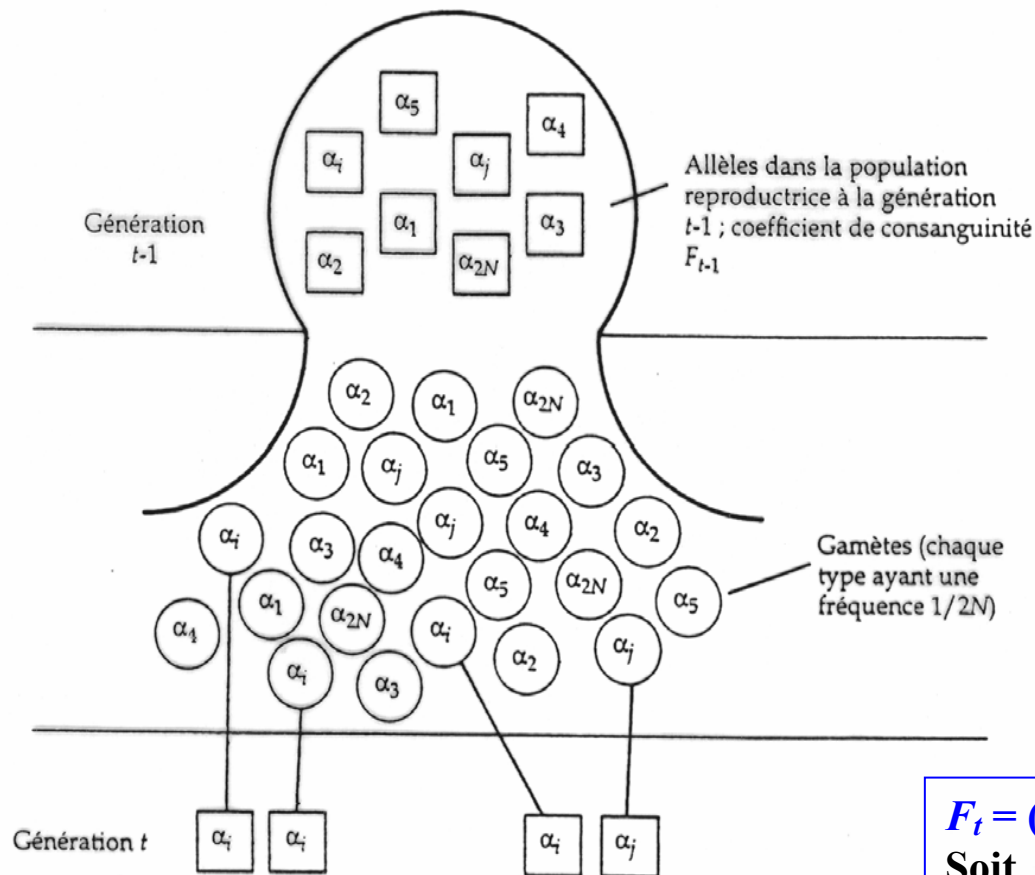
(expérience de Buri, 1956)



Fluctuations aléatoires de la fréquence d'un gène dans une série de populations ayant au départ la même fréquence allélique (p_0).

Dérive génétique et consanguinité

1 pop. panmictique mais d'effectif limité => proba F_t pour 2 allèles d'être identiques $\neq 0$



Proba : $1/2N$

$1 - 1/2N$

Contribution à F_t : 1

F_{t-1}

$$F_t = (1/2N) + (1-1/2N) F_{t-1}$$

Soit

$$(1 - F_t) = (1-1/2N) (1 - F_{t-1})$$

$$(1 - F_t) = (1-1/2N)^t (1 - F_0)$$

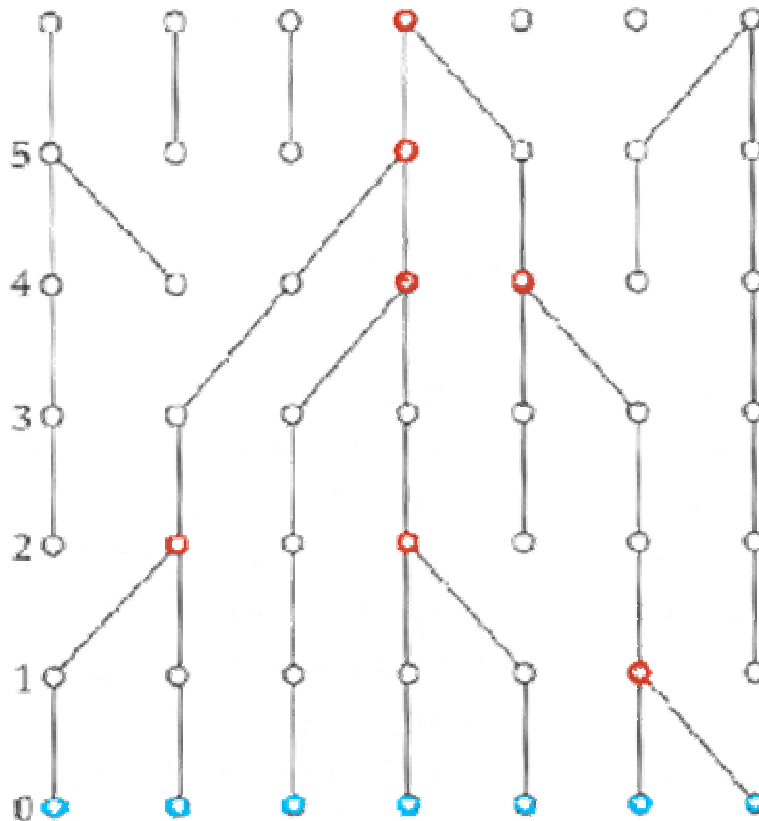
$$\text{ou, si } F_0 \text{ est nul : } F_t = 1 - (1-1/2N)^t$$

A terme : $F = 1$

=> si on remonte assez haut dans la généalogie, 1 seul gène ancêtre de tous les gènes présents à un locus dans une population donnée, car chaque événement de 'coalescence' réduit 2 lignées à une seule (cf « ève mitochondriale »).

La dérive détermine des généalogies de gènes : si chaque gène ne produisait à chaque génération qu'une seule copie, la généalogie serait faite de lignées parallèles : pas de dérive possible.

Dans la réalité : variation aléatoire du nombre de descendants



Théorie de la coalescence :

Reconstitution de l'histoire généalogique d'un échantillon de gènes jusqu'à l'ancêtre commun le plus récent

Effectif efficace (ou génétique) d'une population

➤ Dans une population idéale, chaque individu a la même probabilité de participer à la génération suivante.

⇒ N_e = effectif efficace d'une population réelle => effectif d'une population théorique ('idéale') soumise au même taux de croissance de F_t que la pop réelle

⇒ Différences => non respect d'une ou plusieurs conditions du modèle de dérive (fluctuations d'effectifs, nombres inégaux de mâles et femelles, structure d'âge, etc...)

➤ Exemple 1 : fluctuations d'effectif entre générations

Supposons que la taille d'une population passe de N_1 à N_2 en deux générations successives :

$$(1 - F_2) = (1 - 1/2N_2) (1 - F_1)$$

$$(1 - F_1) = (1 - 1/2 N_1) (1 - F_0)$$

$$\text{soit } (1 - F_2) = (1 - 1/2N_2) (1 - 1/2 N_1) (1 - F_0)$$

par analogie avec le cas où N constant, on écrit : $(1 - F_2) = (1 - 1/2 N)^2 (1 - F_0)$ où N représente cette fois la taille efficace, ce qui conduit à :

$$(1 - 1/2 N)^2 = (1 - 1/2N_2) (1 - 1/2 N_1)$$

avec une bonne approximation donnée par : $1/N = 1/2(1/N_2 + 1/N_1)$

$$\text{cas général : } 1/N_e = (1/t) (1/N_1 + 1/N_2 + \dots + 1/N_t)$$

→ importance des termes les plus petits (cf bottlenecks / effets de fondation)

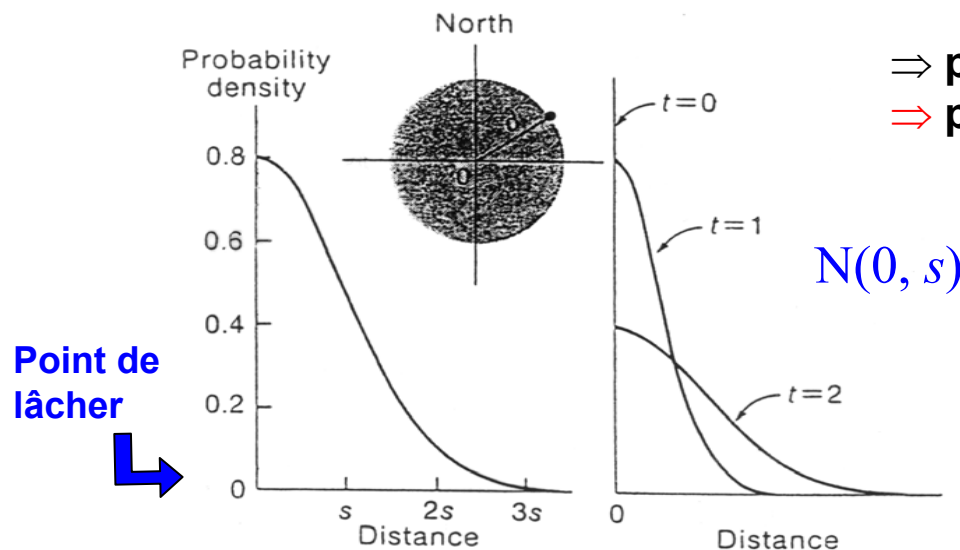
➤ Exemple 2 : Répartition uniforme / isolement par la distance

N_e (taille de voisinage) dépend :

- du nb de reproducteurs par unité de longueur ou de surface (δ)
- de la dispersion 'cumulée': variance σ^2 (s^2) de la distance entre lieux de naissance et de reproduction.

Si la dispersion suit une loi normale :

$$s^2 = (1/(N-1)) \sum d_i^2 \quad d_i^2 = (x_i - m_x)^2$$



⇒ pour une colonie linéaire : $N_e = 2\sqrt{\pi\delta} s(t)$

⇒ pour une surface : $N_e = 4\pi\delta[s(t)]^2$

Le **voisinage** correspond à la surface (reproducteurs inscrits dans un cercle de rayon $2s$) maximale sur laquelle il y a panmixie.

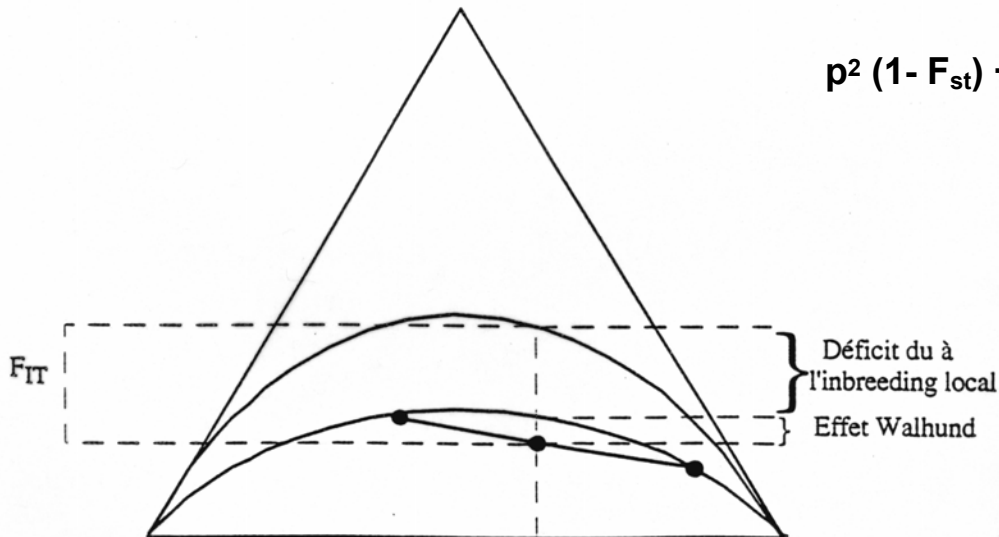
⇒ **Nombreuses limites au modèle**

Subdivision spatiale des populations

Effet Wahlund (rupture d'isolement)

	AA	Aa	aa
Sous-pop. 1	p_1^2	$2 p_1 q_1$	q_1^2
Sous-pop. 2	p_2^2	$2 p_2 q_2$	q_2^2
Moyennes :			
- sous-pops séparées	$(p_1^2 + p_2^2)/2$	$p_1 q_1 + p_2 q_2$	$(q_1^2 + q_2^2)/2$
- comme pop.totale	$p^2 (1 - F_{st}) + p F_{st}$	$2pq(1 - F_{st})$	$q^2 (1 - F_{st}) + q F_{st}$

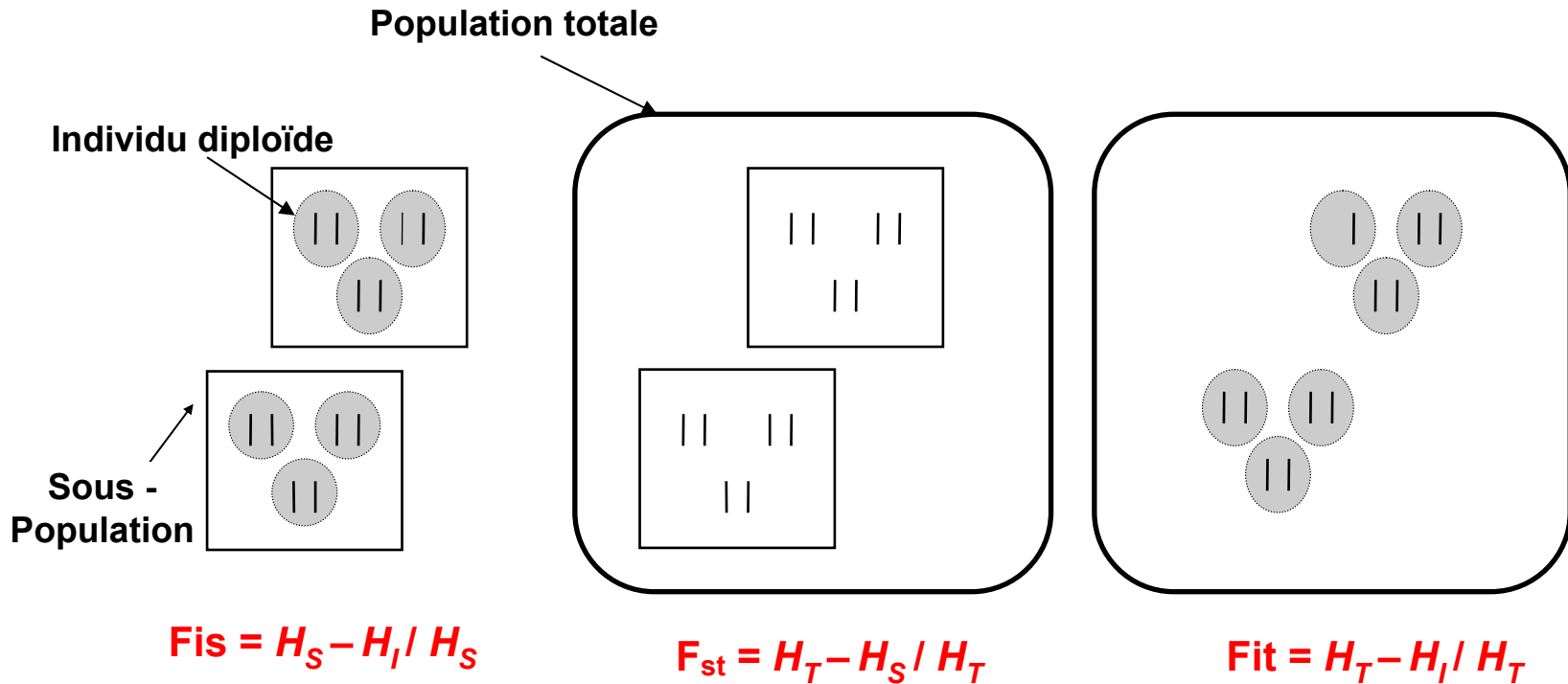
$$p^2 (1 - F_{st}) + p F_{st} = (p_1^2 + p_2^2)/2 \Rightarrow F_{st} = \text{Var}(p)/pq$$



Quand on pioche sans le savoir dans 2 dèmes génétiquement différenciés => déficit en Hz par rapport au résultat attendu dans la population panmictique 'fusionnée'. Ce déficit, dû à l'effet Wahlund, est égal à F_{st} .

Décomposition de l'hétérozygotie selon trois subdivisions hiérarchiques :

- individu (I)
- sous-population (S)
- population totale (T)



$$(1 - F_{it}) = (1 - F_{is}) (1 - F_{st})$$

Migration dans un modèle en îles

- Flux de gènes : transfert de matériel génétique entre populations (mouvements des individus ou de leurs gamètes = migration)
- La migration modifie les fréquences alléliques des populations 'receveuses'

=> difficulté d'une définition non arbitraire de la population

- modèle les plus simples = modèles en « îles »
 - p_i : fréquence d'un allèle dans la population étudiée ('île'),
 - p_0 : fréquence de l'allèle chez les immigrants (en proportion m par génération) :

$$p_{i,t+1} = (1-m) p_{i,t} + m p_0$$

'Au fil du temps', l'écart entre les deux pools diminue :

$$p_{i,t+1} - p_0 = (1-m) (p_{i,t} - p_0)$$

$$\Delta p_i = m(p_0 - p_i)$$

$$\text{Équilibre : } p_0 = p_i$$

- La migration (i) homogénéise les fréquences des pops qui échangent des gènes, (ii) limite la consanguinité :

- $(1 - m)^2$ = proba de tirer 2 allèles non 'migrants' d'où :

$$F_t = [1/2N + (1 - 1/2N) F_{t-1}] (1 - m)^2$$

A l'équilibre : $F_t = F_{t-1}$, d'où, si m suffisamment petit (m^2 négligeable devant m) :

$$F_{st} = 1 / (4Nm + 1)$$

Distance génétique de Nei (1972)

⇒ divergence entre 2 populations = fonction des fréquences alléliques

⇒ signification évolutive

⇒ Fondée sur le calcul préalable de l'identité normalisée ou **identité génétique I** de Nei.

Pour 2 populations A et B et un gène K à i allèles différents :

$$I_k = \sum a_i b_i / \sqrt{(\sum a_i^2 \sum b_i^2)}$$

- I_k : proba que 2 allèles pris au hasard dans chacune des 2 pops soient identiques / proba que 2 allèles pris ds la même population le soient (proba normalisée).

- Pour estimer la différenciation, plusieurs locus considérés => identité génétique des 2 populations fondée sur le calcul des moyennes arithmétiques de chacun des termes précédents :

$$I = I_{ab} / \sqrt{I_a \cdot I_b}$$

$$\text{Distance standard : } D = - \ln (I)$$

⇒ D = nb moyen de substitutions alléliques par locus depuis la séparation des 2 populations sous un modèle mutation-dérive avec un nb infini d'allèles

⇒ $E(D) = \mu t$ => D linéairement proportionnelle au temps de divergence

⇒ limites (horloge, non-métrie)....