



Génétique des populations

Master 1 Biologie Végétale (parcours BPMA) et Master 1
Bioinformatique (parcours BBS)

Maxime Bonhomme

Laboratoire de Recherche en Sciences Végétales



Génétique des populations

Introduction

- qu'est-ce que la génétique des populations ?
- quelques définitions
- mesurer et décrire la variation génétique
- facteurs influençant la variabilité génétique

Loi de Hardy-Weinberg

- principe
- test de conformité

Régimes de reproduction

- régimes fermés
- régimes ouverts

Forces évolutives

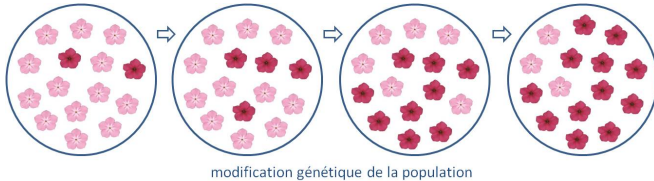
- mutation
- dérive génétique
- sélection naturelle
- migration

Conclusions

References



Qu'est-ce que la génétique des populations ?



Définition

- Etude de la distribution et des changements de la fréquence des versions d'un gène (allèles) dans les populations d'êtres vivants, sous l'influence des pressions évolutives (sélection naturelle, dérive génétique, mutations, et migration).

Epistémologie

- Application des principes fondamentaux de la génétique mendélienne à l'échelle des populations.
- Synthèse entre la génétique mendélienne et la théorie de l'évolution = néo-darwinisme (théorie synthétique de l'évolution).



Qu'est-ce que la génétique des populations ?

Objectifs

- Décrire les **génotypes**, **estimer** leur **fréquence** et celle des **allèles**, déterminer leur **distribution** au sein des individus, des populations, et entre les populations (**descriptif**).
- **Comprendre et prédire** l'évolution des fréquences des allèles dans les populations sous l'effet des **forces évolutives** (**explicatif**).

Les initiateurs

- Discipline qui fait appel aux outils mathématiques et statistiques.
- Initiée dans les années 1920 à 1940 par :



Ronald A Fisher (1890 – 1962)



Sewall Wright (1889 – 1988)



John BS Haldane (1892 – 1964)



Qu'est-ce que la génétique des populations ?

Les différentes approches

- **Théoriques** : nécessaire pour tester des hypothèses avec des données génétiques.
- **Expérimentales** : tester des modèles et leurs hypothèses dans des conditions contrôlées.
- **Empiriques** : décrire la distribution du polymorphisme dans les populations naturelles, et inférer l'histoire démographique et adaptative des populations.



Quelques définitions

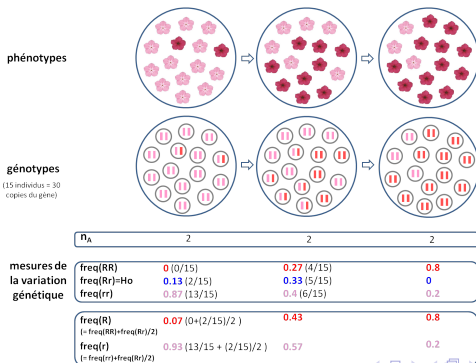
- **locus** = segment d'ADN précisément situé dans le génome.
- **polymorphe** = présente au moins deux états alléliques.
- **génotype** = chez un organisme diploïde, composition allélique d'un individu à un locus donné (ex : AA, Aa, aa).
- génotype avec même état allélique (**homozygote**) ou non (**hétérozygote**).
- **marqueur moléculaire** : locus à **déterminisme mendélien**, polymorphe, codominant (ex : microsatellites, RFLP, AFLP, SNPs)
- **entité étudiée** = la **population**, ensemble d'individus susceptibles de se reproduire entre eux à court terme.
- la variation génétique s'exprime par les fréquences relatives des différents allèles.
- **force évolutive** = processus qui agit sur les changements de fréquence des allèles.
- l'évolution se traduit par une variation des fréquences au cours du temps.



Mesurer la variation génétique

Un gène de la couleur de la fleur

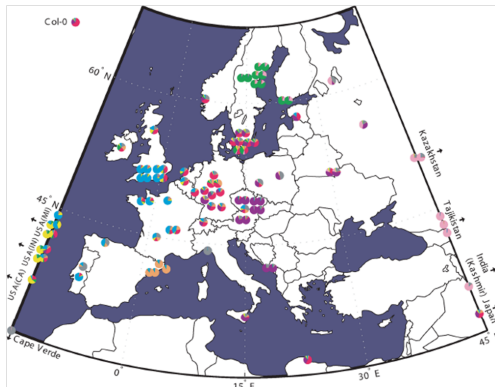
- Gène R avec allèle 'rouge' R et 'rose' r . L'allèle R domine r .
- Un marqueur moléculaire (SNP, microsatellite,...) permet de distinguer les 2 allèles, et de repérer les individus hétérozygotes.
- On génotype tous les individus diploïdes échantillonnés avec ce marqueur.
- Mesures : **nombre d'allèles** (n_A), **fréquences alléliques** ($freq(R), freq(r)$), **pourcentage d'hétérozygotes** ($freq(Rr)=H_o$).





Répartition de la variabilité génétique

Variation entre populations



Arabidopsis thaliana

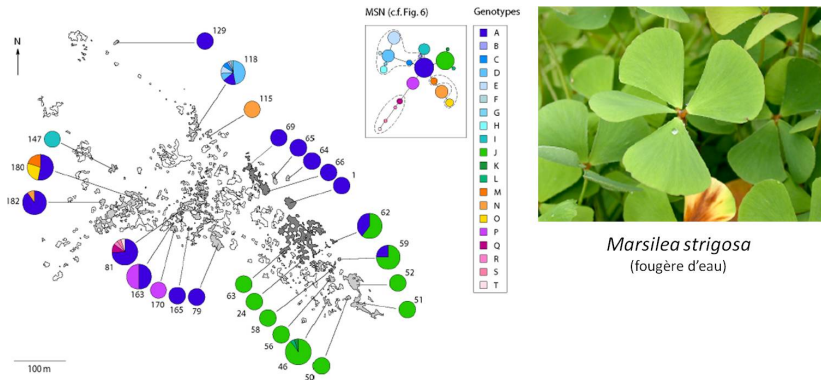
- 17000 polymorphismes (SNPs ou indels) issus de l'analyse de 876 séquences obtenues parmi 96 individus (44 000 000 bp).

(source : cours de Raphaël Leblois)



Répartition de la variabilité génétique

Variation entre sous-populations



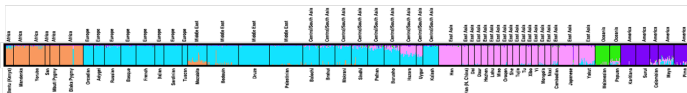
- peu de polymorphisme au sein des mares, mais une forte différenciation entre mares.

(source : cours de Raphaël Leblois)



Répartition de la variabilité génétique

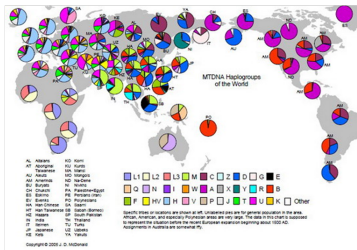
Diversité et structure génétique chez l'homme



Rosenberg et al. 2002 (Science)



source: Rober Grabel (Wikipedia)





Facteurs influençant la variabilité génétique

Le modèle 'nul' : population de Hardy-Weinberg

- croisements aléatoires (panmixie).
- => pas d'évolution génotypique, allélique.

Régimes de reproduction

- homo/hétérogamie, consanguinité, autofécondation.
- => modifient la distribution des différents génotypes.

Forces évolutives

- => modifient la fréquence des allèles.
 - **mutation** : source fondamentale de variation.
 - **migration** : introduction d'allèles (nouveaux) dans une population.
 - **sélection naturelle** : avantage ou désavantage adaptatif d'un allèle (fitness).
 - **dérive génétique** : fluctuations aléatoires des fréquences résultant d'un échantillonnage aléatoire parmi les gamètes, pour générer la génération suivante. Processus important dans les petites populations.



Loi de Hardy-Weinberg

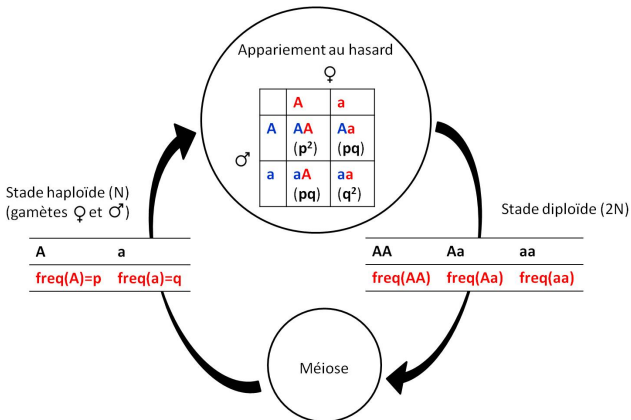
Hypothèses :

- organisme diploïde et reproduction sexuée.
 - **croisement au hasard des individus** (panmixie) [et implicitement des gamètes (pangamie)].
 - **générations non chevauchantes** (tous les individus participant à la reproduction appartiennent à la même génération).
 - **population isolée** (aucun gène introduit par des individus migrants).
 - **population de taille 'infinie'** (de grande taille).
 - pas de mutation.
 - pas de sélection.
-
- **Question : Quelles sont les fréquences génotypiques/alléliques à la génération suivante ?**
 - **REMARQUE : ces hypothèses sont rarement vérifiées en populations naturelles.**



Loi de Hardy-Weinberg

- On considère un locus à 2 allèles A et a de fréquences p et q
- Les fréquences p et q parmi les gamètes sont égales aux fréquences alléliques parmi les adultes qui les ont produits





Loi de Hardy-Weinberg

Formulée en 1908 indépendamment par :



G.H Hardy (1877-1947)
Mathématicien anglais



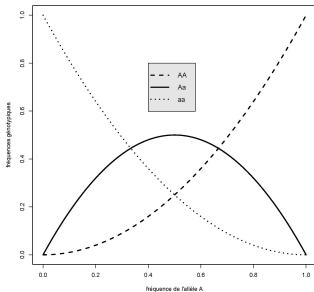
W. Weinberg (1863-1937)
Physicien et obstétricien-gynécologue

- les proportions de Hardy-Weinberg $p^2, 2pq, q^2$ (fréquences génotypiques à l'équilibre en fonction des fréquences alléliques) sont atteintes en une génération pour une espèce diploïde.
- les fréquences alléliques sont constantes au cours du temps : absence d'évolution et maintien du polymorphisme.
- Si les hypothèses précédentes sont respectées, on peut donc prédire exactement les fréquences génotypiques à partir des fréquences alléliques de la population.



Implications de la loi de Hardy-Weinberg

- Hétérozygotie : fréquence **observée** d'individus hétérozygotes (H_o) = mesure du polymorphisme.
- Si hypothèses de HW, hétérozygotie peut être déduite des fréquences alléliques : $H_o = H_e = 2pq$ (H_e = **hétérozygotie attendue sous HW**).



Distribution des fréquences génotypiques sous HW à un locus biallélique.

- Remarque : généralisation du calcul de H_e avec plus de 2 allèles :

$$H_e = 1 - \sum p_k^2 \quad (1)$$

avec p_k la fréquence de l'allèle k au locus considéré.



Test de conformité à la panmixie

Pour un locus bi-allélique A, a : test de χ^2 de conformité.

	Attendu (panmixie)	Observé
AA	Np^2	N_1
Aa	$2Npq$	N_2
aa	Nq^2	N_3

- H_0 : fréq. génot observées = fréq. génot attendues sous HW (p^2 , $2pq$, q^2).
- statistique de test :

$$X = \sum_{i=1}^m \frac{(N_{oi} - N_{ei})^2}{N_{ei}} \quad (2)$$

- avec N_{oi} = effectif observé du génotype i , N_{ei} = effectif attendu du génotype i , m = nombre de génotypes.
- distribution du χ^2 à $ddl = \text{nombre de génotypes possible} - \text{nombre d'allèles} (= 1 \text{ } ddl \text{ pour 2 allèles})$.
- p - value = $\mathbb{P}(X > \chi_{ddl=1}^2)$ proba d'obtenir une valeur au moins aussi extrême par hasard.
- si p - value < 0.05 , on rejette H_0 .
- tout écart significatif de l'équilibre génotypique résulte de l'effet d'un régime de reproduction non aléatoire ...



Quantification de l'écart à la panmixie

Mesure F_{IS}

- Sewall Wright définit le F_{IS} (ou indice de fixation F), tel que :

$$F_{IS} = \frac{H_e - H_o}{H_e} = 1 - \frac{H_o}{H_e} \quad (3)$$

- $0 < F_{IS} < 1$
- $F_{IS} = 1$ si $H_o = 0$
- $F_{IS} = 0$ si $H_o = H_e$, donc si conformité à la panmixie.
- on déduit $H_o = 2pq(1 - F_{IS})$
- remarque : on peut tester la significativité du F_{IS} ('significativement différent de 0?') par un test de permutation des allèles sur les génotypes.

Le F_{IS} mesure l'écart à la panmixie (déficit d'hétérozygote) dû à la consanguinité du régime de reproduction au sein de la population.

Régime de reproduction fermé 'extrême' : autofécondation totale (autogamie)



Medicago

génération 0
autofécondation ↓
génération 1

génération 0	AA	Aa			aa
génération 1	100%	25%	50%	25%	100%

$$H_t = \frac{H_{t-1}}{2} = \frac{H_0}{2^t} \quad (4)$$

- Réduction de l'hétérozygotie de moitié à chaque génération (à l'échelle du génome).
- A l'équilibre $\text{freq}(AA) = p$, $\text{freq}(Aa) = 0$, $\text{freq}(aa) = q$, et $F_{IS} = 1$.
- On tend donc vers des lignées pures, homozygotes à tous les locus.
- **MAIS**, toutes autres hypothèses de HW étant égales par ailleurs, les fréquences alléliques restent constantes au cours du temps.
- RILs (recombinant inbred lines) = croisement de lignées pures et autofécondations successives pour donner des lignées homozygotes mais variables entre elles au niveau de l'allèles "fixé" à chacun des locus.



Régime de reproduction fermé : autofécondation partielle

- En autofécondation partielle, chaque individu produit un taux S de descendants en autofécondation et $1 - S$ en allofécondation.
- On s'attend à un **déficit en hétérozygotes (H_o)** par rapport à HW.
- Expression de H_o en fonction de S :



Pseudoceros bifurcus
(plathelminthe)



Brassica napus

$$H_o = \underbrace{(H_o/2)S}_{\text{hétérozygotie en autofécondation}} + \underbrace{2pq(1-S)}_{\text{hétérozygotie en panmixie}}$$

taux d'autofécondation taux de panmixie
 ↓ ↓

- Expression de H_o en fonction de F_{IS} : $H_o = 2pq(1 - F_{IS})$.
- D'où la relation entre F_{IS} et S :

$$F_{IS} = \frac{S}{2 - S} \quad (5)$$

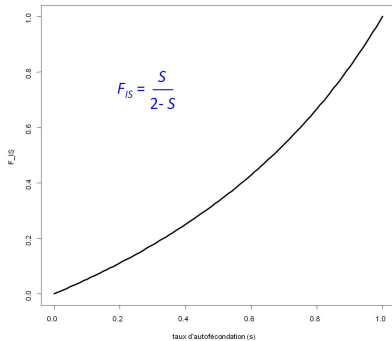
- et donc :

$$S = \frac{2F_{IS}}{1 + F_{IS}} \quad (6)$$



Régime de reproduction fermé : autofécondation partielle

Relation entre S et F_{IS}



- si $S = 0$ (panmixie) $F_{IS} = 0$
- si $S = 1$ (autogamie complète) $F_{IS} = 1$



Régime de reproduction fermé : la consanguinité

Définitions

- **Un individu consanguin** résulte de l'**union entre 2 individus apparentés (ayant au moins un ancêtre commun)**.
- **Coefficient de consanguinité** f_i d'un individu i (G Malécot 1948) : probabilité que 2 gènes homologues de l'individu i soient **identiques**.
- \Rightarrow **2 copies d'un gène sont identiques** ssi ce sont 2 copies sans mutation d'un même gène ancêtre.
- **Coefficient de parenté** ϕ_{ij} de 2 individus I et J : probabilité que 2 gènes homologues tirés au hasard l'un chez I, l'autre chez J, soient identiques.
- **Remarque importante** : le coefficient de consanguinité f_i d'un individu i est égal au coefficient de parenté de ses parents.



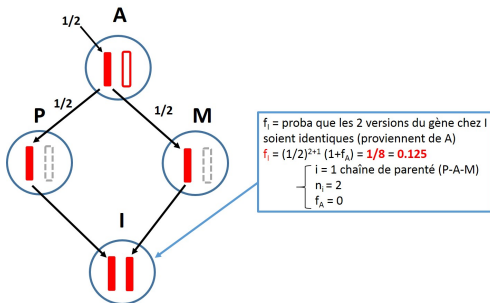
Régime de reproduction fermé : la consanguinité

Calcul du coefficient de consanguinité d'un individu

$$f_i = \phi_{PM} = \sum_i \left(\frac{1}{2}\right)^{n_i+1} (1 + f_A) \quad (7)$$

n_i = nombre de maillons de la chaîne de parenté i entre P et M (passant par un ancêtre commun, et pas deux fois par le même individu), f_A = coefficient de consanguinité de l'ancêtre commun.

Un exemple





Régime de reproduction fermé : la consanguinité

Quelques coefficients...

	coefficient de parenté
père/enfant	1/4
pleins-frères	1/4
demi-frères	1/8
cousins germains	1/16

Consanguinité dans une population : structure de Wright

- F = probabilité de trouver à un locus deux copies identiques chez un individu tiré au hasard dans la population.

$$f(AA) = p^2 + Fpq = p^2(1 - F) + Fp \quad (8)$$

$$f(Aa) = 2pq(1 - F) \quad (9)$$

$$f(aa) = q^2 + Fpq = q^2(1 - F) + Fq \quad (10)$$

- $F (= F_{IS})$ = moyenne des coefficients de consanguinité (f_i) des individus de la population ssi la consanguinité est le seul facteur provoquant un écart à HW.
- Conséquence de la consanguinité : l'augmentation de l'homozygotie a pour conséquence une occurrence de maladies rares récessives plus importante dans les populations consanguines que dans les population proches de la panmixie.



Régime de reproduction "partiellement" fermé : l'homogamie

- Homogamie : appariement préférentiel de génotypes identiques correspondant à un phénotype particulier (couleur, taille,...).
- L'homogamie réduit l'hétérozygotie, augmente l'homozygotie, pour les gènes déterminant le phénotype impliqué dans l'appariement préférentiel.

Locus « A » impliqué dans le phénotype d'appariement préférentiel

génération 0	AA x AA	Aa x Aa			aa x aa
homogamie ↓	AA	AA	Aa	aa	aa
génération 1	100%	25%	50%	25%	100%

Exemples



Podisma pedestris
(Hewitt et al. 1987)

(homogamie pour patterns de couleur)



Heliconus himera
(Merrill et al. al. 2014)

(allofécondation, homogamie pour temps de floraison)



Brassica rapa
(navette)



Régime de reproduction "partiellement" ouvert : hétérogamie

- Hétérogamie (ou allogamie) : appariement préférentiel avec un partenaire dissemblable (sur au moins un caractère).
- L'hétérogamie augmente l'hétérozygotie pour les gènes déterminant le phénotype impliqué dans l'appariement préférentiel.

Locus « A » impliqué dans le phénotype
d'appariement préférentiel

génération 0	AA x aa	Aa x AA	Aa x aa	aa x AA
hétérogamie ↓	Aa	AA Aa	aa Aa	Aa
génération 1	100%	50% 50%	50% 50%	100%

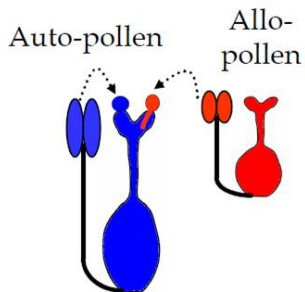
Exemple : systèmes d'auto-incompatibilité chez les plantes

- mécanismes les plus répandus et efficaces d'évitement de l'autofécondation chez les angiospermes.
- si pollen même phénotype que pistil ou stigmate => germination ou croissance inhibées : le croisement est incompatible et la fécondation ne peut avoir lieu.
- si génotypes/allèles différents, réaction compatible et fécondation possible.



Régime de reproduction "partiellement" ouvert : hétérogamie

Exemple : systèmes d'auto-incompatibilité chez les plantes





Les différentes interprétations de l'écart à HW

- **RAPPEL** : les régimes de reproduction affectent la distribution des génotypes et pas la distribution des fréquences d'allèles !!
- Cas 1 : déséquilibre de HW pour une forte majorité de loci.
 - * $H_o < H_e$.
 - **autofécondation** : réduction de moitié de l'hétérozygotie à chaque génération (ex : plantes). Ou **autofécondation** partielle.
 - **consanguinité** : croisement entre individus apparentés.
 - biais d'échantillonnage des populations : un échantillon composé en fait d'individus provenant de populations différentes en terme de fréquences alléliques.
- Cas 2 : déséquilibre de HW pour une petite fraction de loci.
 - * $H_o < H_e$, déficit d'hétérozygotes = régime de reproduction fermé au locus considéré.
 - **homogamie** : appariement préférentiel de génotypes identiques correspondant à un phénotype particulier (couleur, taille,..)
 - OU allèles nuls : biais technologique lors du typage du marqueur moléculaire (mésappariement d'une amorce PCR : un hétérozygote est repéré homozygote).
 - * $H_o > H_e$, excès d'hétérozygotes = régime de reproduction ouvert au locus considéré.
 - **hétérogamie** : appariement préférentiel de génotypes différents conduisant à des phénotypes particuliers (gènes d'auto-incompatibilité, gènes des récepteurs olfactifs, de l'immunité, chromosomes sexuels mammifères).



Les forces évolutives

Elles **modifient les fréquences alléliques** dans une population au cours du temps.

- **Mutation**
- Dérive génétique
- Sélection naturelle
- Migration



Mutation

- Changement héréditaire dans le matériel génétique.
- => distinguer mutations somatiques (cancers, ...) et **germinales** (cellules sexuelles, donc transmissibles) : **celles qui nous importent en génétique évolutive et quantitative.**
- **Source fondamentale de variation génétique.**
 - mutations géniques : changement dans la séquence nucléotidique (substitutions, insertions-délétions, transposons).
 - mutations chromosomiques : réarrangements chromosomiques (inversions, translocations).
 - mutations génomiques : polypléidisation
- Les mutations peuvent être "**neutres**" (sans impact sur le phénotype par rapport aux autres allèles) ou "**favorables/délétères**" (dépend des conditions de l'environnement)



Mutation

Événement rare...

- **substitutions** : 10^{-8} à 10^{-9} nouvelles mutations par nucléotide (base) par génération.
- ou : 10^{-4} à 10^{-6} nouvelles mutations par copie de gène par génération.
- **indels** : chez les primates : $\simeq 10^{-9}$ insertion/délétion d'une base par nucléotide par génération (Saitou & Ueda 1994).
- réarrangements génomiques/chromosomiques : variable suivant les organismes.
- **polyploïdisation** : très fréquente chez les végétaux (beaucoup chez les espèces cultivées). 70% des Angiospermes ont connu au moins un évènement de polyploïdisation.

.. mais permet l'innovation génétique à l'échelle du génome dans les espèces

- si $2Nu$ mutations apparaissent à un nucléotide donné par génération (dans une population diploïde de taille N , pour un taux de mutation génique u)...
- .. alors si $N = 10000$ et $u = 10^{-9}$, il y a 0.00002 mutation par nucleotide dans la population..
- .. et donc si l'on considère un génome de taille $\simeq 100$ Mpb chez *Arabidopsis thaliana* (ou $\simeq 3400$ Mpb chez *Homo sapiens*)..
- .. on obtient 2000 (68000) nouvelles substitutions dans la population en une génération !
- .. soit 0.2 (6.8) substitutions dans un zygote formé



Mutation

Variation du taux de mutation entre organismes

Espèce	Taille du génome (pb)	Taux de mutation par pb et par réplication	Taux de mutation par génome et par réplication
<i>Escherichia coli</i>	4.6×10^6	5.4×10^{-10}	0.0025
Bactériophage λ	4.9×10^4	7.7×10^{-8}	0.0038
<i>Caenorhabditis elegans</i>	8.0×10^7	2.3×10^{-10}	0.018
Souris	2.7×10^9	1.8×10^{-10}	0.49
Homme	3.2×10^9	5.0×10^{-11}	0.16

D'après Drake *et al.* 1998

source: cours Raphaël Leblois



Mutation

Devenir d'une mutation en l'absence d'autres forces évolutives

- Hypothèses :

- seule la mutation modifie les fréquences alléliques.
- locus biallélique (ex : SNP) : allèles A ('sauvage') et a ('mutant').
- u = taux de mutation de A vers a .
- p_t = fréquence de l'allèle A à la génération t .

$$p_1 = (1-u)p_0$$

$$p_2 = (1-u)p_1$$

$$p_2 = (1-u)(1-u)p_0 = (1-u)^2 p_0$$

$$p_t = (1-u)^t p_0 \quad (11)$$

- et variation de la fréquence p :

$$\Delta p = p_{t+1} - p_t = -up \quad (12)$$

- Application numérique : combien de générations t pour que p_0 (fréquence initiale de l'allèle sauvage A) diminue de moitié ?

$$p_t = \frac{1}{2}p_0 = (1-u)^t p_0$$

$$(1-u)^t = \frac{1}{2}$$

$$t = -\ln(2)/\ln(1-u) \approx -\ln(2)/-u \approx 0.7/u$$

- pour $u = 10^{-6}$, $t = 700000$ (14 à 17 Millions d'années chez l'homme !).
- la mutation a un rôle négligeable dans l'évolution des fréquences alléliques.



Mutation

Devenir d'une mutation en l'absence d'autres forces évolutives

- Hypothèses :
 - mutations réverses.
 - u = taux de mutation de A vers a .
 - v = taux de mutation de a vers A .

$$\Delta p = -up + v(1 - p) = v - p(u + v) \quad (13)$$

- s'annule pour :

$$p = \frac{v}{u + v} \quad (14)$$

- donc théoriquement équilibre stable (A et a vont persister à des fréquences stables), mais en réalité les fréquences à l'équilibre resteront proches des fréquences initiales : évolution très lente.



Mutation

Devenir d'une mutation

- **Théorie neutraliste de l'évolution moléculaire (Kimura, 1968, 1969, 1983) :**
 - dans une population diploïde de taille N , il y a $2N$ copies d'un gène (2 versions par individus).
 - fréquence initiale de la mutation dans la population = $\frac{1}{2N}$.
 - probabilité de fixation dans la population = $\frac{1}{2N}$.
 - probabilité d'élimination = $1 - \frac{1}{2N}$, forte probabilité dans les premières générations.
 - plus généralement (sous **dérive génétique uniquement**, voir ci-après) : probabilité de fixation = fréquence de l'allèle.
 - dans les populations de petites taille, probabilité de fixation plus importante.
 - temps moyen de fixation d'une mutation : $4N$ générations.
- Selon la théorie, la majorité des polymorphismes moléculaires résulte de l'évolution par dérive génétique d'allèles mutants sélectivement neutres (ex : ADN non codant majoritaire, 3ème position des codons - mutation synonyme -).

Devenir d'une mutation EN PRATIQUE

- **mutation neutre** : le plus souvent éliminée de la population, mais peut aussi se substituer à l'allèle sauvage, à cause des effets aléatoires de la **dérive génétique** dans les petites populations.
- **mutation défavorable** : diminue en fréquence (**sélection négative**).
- **mutation favorable** : augmente en fréquence (**sélection positive**).



Les forces évolutives

- Mutation
- **Dérive génétique**
- Sélection naturelle
- Migration



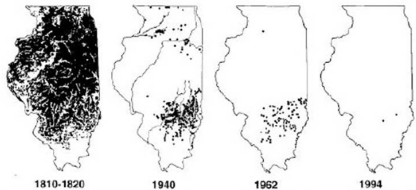
Dérive génétique

Evolution en populations finies

- On a supposé que les populations ont une taille **infinie**.
- Or les populations naturelles sont rarement de taille infinie.



Tympanuchus cupido
Tétra des prairies



Illinois

source: cours Raphaël Leblois



Dérive génétique

Evolution en populations finies

- On a supposé que les populations ont une taille **infinie** : par exemple on a supposé que les fréquences alléliques chez les descendants étaient **exactement égales** à celles des parents.
- Or dans les populations de taille finie, **les fréquences peuvent varier** d'une génération à l'autre du simple fait du **hasard** (modèle stochastique).
- Cette variation de fréquence aléatoire due à **l'effet d'échantillonnage** dans une population de taille finie s'appelle la **dérive génétique**.



Dérive génétique

Evolution en populations finies : le modèle de Wright-Fischer

- De la même façon que dans une population de taille **finie** et constante chaque individu ne donne **pas exactement un descendant et un seul**, chaque gène ne contribue **pas par une copie et une seule** à la composition génétique de la génération suivante, mais par **un nombre variable de copies** qui suit une loi de probabilité d'espérance égale à 1 (1 descendant en moyenne, certains n'en produisent pas, d'autres plus que 1...).



Ronald A Fisher (1890 – 1962)

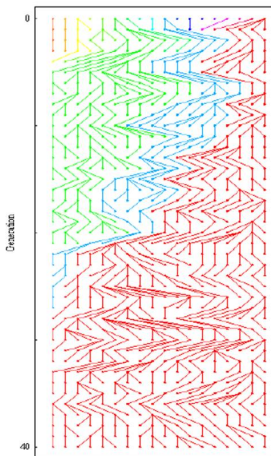


Sewall Wright (1889 – 1988)



Dérive génétique

Evolution en populations finies : le modèle de Wright-Fischer

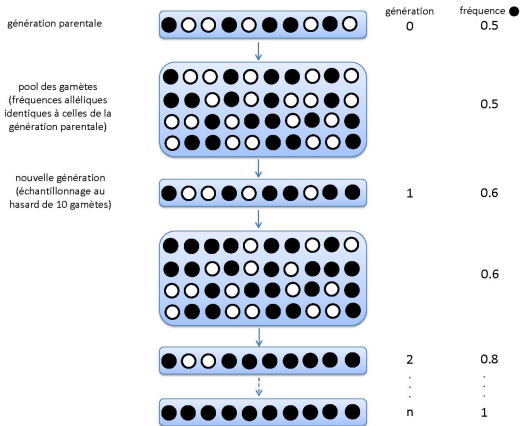


- Une population de **taille constante**, où les individus se **reproduisent une seule fois** et **au même moment** (générations non chevauchantes).
- Chaque gène à **une génération** est **la copie** d'un gène de la **génération précédente**.
- En l'absence de mutation et de sélection, **les fréquences alléliques dérivent** (augmentent et diminuent) inévitablement jusqu'à la **fixation** d'un allèle.
- La dérive conduit donc à la **perte de variation génétique à l'intérieur des populations** (et à la différenciation entre les populations).



Dérive génétique

Evolution en populations finies : comme un 'jeu du tirage' dans une urne



Echantillonnage des gamètes et changement des fréquences alléliques.



Dérive génétique

Un modèle d'évolution des fréquences alléliques

Hypothèses :

- un locus, 2 allèles A et a .
- panmixie, population de taille finie, ni mutation, ni sélection.
- fluctuation des fréquences assimilable à l'échantillonnage de $2N$ gamètes à chaque génération.
- loi de probabilité associée : loi binomiale $B(2N, p)$, avec $\mathbb{E}(X) = 2Np$ (nombre moyen de gamètes A), $\text{Var}(X) = 2Npq$.
- la proportion des gènes A dans la nouvelle génération a donc pour espérance

$$\mathbb{E}(p_{t+1}) = \mathbb{E}\left(\frac{X}{2N}\right) = \frac{1}{2N}\mathbb{E}(X) = p_t \quad (15)$$

et variance

$$\text{Var}(p_{t+1}) = \text{Var}\left(\frac{X}{2N}\right) = \frac{1}{4N^2}\text{Var}(X) \quad (16)$$

$$\text{Var}(p_{t+1}) = \frac{p_t(1-p_t)}{2N} \quad (17)$$

- par récurrence, on montre qu'après t générations :

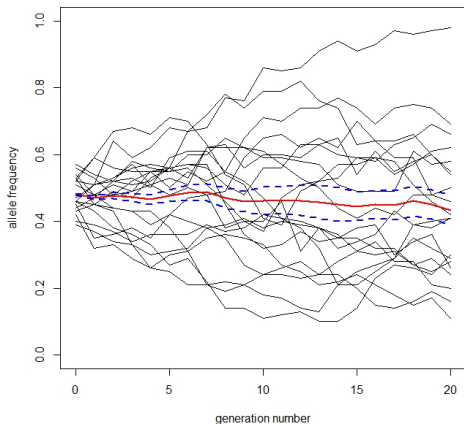
$$\mathbb{E}(p_t) = p_0 \quad (18)$$

$$\text{Var}(p_t) = p_0(1-p_0)\left(1 - \left(1 - \frac{1}{2N}\right)^t\right) \quad (19)$$



Dérive génétique

Simulations : évolution des fréquences alléliques dans plusieurs populations indépendantes



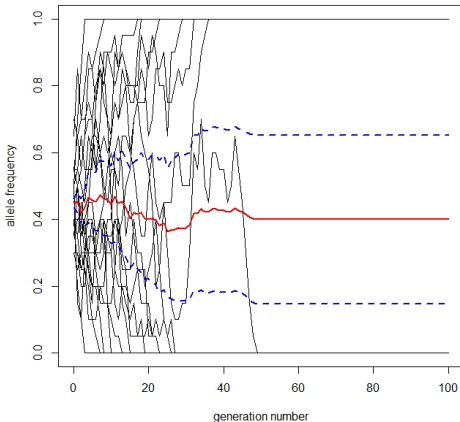
Fréquence initiale = 0.5, taille de la population = 100, 20 simulations (rouge = moyenne, bleu = variance).





Dérive génétique

Simulations

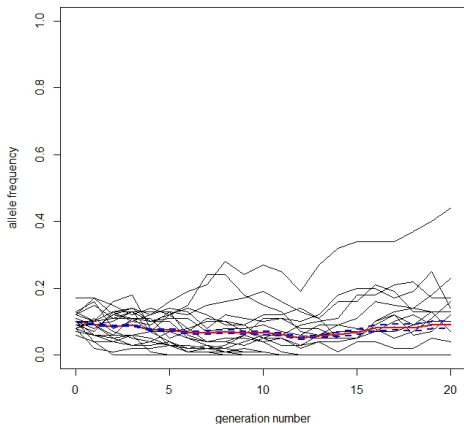


Fréquence initiale = 0.5, taille de la population = 20, 20 simulations.



Dérive génétique

Simulations

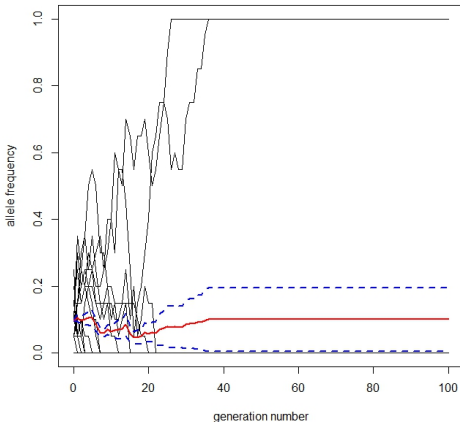


Fréquence initiale = 0.1, taille de la population = 100, 20 simulations.



Dérive génétique

Simulations



Fréquence initiale = 0.1, taille de la population = 20, 20 simulations.



Dérive génétique

Quelques caractéristiques du modèle

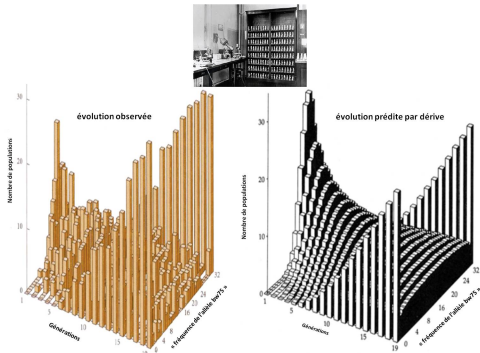
- Sur un grand nombre de populations indépendantes issues d'une même population 'mère' (ou si l'on répétait l'expérience un grand nombre de fois) :
 - en moyenne la fréquence d'un allèle ne change pas d'une génération à l'autre.
 - la variance de la fréquence d'un allèle augmente avec le temps (t), d'autant plus que N est petit (populations de petite taille), et d'autant plus que la fréquence initiale (p_0) est forte.
- Au bout d'un certain nombre de générations, chaque population va nécessairement fixer un allèle (A ou a , avec $p_t = 0$ ou $p_t = 1$).
- Ceci génère de la différenciation génétique, ou structuration génétique entre les populations.



Dérive génétique

Conséquences sur la diversité génétique : expérience de Buri (1956)

- 107 populations de drosophiles, chacune fondée avec 16 individus hétérozygotes pour la mutation 'brown eye' (*bw75*).



- au fil du temps la variation à l'intérieur des populations diminue, et la différenciation entre populations augmente.
- il y a une diminution globale de l'hétérozygotie (H_e) (la probabilité que 2 copies d'un gène tirées au hasard dans la population soient différents allèles).
- et donc il y a une augmentation globale de l'homozygotie, et une perte d'allèle (n_A).



Dérive génétique

Conséquences sur la diversité génétique : dérive et consanguinité

- La **perte de variation** due à la dérive entraîne une **augmentation de l'identité entre gènes** (de façon ultime tous les gènes d'une population finie sont des copies d'un même gène ancêtre ; ils sont tous **identiques par descendance** ; la population est entièrement **consanguine**).
- On peut mesurer la perte de variation à l'intérieur d'une population par l'**hétérozygotie (H_e)**, mais le plus souvent par le **coefficient de consanguinité, ou homozygotie, de la population (F)**.
 - **évolution de l'hétérozygotie H_e de la population** :

$$H_t = H_0 \left(1 - \frac{1}{2N}\right)^t \quad (20)$$

- $H_t = H_e$ à la génération t .

- **évolution du coefficient de consanguinité F de la population** (on suppose $F_0 = 0$) :

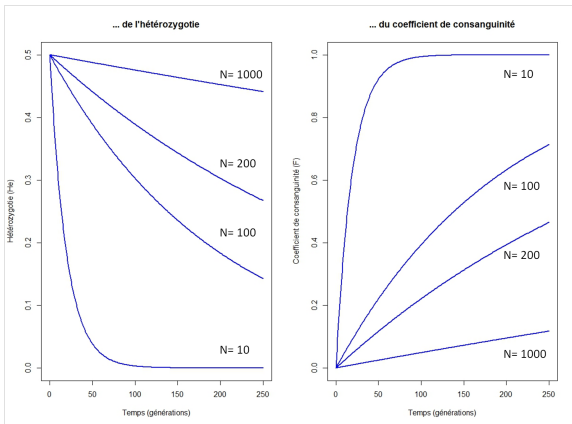
$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t \quad (21)$$

- $H_e \Rightarrow 0$, et $F \Rightarrow 1$.



Dérive génétique

Evolution...



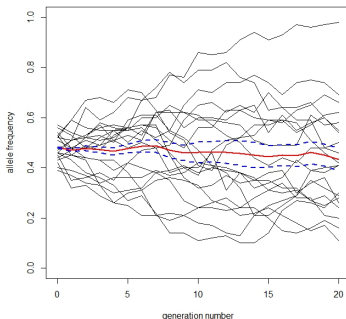
- La population tend d'autant plus vite vers l'identité (**homozygotie**) qu'elle est de petite taille.
- L'**hétérozygotie** diminue d'autant plus que la population est de petite taille.



Dérive génétique

Conséquences sur la différenciation génétique

- Evolution indépendante de plusieurs populations avec les mêmes conditions initiales.



- \Rightarrow forte variance de p sur l'ensemble des populations, donc différenciation inter-population.



Dérive génétique

Conséquences sur la différenciation génétique

- on sait que :
 - $\text{Var}(p_t) = p_0(1 - p_0)(1 - (1 - \frac{1}{2N})^t)$
 - $F_t = 1 - (1 - \frac{1}{2N})^t$

- donc :

$$F_t = F_{ST} = \frac{\text{Var}(p_t)}{p_0(1 - p_0)} \quad (22)$$

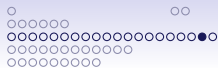
- F_t (ou F_{ST}) varie entre 0 et 1 ; il mesure le degré de différenciation entre populations divergeant uniquement sous l'effet de la dérive.
- Si $F_{ST} \Rightarrow 0$ il n'y a aucune différenciation génétique entre les populations, si $F_{ST} \Rightarrow 1$ alors il y a une différenciation extrême entre les populations.



Dérive génétique

Conclusion :

- par dérive génétique on perd la diversité allélique au locus considéré...
- ... mais en plus tous les gènes de la population au locus considéré finissent par être des **copies d'un seul des gènes de la population de départ** !
- si l'on remonte dans le temps, on trouvera un seul gène ancêtre de tous les gènes actuellement présents au locus dans une population (notion de coalescence).
- **il en est de même pour tous les locus du génome !**



Notion d'effectif efficace

Tous les individus ne participent pas forcément au processus reproductif :

- La taille (ou l'effectif) efficace N_e de la population est rarement égale à celle de la population réelle N que l'on pourrait recenser.
- On définit la taille efficace N_e d'une population comme le nombre d'individus d'une population "idéale" de type Wright-Fisher où la dérive génétique aurait la même intensité que dans la population qui nous intéresse.
- $N_e \leq N$



Notion d'effectif efficace

Taille efficace d'une population en sex-ratio déséquilibré

- Population à sexes séparés :

- taille réelle :

$$N = N_f + N_m \quad (23)$$

- taille efficace :

$$N_e = \frac{4N_m N_f}{N_m + N_f} \quad (24)$$

Exemple...

- si $N_m = 100$ et $N_f = 100 \Rightarrow N_e = 200$
- si $N_m = 100$ et $N_f = 300 \Rightarrow N_e = 300$ (pas 400 !!)



Les forces évolutives

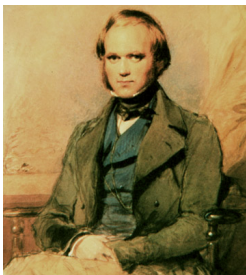
- Mutation
- Dérive génétique
- **Sélection naturelle**
- Migration



Sélection naturelle

Principes

- Charles Darwin (1809 – 1882) : "l'origine des espèces" (1859)



- **Principe 1 = variabilité** : les individus d'une espèce diffèrent les uns des autres pour certains caractères (phénotypes et donc génotypes).
- **Principe 2 = adaptation / survie** : les individus les plus adaptés au milieu survivent et se reproduisent davantage.
- **Principe 3 = transmission des caractères avantageux** : les caractéristiques avantageuses doivent être héréditaires.



Sélection naturelle

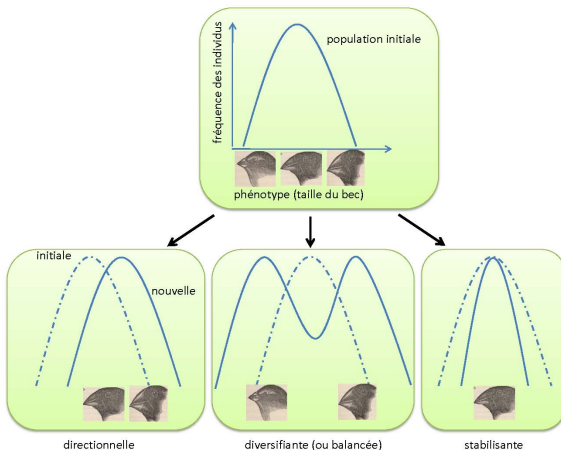
Principes de la sélection naturelle : les mots de Darwin

- Les individus diffèrent par leur capacité à survivre et à se reproduire, en partie en raison de leurs différences phénotypiques (= **variations inter-individuelles**).
- Dans chaque espèce, plus de descendants sont produits que ce qui pourra survivre et se reproduire (= **implique l'adaptation aux contraintes environnementales**).
- A chaque génération, les phénotypes (..et donc les génotypes) favorisant la survie et l'accès à la reproduction dans l'environnement actuel sont sur-représentés à l'âge de la reproduction, et contribuent de façon disproportionnée à la descendance de la génération suivante (= **sur-transmission des caractéristiques avantageuses**).



Sélection naturelle dans une population

Ce que l'on observe au niveau du phénotype (caractère)

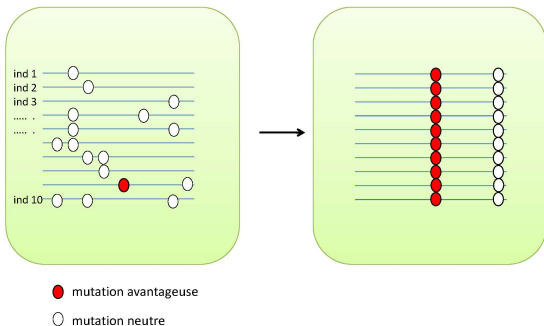




Sélection naturelle dans une population

Ce que l'on observe au niveau moléculaire : **sélection positive**

Sélection positive (darwinienne, directionnelle)



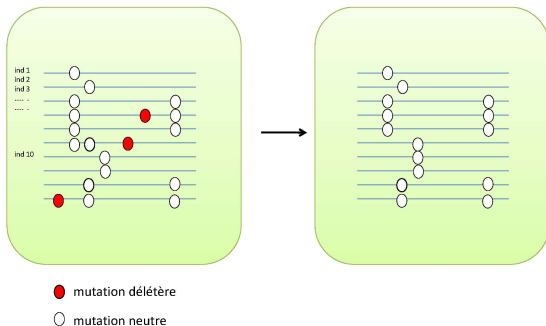
- gènes de l'adaptation (ex : résistance aux insecticides/pesticides, adaptation à la sécheresse).



Sélection naturelle dans une population

Ce que l'on observe au niveau moléculaire : **sélection purifiante**

Sélection purifiante (stabilisante, background selection)

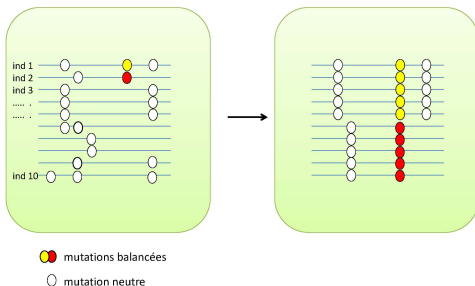




Sélection naturelle dans une population

Ce que l'on observe au niveau moléculaire : **sélection balancée**

Sélection balancée (diversifiante)



- avantage à l'hétérozygote (overdominance).
- sélection fréquence dépendante (sélection de l'allèle rare, dynamique de fréquences cyclique).
- ex : anémie falciforme chez l'homme, gènes de l'immunité (maintien d'un fort polymorphisme).



Sélection naturelle : formulation mathématique

Evolution de la fréquence des génotypes / allèles soumis à la sélection :

- w = fitness (valeur sélective), v = viabilité, f = fertilité ; $w = vf$.
- w_1, w_2, w_3 = fitness des génotypes AA, Aa, aa .
- on suppose que les génotypes se croisent au hasard (panmixie).
- \Rightarrow modifications des fréquences des génotypes (et donc des allèles) de génération en génération, dépendantes de w_1, w_2, w_3 et des fréquences p et q initiales :

	génération	AA	Aa	aa	A
		Fréquences génotypiques			Fréquence allèle A
sélection sur la viabilité	g_0 (zygotes)	p^2	$2pq$	q^2	$\text{freq}(AA)+0.5\text{freq}(Aa) = p^2 + pq = p(p+q) = p$
	g_0 (adultes)	p^2	$2pq$	q^2	$\text{freq}(AA)+0.5\text{freq}(Aa) = p$
sélection sur la fertilité		w_1	w_2	w_3	
	g_1 (zygotes)	$\frac{w_1 p^2}{\bar{w}}$	$\frac{w_2 2pq}{\bar{w}}$	$\frac{w_3 q^2}{\bar{w}}$	$\frac{w_1 p^2 + 0.5w_2 2pq}{\bar{w}}$
	g_1 (adultes)	$\frac{w_1 p^2}{\bar{w}}$	$\frac{w_2 2pq}{\bar{w}}$	$\frac{w_3 q^2}{\bar{w}}$	$\frac{w_1 p^2 + 0.5w_2 2pq}{\bar{w}}$
sélection sur la fertilité					
	g_2 (zygotes)

- \Rightarrow fréquence de l'allèle A après une génération de sélection :

$$p_{t+1} = \frac{w_1 p^2 + 0.5 w_2 2pq}{\bar{w}} \quad (25)$$

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



Sélection naturelle : formulation mathématique

Variation de la fréquence d'un allèle soumis à la sélection :

- variation de p entre 2 générations.

$$\Delta p = p_{t+1} - p_t = \frac{w_1 p^2 + 0.5 w_2 2pq}{\bar{w}} - p \quad (26)$$

$$\Delta p = pq \frac{(w_1 - w_2)p + (w_2 - w_3)q}{w_1 p^2 + 2w_2 pq + w_3 q^2} \quad (27)$$

- **on donne souvent $w = 1$ au génotype ayant la plus grande valeur sélective, et on exprime les autres par rapport à celle-ci.**
- du coup s ($w = 1 - s$) n'est pas un avantage sélectif mais un coefficient de contre-sélection du génotype.



Sélection naturelle : formulation mathématique

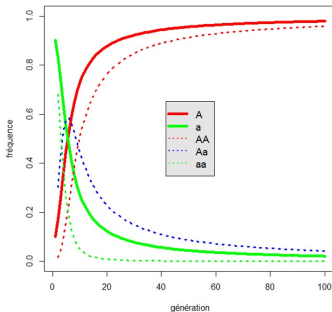
Fréquences alléliques d'équilibre (quand $\Delta_p = 0$) :

- le signe de Δ_p renseigne sur le sens de l'évolution : > 0 = augmentation de fréquence de l'allèle A. $= 0$ signifie une fréquence d'équilibre.
- le signe de Δ_p dépend du numérateur de l'équation (27), donc des valeurs sélectives.
 - $w_1 > w_2 > w_3$ ou $w_1 = w_2 > w_3$ ou $w_1 > w_2 = w_3$: fréquence d'équilibre $p = 1$, fixation de l'allèle A.
 - $w_1 < w_2 < w_3$: fréquence d'équilibre $p = 0$, fixation de l'allèle a.
 - $w_1 < w_2 > w_3$: avantage à l'hétérozygote, maintien du polymorphisme.
 - $w_1 > w_2 < w_3$: avantage aux homozygotes, maintien du polymorphisme.
- la valeur d'équilibre de la fréquence de l'allèle A est $p = \frac{w_3 - w_2}{w_1 - 2w_2 + w_3}$.



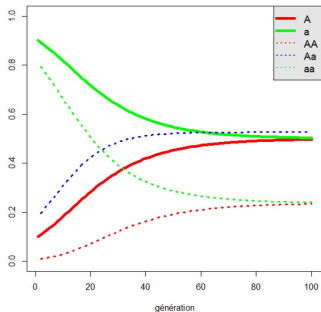
Sélection naturelle : formulation mathématique

Trajectoires de fréquences suivant le mode de sélection :



$w_1=1, w_2=1, w_3=0.5$
 freq(A) initiale = 0.1
 freq(a) initiale = 0.9

avantage à l'allèle A
(sélection positive sur A)
(sélection négative sur a)



$w_1=0.9, w_2=1, w_3=0.9$
 freq(A) initiale = 0.1
 freq(a) initiale = 0.9

avantage aux 2 allèles
(sélection balancée)
(coexistence des 2 allèles, maintien
'actif' du polymorphisme)



Sélection naturelle

Conclusion :

- la sélection naturelle conduit à une **évolution déterministe** des fréquences alléliques, à l'opposé de la dérive (aléatoire).
- la sélection naturelle **agit sur certains gènes du génome**, pas tout le génome à l'opposé de la dérive.
- la sélection naturelle peut **réduire la diversité allélique d'un gène** (sélection positive, négative) mais aussi **maintenir activement le polymorphisme** (sélection balancée, par exemple pour les gènes des récepteurs de l'immunité).
- afin d'identifier des gènes importants dans l'adaptation d'une espèce donnée à son environnement (biotique ou abiotique), des méthodes de génétique des populations ont été développées, qui permettent de déterminer si un locus est sous l'influence de la sélection plus que de la dérive.
- la **sélection artificielle sur des espèces domestiquées** (cultures, bétail) a des **effets similaires à la sélection positive** (directionnelle) et/ou stabilisante.



Les forces évolutives

- Mutation
- Dérive génétique
- Sélection naturelle
- **Migration**



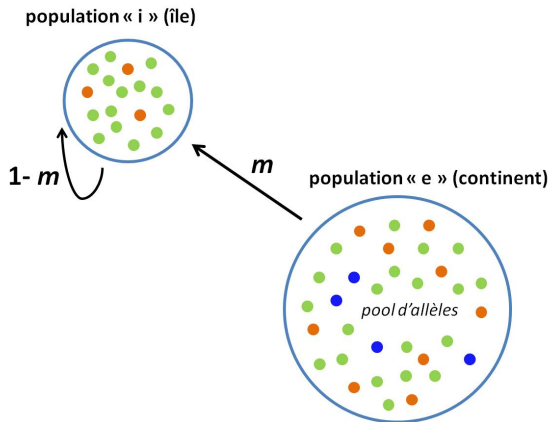
Migration

l'échange d'individus (ou de gamètes) entre sous populations permet les flux de gènes



Migration

Le modèle le plus simple : modèle en "île" (ou archipel)



- on suppose l'absence de mutation, de sélection, et de dérive génétique.



Migration : formulation mathématique

Le modèle le plus simple : modèle en "île" (ou archipel)

- ce modèle considère que la migration s'effectue d'une population 'source' ("e") vers une population 'puit' ("i").
- à chaque génération une fraction $1 - m$ des copies de l'allèle A de fréquence p_i de la population i est remplacée par une fraction m provenant d'immigrants chez qui la fréquence est p_e :
- la fréquence de l'allèle A à la génération $t + 1$ en fonction de la génération t s'exprime :

$$p_{t+1}^i = (1 - m)p_t^i + mp_e \quad (28)$$

- par récurrence, la fréquence de l'allèle A à la génération t en fonction de la génération 0 s'exprime :

$$p_t^i = (1 - m)^t (p_0^i - p_e) + p_e \quad (29)$$

- la variation de p entre 2 générations dans la population i qui reçoit des migrants :

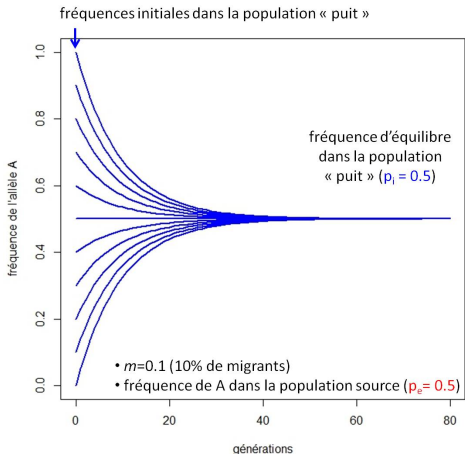
$$\Delta p_i = p_{t+1}^i - p_t^i = m(p_e - p_i) \quad (30)$$

$m = \text{taux de migration}$ (fraction de gènes provenant des migrants), p_e est supposé constant dans la population source.



Migration : formulation mathématique

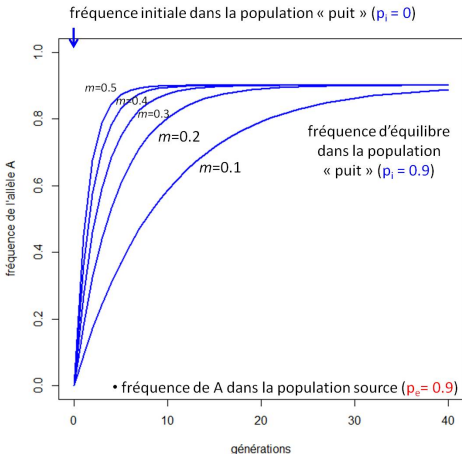
Trajectoires de fréquence de l'allèle A en fonction de p_0 :





Migration : formulation mathématique

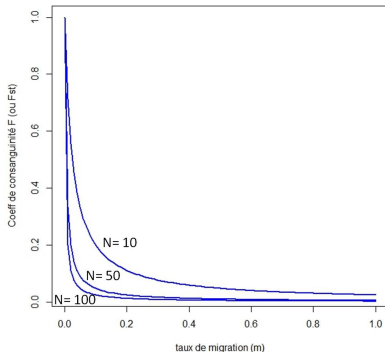
Trajectoires de fréquence de l'allèle A en fonction de m :





Migration : formulation mathématique

Effet de la migration sur le coefficient de consanguinité F de la population (et la différenciation) :



- la différenciation entre les populations (ou le coefficient de consanguinité dans la population "puit") décroît avec m et augmente avec la dérive (N petit).



Migration : formulation mathématique

Effet de la migration sur le coefficient de consanguinité F de la population (et la différenciation) :

- Un résultat remarquable prédit par ce type de modèle est que à l'équilibre entre migration et dérive :

$$F \simeq \frac{1}{1 + 4Nm} \quad (31)$$

- N = la taille de la population (ou des populations connectées), m = le taux de migration.
- un nouveau migrant ($Nm = 1$) s'installant à chaque génération dans la population suffit à limiter sa consanguinité à une valeur maximale $F = 0.2$ au lieu de $F = 1$.
- en calculant F avec des données génétiques, on peut avoir une estimation du taux de migration entre populations.



Migration

Propriétés génétiques de la migration comme force évolutive :

- comme la mutation, la migration **apporte de nouveaux allèles dans une population**, c'est une source d'innovation génétique.
- mais la migration est une force de **bien plus forte intensité que la mutation**.
- En l'absence d'autres forces évolutives, **la migration tend à homogénéiser les fréquences alléliques entre populations qui échangent des gènes (= mélange)**.
- Elle **s'oppose donc à la différenciation génétique** qui tend à s'installer entre des populations sous l'effet de la dérive notamment.
- Le modèle de migration (la façon dont les individus se déplacent dans l'espace, la dispersion «en groupe» ou solitaire, etc..) influence beaucoup la distribution spatiale du polymorphisme.
- d'autres modèle que le modèle en îles sont étudiés (isolement par la distance, populations subdivisées (structurées), métapopulations, admixture, hétérogénéité des taux de migration etc...).



Concepts importants du chapitre génétique des populations

La théorie...

- la façon dont les organismes se reproduisent (régimes de reproduction) influence principalement les fréquences des génotypes dans une population au cours des générations, et notamment l'hétérozygotie observée.
- la loi de Hardy-Weinberg et la statistique F_{is} permettent d'évaluer le régime de reproduction le plus probable dans une espèce donnée.
- les forces évolutives (mutation, dérive, sélection, migration) influencent les fréquences des allèles au cours des générations.
- les forces évolutives ont des effets variables :
 - **la mutation** est l'ultime source d'innovation, mais ne peut à elle seule permettre une évolution rapide.
 - **la dérive génétique** a un effet majeur sur les fréquences alléliques (à tous les locus du génome) surtout dans les petites populations. Elle est à l'origine de la différenciation génétique des populations.
 - **la sélection** est la force qui permet l'adaptation des organismes à leur environnement, en favorisant les allèles conférant les caractères les plus adaptés. Elle peut réduire comme stabiliser la variation génétique (à certains gènes d'un génome).
 - **la migration** s'oppose à la différenciation génétique et à la consanguinité générée par la dérive, elle tend à homogénéiser les populations.



Concepts importants du chapitre génétique des populations

... et la réalité

- nous avons vu l'**effet** de tous ces processus de manière "**séparée**" ...
- ... mais **dans la réalité tous ces processus s'entremêlent** plus ou moins ...
- ... et l'objectif de la **génétique des populations moderne** est de développer des méthodes statistiques et des modèles pour essayer de **tirer des informations sur l'importance relative de ces divers processus (régimes de reproduction et forces évolutives)** dans les populations, et de **reconstruire leur histoire démographique et adaptative**, à partir des données de polymorphisme génétique (marqueurs moléculaires) obtenus par séquençage/génotypage.



References



Précis de génétique des population. JP Henry, PH Gouyon



Principles of Population Genetics, 4th Edition. Hartl DL, Clark AG



cours de Laurent Excoffier,

http://genet.univ-tours.fr/EXCOFFIER/Laurent/GMDP_intro.htm.



cours de Raphaël Leblois, raphael.leblois.free.fr/ressources/cours/Cours_DeriveMigrationMetapopFstat_ENS_2009.pdf.