

Introduction à la génétique des populations

M1 : MABS

Parcours : Bioinformatique et Biologie des Systèmes

UE : Génomique et Génétique Statistiques

Maxime Bonhomme

UMR CNRS-UPS 5546, Laboratoire de Recherche en Sciences Végétales,
Castanet-Tolosan

11 octobre 2011

Génétique des populations

- 1 Introduction
- 2 Population de Hardy-Weinberg
 - modèle et implications
 - régimes de reproduction et écart au modèle
- 3 Les forces évolutives
 - mutation
 - dérive génétique
 - sélection naturelle
 - migration
- 4 Notations pour GGS
- 5 References

Qu'est-ce que la génétique des populations ?

- Etude de la distribution et des changements de la fréquence des versions d'un gène (allèles) dans les populations d'êtres vivants.
- Influence des pressions évolutives (sélection naturelle, dérive génétique, mutations, et migration).
- Initiée dans les années 1920 à 1940 par Ronald Fisher, J.B.S. Haldane et Sewall Wright.
- Application des principes fondamentaux de la génétique mendelienne à l'échelle des populations.
- A permis de faire la synthèse entre la génétique mendelienne et la théorie de l'évolution = néo-darwinisme (théorie synthétique de l'évolution).
- Les changements de fréquence des allèles peuvent conduire au processus de spéciation.

La diversité génétique : fondements de la génétique des populations

- 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).
- la variation génétique s'exprime par les fréquences relatives des différents allèles.
- entité étudiée = la population, ensemble d'individus susceptibles de se reproduire entre eux à court terme.
- l'évolution se traduit par une variation des fréquences au cours du temps.
- force évolutive = processus qui agit sur les changements de fréquences.

Systèmes de reproduction et forces évolutives

- les systèmes de reproduction (homo/hétérogamie, consanguinité, autofécondation) modifient la distribution des différents génotypes.
- les 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.

Objectifs de la génétique des populations

- Mesurer la variation génétique dans les populations et décrire les patrons de distribution de la variation (statistiques descriptives).
- Expliquer l'origine, le maintien et l'évolution de la variation génétique par l'effet des forces évolutives (modélisation et statistique inférentielle).
- Discipline essentiellement quantitative qui fait appel aux outils mathématiques et statistiques.

Hypothèses du modèle Hardy-Weinberg

- Hypothèses du modèle de population de Hardy-Weinberg
 - organisme diploïde et reproduction sexuée.
 - générations non chevauchantes (individus de chaque génération meurent avant la naissance des membres de la génération suivante).
 - fréquences alléliques identiques chez mâles et femelles.
 - panmixie (unions au hasard des individus) et pangamie (unions au hasard des gamètes).
 - très grande taille de la population (N).
 - mutation négligeable.
 - migration entre populations négligeable.
 - la sélection naturelle n'agit pas au locus considéré.

Loi de Hardy-Weinberg

Formulée en 1908 indépendamment par :

- G.H. Hardy (1877-1947) mathématicien anglais.
- W. Weinberg (1862,1937) physiologiste allemand.

Si les hypothèses précédentes sont respectées, on peut prédire exactement les fréquences génotypiques à partir des fréquences alléliques de la population.

Exemple : si locus à deux allèles A et a de fréquences p et q : croisements aléatoires des gamètes ♂ et ♀ de même fréquences et fréquences génotypiques attendues.

gamètes ♂ gamètes ♀	$A(p)$	$a(q)$
$A(p)$	$AA(p^2)$	$Aa(pq)$
$a(q)$	$Aa(pq)$	$aa(q^2)$

$$f(AA)=p^2, f(Aa)=2pq, f(aa)=q^2$$

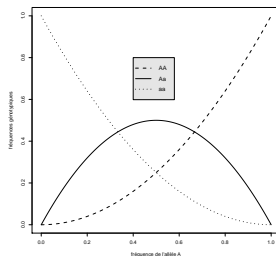
remarque : la reproduction aléatoire entre génotypes est équivalente à l'union aléatoire des gamètes.

Implications de la loi de Hardy-Weinberg

- une seule génération de panmixie suffit à atteindre les fréquences génotypiques de HW.
- fréquences alléliques constantes selon les hypothèses ($p_{t+1} = p_t$), donc absence d'évolution et maintien du polymorphisme.
- base pour un modèle plus complexe avec séparation en 2 phases :
 - gamètes à zygotes : fréquences génotypiques à la naissance en génération t inchangées par rapport aux adultes de $t - 1$.
 - zygotes à adultes : ajout de l'effet de la sélection ou de la migration (changement des fréquences alléliques).

Implications de la loi de Hardy-Weinberg

- Hétérozygotie : fréquence 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 bi-allélique.

- Mesure et test de l'écart : test de χ^2 de conformité.
 - H_0 = fréquences génotypiques observées = fréquences attendues sous HW (p^2 , $2pq$, q^2).
 - statistique de test :

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

avec N = effectifs genotypiques, m = nombre de génotypes

- Distribution du χ^2 à $ddl = n_{alleles} - 1$ ($= 1$ ddl pour 2 allèles).
- $P - valeur = \mathbb{P}(X > \chi^2_{ddl=1})$ proba d'obtenir une valeur au moins aussi extrême par hasard

Ecart au modèle de Hardy-Weinberg

- Généralisation de l'hétérozygotie à plus de 2 allèles :

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

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

- Autre mesure de l'écart à HW (plus interprétable) :
 - F_{IS} de Sewall Wright (ou indice de fixation F) :

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

- P-valeur nominale : tests de permutation des allèles sur les génotypes.

Interprétation de l'écart à HW

- Cas 1 : déséquilibre de HW pour une petite fraction de loci.
 - * $H_o < H_e$, déficit d'hétérozygotes = régime de reproduction fermé au locus considéré.
 - **homogamie** : apparemment préférentiel de génotypes identiques correspondant à un phénotype particulier (couleur du plumage, taille, ornements ...)
 - **allèles nuls** : biais technologique lors du typage du marqueur moléculaire (misappariement d'une amorce PCR).
 - * $H_o > H_e$, excès d'hétérozygotes = régime de reproduction ouvert au locus considéré.
 - **hétérogamie** : apparemment préférentiel de génotypes différents conduisant à des phénotypes particuliers (gènes des récepteurs olfactifs, de l'immunité, chromosomes sexuels).
- Cas 2 : déséquilibre de HW pour une forte majorité de loci.
 - * $H_o < H_e$.
 - **autogamie** : réduction de moitié de l'hétérozygotie à chaque génération (ex : plantes).
 - **consanguinité** : croisement entre individus apparentés (ayant au moins un ancêtre en commun).
 - **biais d'échantillonnage** des populations : un échantillon composé en fait d'individus provenant de populations différentes (fréq all).

Régime de reproduction fermé : exemple de l'autogamie

AA	Aa	aa
AA	AA Aa aa	aa
AA	0.25 0.5 0.25	aa

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

- Réduction de l'hétérozygotie de moitié à chaque génération (à l'échelle du génome).
- A l'équilibre $\text{freq}(\text{AA}) = p$ et $\text{freq}(\text{aa}) = q$.
- On tend donc vers des lignées pures, homozygotes à tous les locus.
- RILs (recombinant inbred lines) = croisement de lignées pures et autofécondation pour donner des lignées pures mais variables entre elles au niveau de l'allèles "fixé" à chacun des locus.

Régime de reproduction fermé : exemple de la consanguinité

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

$$f_i = \phi_{PM} = \sum_i \left(\frac{1}{2}\right)^{ni+1} (1 + f_A) \quad (5)$$

- ni = nombre de chaînes de parentés 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.

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.

$$f(AA) = p^2 + Fpq = p^2(1 - F) + Fp \quad (6)$$

$$f(Aa) = 2pq(1 - F) \quad (7)$$

$$f(aa) = q^2 + Fpq = q^2(1 - F) + Fq \quad (8)$$

- $F (= F_{IS})$ = moyenne des coefficients de consanguinité (f_i) des individus de la population ssi la consanguinité est le seul facteur provoquant un écart à HW.

Effet Wahlund : biais d'échantillonnage

- La structure génotypique observée sur l'ensemble correspond aux fréquences génotypiques moyennes :

$$AA = \sum (p_i^2) \cdot (n_i / N) = \mathbb{E}(p^2) \quad (9)$$

$$Aa = \sum (2p_i q_i) \cdot (n_i / N) = 2\mathbb{E}(pq) \quad (10)$$

$$aa = \sum (q_i^2) \cdot (n_i / N) = \mathbb{E}(q^2) \quad (11)$$

avec p_i et n_i la fréquence de l'allèle A dans la population i et la taille de la population i .

- Si l'ensemble avait été une seule grande unité panmictique (comme on le croit lors de l'échantillonnage), on aurait attendu :

$$AA = P^2 = [\mathbb{E}(p)]^2 \quad (12)$$

$$Aa = 2PQ = 2\mathbb{E}(p)\mathbb{E}(q) \quad (13)$$

$$aa = Q^2 = [\mathbb{E}(q)]^2 \quad (14)$$

avec $E(p)$ la moyenne des fréquences p_i de l'allèle A sur k populations.

- On obtient donc :

$$F_{IS} = \frac{H_e - H_o}{H_e} = \frac{2PQ - 2\mathbb{E}(pq)}{2PQ} = 1 - \frac{H_o}{2PQ} \quad (15)$$

- Ici $F_{IS} = 0$ si p_i sont égaux (une seule population), $F_{IS} > 0$ si différentes populations.

Régime de reproduction ouvert : les chromosomes sexuels chez les mammifères

- On considère X et Y comme des allèles, $\text{freq}(X) = \frac{3}{4}$ et $\text{freq}(Y) = \frac{1}{4}$. En panmixie on aurait $H_e = \frac{3}{8}$, or on a $H_o = \frac{1}{2}$ (proportion des mâles, sexe hétérogame)
- **ATTENTION** : les régimes de reproduction affectent la distribution des génotypes et pas la distribution des fréquences d'allèles !!

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.
- Source fondamentale de variation génétique.
 - mutations géniques : changement dans la séquence nucléotidique (mutation ponctuelle, indels, transposons).
 - mutations chromosomiques : réarrangements chromosomiques (inversions, translocations).
 - mutations génomiques : polyploïdisation
- Distinguer mutations somatiques (cancers, ...) de germinales (cellules sexuelles, donc transmissibles).
- Distinguer mutations neutres (sans impact sur le phénotype par rapport aux autres allèles) de favorable/délétère (dépend des conditions du milieu)

Mutation

- Événement rare.
 - 10^{-8} à 10^{-9} nouvelles mutations par nucléotide par génération.
 - 10^{-4} à 10^{-6} nouvelles mutations par copie de gène par génération.
 - mutations génomiques : polyploïdisation
- Innovation génétique.
 - Nombre de mutation par génération = $2Nu$, non négligeable si population grande.
 - Nombre important de gènes dans les génomes (ex : 20 000 -30 000 chez l'homme), donc plusieurs gènes mutés chez un zygote.

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 et a .
- u = taux de mutation de A vers a .

$$p_{t+1} = (1 - u)p_t \quad (16)$$

$$\Delta_p = p_{t+1} - p_t = -up \quad (17)$$

$$p_t = (1 - u)^t p_0 \quad (18)$$

- Application numérique : nombre de générations t pour que p diminue de moitié.

$$p_t = \frac{1}{2} p_0 = (1 - u)^t p_0 \quad (19)$$

$$(1 - u)^t = \frac{1}{2} \quad (20)$$

$$t = \frac{-\ln 2}{\ln(1 - u)} \simeq \frac{-\ln 2}{u} \simeq \frac{0.7}{u} \quad (21)$$

- pour $u = 10^{-6}$, $t = 700000$ (10 à 20 Ma chez l'homme!).
- rôle négligeable dans l'évolution des fréquences alléliques.

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 .

$$\Delta p = -up + v(1 - p) = v - p(u + v) \quad (22)$$

- s'annule pour :

$$p = \frac{v}{u + v} \quad (23)$$

- donc théoriquement équilibre stable, mais en réalité une faible influence sur les fréquences alléliques.

Devenir d'une mutation en pratique

- mutation défavorable : diminue en fréquence (**sélection négative**).
- mutation favorable : augmente en fréquence (sélection positive).
- 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.
- Théorie neutraliste de l'évolution moléculaire (Kimura, 1968,1969,1983) :
 - fréquence initiale de la mutation = $\frac{1}{2N}$.
 - probabilité de fixation = $\frac{1}{2N}$.
 - probabilité d'élimination = $1 - \frac{1}{2N}$, forte probabilité dans les premières générations.
 - plus généralement (sous dérive génétique uniquement) : 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 -).

Les forces évolutives

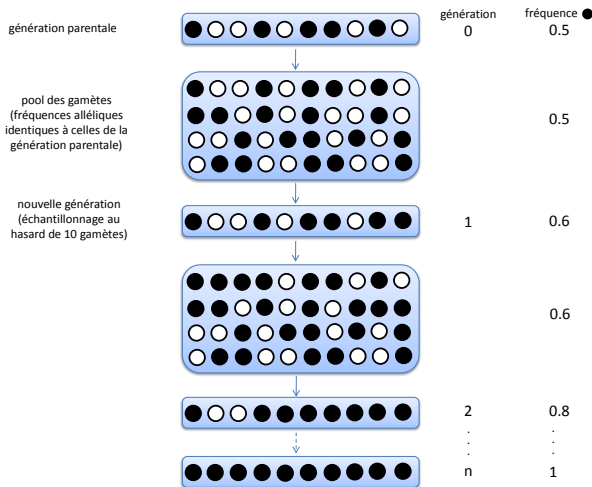
- Mutation
- **Dérive génétique**
- Sélection naturelle
- Migration

Dérive génétique

Fluctuation des fréquences alléliques de générations en générations du fait d'un échantillonnage aléatoire des gamètes dans une population de taille finie (non transmission de certains allèles à la descendance, individus ne se reproduisant pas).

- Effet sur la diversité génétique intra-population (H_e , nombre d'allèles).
- Effet sur la diversité génétique inter-population (variance des fréquences alléliques - F_{ST}).

Dérive génétique



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, ni migration.
- pas de distorsion de ségrégation lors de la formation des gamètes (méiose).
- 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

$$\mathbb{E}(p_{t+1}) = \mathbb{E}\left(\frac{X}{2N}\right) = \frac{1}{2N} \mathbb{E}(X) = p_t \quad (24)$$

et variance

$$\text{Var}(p_{t+1}) = \text{Var}\left(\frac{X}{2N}\right) = \frac{1}{4N^2} \text{Var}(X) \quad (25)$$

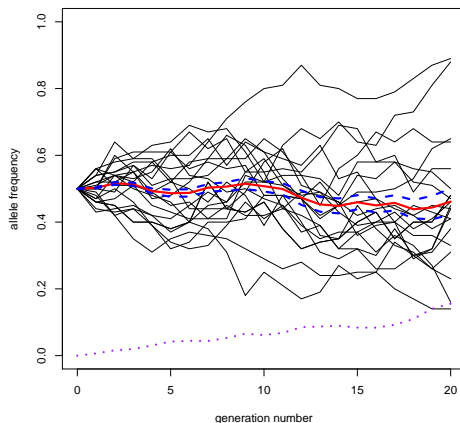
$$\text{Var}(p_{t+1}) = \frac{p_t(1-p_t)}{2N} \quad (26)$$

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

$$\mathbb{E}(p_t) = p_0 \quad (27)$$

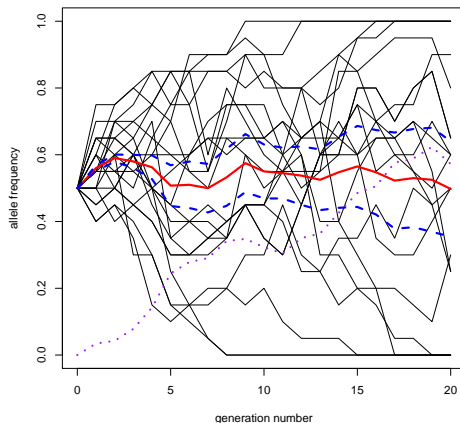
$$\text{Var}(p_t) = p_0(1-p_0)\left(1 - \left(1 - \frac{1}{2N}\right)^t\right) \quad (28)$$

Dérive génétique : simulations



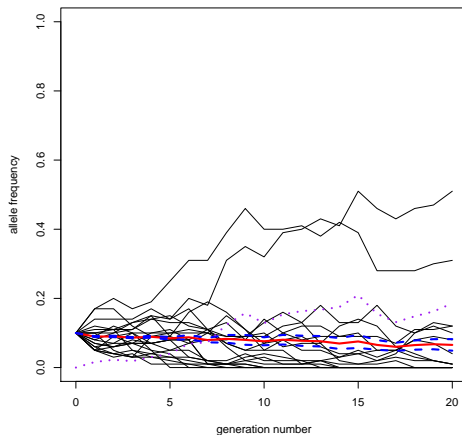
Fréquence initiale = **0.5**, taille de la population = **100**, 20 simulations (rouge = moyenne, bleu = variance, violet = coefficient de consanguinité moyen).

Dérive génétique : simulations



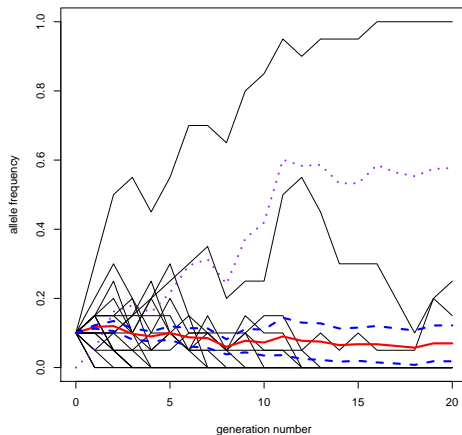
Fréquence initiale = 0.5, taille de la population = 20, 20 simulations.

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 : effet sur la diversité intra-population

- fluctuations aléatoires (variance) plus importantes si N petit.
- fluctuations aléatoires (variance) moins importantes si p_0 petit.
- effet de la dérive génétique sur une nouvelle mutation (cf : devenir d'une mutation en pratique).
- réduction de l'hétérozygotie H_e (maximale pour $p=q=0.5$).
- production d'homozygotes de deux manières, à la génération t :
 - tirage de 2 gamètes provenant de la même copie de gène

$$\frac{1}{2N} \cdot \frac{1}{2N} \cdot 2N = \frac{1}{2N} \quad (29)$$

- tirage de 2 gamètes provenant de 2 copies de gène différentes mais ayant le même allèle

$$\left(1 - \frac{1}{2N}\right) f_{t-1} \quad (30)$$

avec f_{t-1} la proba que 2 copies d'un gène soient identiques dans la population parentale.

- variation de l'homozygotie entre 2 générations ($t-1$ et t) :

$$f_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) f_{t-1} \quad (31)$$

Dérive génétique : effet sur la diversité intra-population

- variation de l'hétérozygotie entre 2 générations ($t - 1$ et t) :

$$H_t = \left(1 - \frac{1}{2N}\right) H_{t-1} \quad (32)$$

avec $H_t = 1 - f_t$.

- évolution du coefficient de consanguinité f de la population :

$$f_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) f_{t-1} \quad (33)$$

$$(1 - f_t) = \left(1 - \frac{1}{2N}\right) (1 - f_{t-1}) = \left(1 - \frac{1}{2N}\right)^t (1 - f_0) \quad (34)$$

$$f_t = 1 - \left(1 - \frac{1}{2N}\right)^t \quad (35)$$

on suppose $f_0 = 0$.

Dérive génétique : effet sur la diversité intra-population

- évolution de l'hétérozygotie H_e de la population :

$$\mathbb{E}(H_t) = \mathbb{E}(2p_t(1 - p_t)) = 2(\mathbb{E}(p_t) - \mathbb{E}(p_t^2)) \quad (36)$$

or

$$\text{Var}(p_t) = \mathbb{E}(p_t^2) - [\mathbb{E}(p_t)]^2 \quad (37)$$

et

$$\mathbb{E}(p_t) = p_0 \quad (38)$$

$$\text{Var}(p_t) = p_0(1 - p_0)(1 - (1 - \frac{1}{2N})^t) \quad (39)$$

d'où

$$\mathbb{E}(p_t^2) = p_0(1 - p_0)(1 - (1 - \frac{1}{2N})^t) + p_0^2 \quad (40)$$

$$\mathbb{E}(H_t) = 2p_0(1 - p_0)(1 - \frac{1}{2N})^t \quad (41)$$

$\mathbb{E}(H_t)$ tend donc vers 0, et f vers 1.

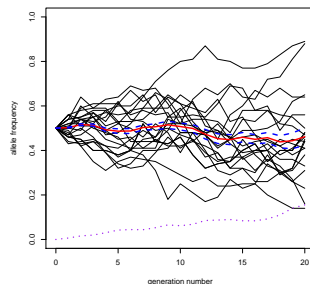
Dérive génétique : effet sur la diversité intra-population

Conclusion :

- 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 : voir cours coalescence !

Dérive génétique : effet sur la diversité génétique inter-population

Evolution indépendante de plusieurs populations avec les mêmes conditions initiales.



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

Dérive génétique : effet sur la diversité génétique inter-population

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

$$\text{Var}(p_t) = p_0(1 - p_0)\left(1 - \left(1 - \frac{1}{2N}\right)^t\right) \quad (42)$$

$$f_t = 1 - \left(1 - \frac{1}{2N}\right)^t \quad (43)$$

$$f_t = \frac{\text{Var}(p_t)}{p_0(1 - p_0)} \quad (44)$$

- f_t (ou F_{ST}), compris entre 0 et 1, mesure aussi le degré de différenciation entre populations divergeant uniquement sous l'effet de la dérive.
- notons que cette statistique est la variance des fréquences alléliques "normalisée" par l'hétérozygotie de la population ancestrale.

Dérive génétique : notion d'effectif génétique

Tous les individus ne participent pas forcément au processus reproductif : la taille efficace N_e de la population réelle est le nombre d'individus d'une population idéale de type Wright-Fisher (population de taille finie, nombre infini de gamètes) pour lequel on aurait un degré de dérive génétique équivalent à celui de la population réelle ($N_e < N$).

- Population changeant de taille :

$$\frac{1}{N_e} = \frac{1}{t} \sum_{k=1}^t \frac{1}{N_k} \quad (45)$$

N_e est la moyenne harmonique, N_k la taille de la population à la génération k .

Dérive génétique : notion d'effectif génétique

- Population à sexes séparés :

- taille réelle :

$$N = N_f + N_m \quad (46)$$

- taille efficace :

$$\frac{1}{2N_e} = \frac{1}{8N_f} + \frac{1}{8N_m} \quad (47)$$

$$N_e = \frac{4N_m N_f}{N_m + N_f} \quad (48)$$

Les forces évolutives

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

La sélection naturelle

Formulation moderne de la sélection naturelle :

Dans chaque espèce, plus de descendants sont produits que ce qui pourra survivre et se reproduire.

Les individus diffèrent par leur capacité à survivre et à se reproduire, en partie en raison de leurs différences phénotypiques et en relation avec leurs caractéristiques génotypiques (la relation entre les deux pouvant être complexe).

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.

La sélection naturelle

Evolution d'une population soumise à la sélection :

- w = fitness (valeur sélective), v = viabilité, f = fertilité; $w = v.f$
- w_1, w_2, w_3 = fitness des génotypes AA, Aa, aa
- fréquence des zygotes produits : $p^2, 2pq, q^2$

$$p_{t+1} = \frac{w_1 p^2 + 0.5 w_2 2pq}{\bar{w}} \quad (49)$$

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).

$$\Delta p = p_{t+1} - p_t = \frac{w_1 p^2 + 0.5 w_2 2pq}{\bar{w}} - p \quad (50)$$

$$\Delta p = pq \frac{(w_1 - w_2)p + (w_2 - w_3)q}{w_1 p^2 + 2w_2 pq + w_3 q^2} \quad (51)$$

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 plus un avantage sélectif mais un coefficient de contre-sélection du génotype.

La sélection naturelle

Evolution d'une population soumise à la sélection :

- 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 (51), 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 $p = \frac{w_3 - w_2}{w_1 - 2w_2 + w_3}$.

La sélection naturelle : équilibre mutation-sélection

Cryptopolymorphisme : équilibre entre introduction récurrentes de mutations le plus souvent défavorables dans une séquence codante (ex : "'maladies génétiques" : mucoviscidose, hémophilie...), et contre-sélection. Exemple d'une mutation récessive défavorable :

Les génotypes AA , Aa , aa ont les valeurs sélectives $w_1 = 1$, $w_2 = 1$, $w_3 = 1 - s$. Dans une population panmictique, la pression de sélection est :

$$\Delta p = pq \frac{sq}{1 - sq^2} \quad (52)$$

La pression de mutation est :

$$\Delta p = -up \quad (53)$$

La pression d'équilibre est :

$$\Delta_s p + \Delta_m p = 0 \quad (54)$$

$$\frac{sq^2}{1 - sq^2} = u \quad (55)$$

$$sq^2 = \frac{u}{1 + u} \simeq u \quad (56)$$

Fréquence des individus atteints : $q^2 = \frac{u}{s}$.

Les forces évolutives

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

La migration (1 population qui reçoit des migrants)

Modèle simple en "île" : à chaque génération une fraction des copies de l'allèle A de la population i est remplacée par une fraction provenant d'immigrants chez qui la fréquence est p_e :

$$p_{t+1}i = (1 - m)p_t i + mp_e \quad (57)$$

$$\Delta p_i = p_{t+1}i - p_t i = m(p_e - p_i) \quad (58)$$

m = taux de migration (fraction de gènes provenant des migrants), p_e est constant dans la population source.

La migration homogénéise les fréquences entre les populations qui échangent des gènes. Elle s'oppose donc à la différenciation qui tend à s'installer entre des populations sous l'effet de la dérive notamment.

La migration (1 population qui reçoit des migrants)

Effet sur le coefficient de consanguinité f de la population :

$$f_t = \frac{1}{2N} + (1 - \frac{1}{2N})(1 - m)^2 f_{t-1} \quad (59)$$

$$f_{eq} = \frac{1}{2N} + (1 - \frac{1}{2N})(1 - m)^2 f_{eq} \quad (60)$$

$$f_{eq} = \frac{1}{2N(1 - (1 - 1/2N)(1 - m)^2)} \quad (61)$$

si m est faible, m^2 est négligeable devant m et :

$$f_{eq} \simeq \frac{1}{1 + 4Nm - 2m} \quad (62)$$

$$f_{eq} \simeq \frac{1}{1 + 4Nm} \quad (63)$$

Résultat remarquable : un nouveau migrant ($Nm = 1$) s'installant à chaque génération dans la population suffit à limiter sa consanguinité à une valeur maximale $f_e = 0.2$ au lieu de $f_e = 1$.

Notations cours et TD

AA, Aa, aa

$freq(AA), freq(Aa), freq(aa)$

p, q

t

p_t

H

H_e

H_o

$h = 1 - H$

$\chi^2_{ddl=}$

ddl

p_i

$p_{i,j}$

$j = 1, \dots, n$

Génotypes d'un locus biallélique, allèles codominants

Fréquence des génotypes d'un locus biallélique

Fréquences des allèles A et a d'un locus biallélique

Génération t , ou temps en nombre de générations

Fréquence de l'allèle A à la génération t

Hétérozygotie

Hétérozygotie attendue (expected)

Hétérozygotie observée

Homozygotie

Loi de chi deux

degré(s) de liberté

Fréquence de l'allèle A dans la population i

Fréquence de l'allèle i dans la population j

Indices des populations (ou des individus)

Notations cours et TD

n_i	Nombre d'individus échantillonnés dans la population i
ϕ_{ij}	Coefficient de parenté entre les individus i et j
f_i	Coefficient de consanguinité d'un individu i
$\ell = 1, \dots, L$	Indices des locus
\bar{p}	Moyenne empirique de l'échantillon $p_1, \dots, p_i, \dots, p_n$
s_p^2	Variance empirique de l'échantillon $p_1, \dots, p_i, \dots, p_n$
$\mathbb{E}(X)$	Espérance mathématique de la variable aléatoire X
$\text{Var}(X)$	Variance de la variable aléatoire X
$\text{Cov}(X, Y)$	Covariance entre les variables aléatoires X et Y
$\mathbb{P}(X)$	Probabilité de l'événement X
$\text{♀}, \text{♂}$	Femelle, Mâle
P, M	Père, mère
f	Coefficient de consanguinité de la population dû à la dérive (suppose HW)
$F = F_{IS}$	Indice de fixation dans une population (écart aux fréquences de HW) = coeff de consanguinité de la pop si toutes les autres conditions de la loi
F_{ST}, F_{IS}, F_{IT}	Indices de fixation de Wright (F-statistics)

Notations cours et TD

m	Taux de migration par génération
u	Taux de mutation par génération
w	Valeur sélective (fitness), nb moyen de descendants
s	Avantage sélectif
N	Taille de la population (diploïde)
N_e	Taille efficace de la population
IBD	Identique par descendance
IBS	Identique par état
DL	Déséquilibre de liaison
D, D', r^2	Mesures du déséquilibre de liaison
c	Probabilité de crossing-over sur un segment par génération
$\rho = 4N_e c$	Nombre de recombinaisons sur un segment dans l'ensemble de la pop
r	Taux de recombinaison entre paires de bases adjacentes
$\theta = 4N_e u$	Nombre de mutations sur un segment dans l'ensemble de la pop =taux de mutation "rescaled" en coalescence
p_{AB}	Fréquence de l'haplotype -ou gamètes- AB pour 2 loci (allèles A,a et B,b)

Notations cours et TD - coalescence

$$\tau = \frac{t}{2N}$$

Mesure du temps en unité de $2N$ générations dans la version continue du coales

t

temps en unités de générations

N

taille haploïde de la population

n

taille haploïde de l'échantillon

T_k^N

temps de coalescence pour un échantillon de taille k et une population de taille

T_k

temps de coalescence pour un échantillon de taille k dans la limite $N \rightarrow +\infty$

u

taux de mutation par génération

References



Précis de génétique des population. JP Henry, PH Gouyon



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



pour aller plus loin...

[http ://genet.univ-tours.fr/EXCOFFIER/Laurent/GMDP;ntro.htm](http://genet.univ-tours.fr/EXCOFFIER/Laurent/GMDP;ntro.htm)