

UE11 – PARCOURS - Génétique - n°7 03/04/2019 Audrey SABBAGH	RT : Noémie Lemainque, Lucas Lescure RL : Guillaume Le Péron
---	--

Génétiques des populations : les forces évolutives

Plan :

I. Mutations

- A- Les mutations : source de variabilité
- B- Différents types de mutations
- C- Évolution des théories
- D- Taux de mutation
- E- Différentes expressions du taux de mutation
- F- Estimation du taux de mutation

II. Dérive génétique

- A- Définition
- B- Échantillonnage
- C- Notion d'effectif efficace
- D- Sex ratio et effectif efficace
- E- Calcul de l'effectif efficace
- F- Fixation et élimination d'un nouvel allèle
 - 1) Fixation
 - 2) Élimination
- G- Conséquences de la dérive génétique
 - 1) Au sein de la population
 - 2) Entre les populations
- H- Effet fondateur

III. Migrations

- A- Modèle de l'île
- B- Migrations multidirectionnelles
- C- Modèle d'isolement par la distance

IV. Sélection naturelle

A- Fitness absolue (W)

B- Fitness relative (w)

C- Coefficient de sélection (s)

D- Différents types de sélection

E- Sélection différentielle

3) Définition

4) Exemples

Dans ce cours, 4 forces évolutives seront évoquées : les mutations, la dérive génétique (hasard), la migration et la sélection naturelle (déterministe).

I. Mutations

A- Les mutations : source de variabilité

La mutation est une force évolutive très importante car elle permet de créer de la **diversité génétique** et de la **variabilité allélique**. En effet, une mutation sur un gène donné peut créer un nouvel allèle et donc une nouvelle protéine ayant un effet différent de la protéine originelle. Cela peut se faire de deux façons : soit par mutation simple, c'est-à-dire par modification de la séquence nucléotidique, soit par recombinaison méiotique. Dans ce deuxième cas, au niveau d'un locus, il peut y avoir un crossing-over entre les deux allèles d'un même gène (celui issu du père et celui issu de la mère) ce qui donne un gamète haploïde ayant des chromosomes « patchwork » ce qui crée des nouveaux allèles.

Ces mutations créatrices de nouveaux allèles permettent de déterminer un **potentiel adaptatif**. Ces nouveaux allèles peuvent ainsi procurer des avantages aux personnes les possédant s'il apparait une modification de leur environnement et peuvent leur permettre de s'adapter.

Les mutations peuvent être spontanées (erreurs récurrentes au niveau de la réplication de l'ADN ou à cause d'hydrolyses spontanées mais réparées généralement par les mécanismes de réparation de l'ADN) ou provoquées par des agents physiques (ex : rayons du soleil) ou chimiques.

Ces nouveaux allèles peuvent être neutres (pas d'avantage ou de désavantage vis-à-vis de la sélection naturelle), avantageux ou désavantageux.

B- Différents types de mutations

Il existe :

- ➔ Les **mutations ponctuelles** : variation de la séquence nucléotidique de type SNP (Simple Nucleotide Polymorphism = substitution, les plus fréquentes), les petites insertions-délétions de quelques bases.
- ➔ Les **variations du nombre de copies**, de plus en plus étudiées. Ils peuvent être microsatellites, minisatellites ou de type CNV.

Pour les **microsatellites**, il s'agit d'un motif d'1 à 4 nucléotides qui se répètent en tandem (=consécutif). La séquence reste la même mais le nombre de répétitions peut varier en fonction de l'allèle. Ils sont présents un peu partout dans le génome mais plus rarement près de régions codantes car ils peuvent poser problème si la variation correspond à un ajout ou à un retrait de bases non multiples de 3 (=décalage du cadre de lecture).

Pour les **minisatellites**, la séquence est plus longue (quelques dizaines de bases) et il peut y avoir variation du nombre de répétition ET de la séquence. Ils sont situés juste avant les télomères et sont moins abondants que les microsatellites.

Pour les **CNV (Copy Number Variation)**, la séquence variante est beaucoup plus grande et peut mesurer de 1kb à 3Mb. Ils peuvent contenir des gènes et on se retrouve donc avec des individus possédant ce gène en plusieurs exemplaires et d'autres ne le possédant pas du tout.

- Les **remaniements chromosomiques** qui correspondent à une modification de la structure des chromosomes par changement de place d'un fragment chromosomique de taille variable. On y retrouve :
- Les **translocations** : échange de matériel génétique entre chromosome non-homologues. Apparition de chromosomes mosaïques aboutissant à des phénomènes de délétions de grande échelle, de duplication ou d'inversion (soit en dehors du centromère soit au niveau du centromère = inversion péri-centrique). On n'a pas de perte ou de gain de matériel génétique mais une modification de l'ordre des gènes. Cela peut aboutir à l'inactivation d'un gène ou à la création d'un nouveau gène et d'une nouvelle protéine.
 - **Modification du nombre de chromosomes** : gain ou perte d'un chromosome unique (= aneuploïdie) ce qui aboutit à une monosomie ou à une trisomie. Certaines de ces modifications du nombre de chromosomes sont viables (trisomie 21 par exemple) alors que d'autres sont létales dès le stade embryonnaire. À noter que la duplication totale du génome (polyploïdie) ou la perte totale d'une paire de chromosomes (nullisomie) ne sont pas viables chez l'Homme.

Le taux de mutation est faible chez l'Homme, de l'ordre de 10^{-8} à 10^{-9} mutations par site et par génération pour les SNP et 10^{-5} à 10^{-6} pour les marqueurs microsatellites. Cela peut paraître très peu mais si on prend en compte le génome entier d'une personne (3 milliards de pb) ainsi que l'ensemble de la population humaine (7 milliards d'individus), cela représente une quantité non-négligeable de mutations. Ainsi, pour chaque gamète créé, on aura $3 \times 10^9 \times 10^{-9} = 3$ nouveaux allèles. Pour un zygote, cela représente donc 6 nouveaux allèles. Pour l'intégralité de la population humaine, on se retrouve à chaque génération avec environ 42×10^9 nouveaux allèles qui apparaissent. Ces nouveaux allèles ont un potentiel adaptatif très important.

Ces mutations ne peuvent être transmises à la descendance que si elles touchent des cellules germinales. Ainsi, si la mutation est déjà présente au moment de la méiose et est transmise à un gamète fécondant (=mutation héritée), alors toutes les cellules de l'individu seront touchées. Maintenant, s'il s'agit d'une mutation *de novo* apparaissant avant la fécondation, le résultat sera le même. Enfin, si la mutation apparaît plutôt à un stade de mitose plus tardif dans le développement dans certains types cellulaires, on se retrouvera avec un individu dont certains tissus seront touchés par la mutation : c'est une mosaïque somatique.

C- Évolution des théories

Pour Lamarck, les populations s'adaptent à leur environnement ce qui induit l'apparition de mutations suite à ces efforts. Les mutations apparaîtraient donc suite à la modification de l'environnement.

L'exemple type est celui de la girafe. Au départ on a plutôt des girafes avec des cous courts. Si on a une modification de l'environnement avec des feuilles d'acacias de plus en plus hautes, les girafes doivent fournir un effort au niveau du cou pour les atteindre. Ainsi, le cou s'allonge et, selon la théorie de Lamarck, cela déclenche l'apparition de mutations qui peuvent être transmises à la descendance.

Pour Darwin, la mutation est déjà présente dans la population avant la modification de l'environnement. Ainsi, si on reprend le modèle de la girafe, selon Darwin, il existait avant toute modification de l'environnement des populations de girafes avec des cous plus ou moins longs et avec une moyenne des cous plutôt courts.

Avec la modification de l'environnement évoquée, les girafes aux cous longs seront favorisées, elles pourront mieux manger, mieux survivre et mieux se reproduire. L'allèle qui leur fournit le cou plus long peut ainsi être transmis à la descendance et progressivement au fil des générations, cet allèle devient majoritaire dans la population. Le polymorphisme de la longueur des cous existe toujours mais la moyenne tend ainsi à augmenter petit à petit.

La démonstration de ces hypothèses peut se faire par expérimentation sur des souches de bactéries. Ces bactéries ont été exposées à un agent sélectif, un antibiotique par exemple, qui entraîne l'apparition de mutation de résistance à cet agent. Au départ, on a plusieurs souches de bactéries qu'on laisse se développer sur plusieurs générations. Ensuite, on les soumet à un agent sélectif. Deux possibilités : soit l'agent est responsable de l'apparition de la mutation et on ne devrait l'observer qu'à la dernière génération, soit cette mutation était déjà présente avant l'introduction de l'agent et on pourra la retrouver dans une cellule et dans les cellules mères dont elle est issue tandis qu'on assistera aussi à une propagation de la mutation car seules les cellules l'ayant possédée avant l'introduction de l'agent devraient survivre. Ces modèles sont contradictoires et il a été prouvé que c'est le deuxième modèle qui est valable.

D- Taux de mutation

Considérons un locus autosomique, bi-allélique avec deux allèles A et a. u correspond au taux (ou probabilité) de mutation de A vers a et v le taux de mutation de a vers A à chaque génération.

À la génération i , notre allèle A est retrouvé à une fréquence p_i et notre allèle a est retrouvé à une fréquence q_i . On cherche à déterminer la fréquence de ces allèles à une génération $i+1$.

up_i représente donc la fréquence des allèles qui ont mutés de A vers a
Même chose pour vq_i

- $p_{i+1} = (1-u)p_i + vq_i$ ($1-u$ correspond à la probabilité pour A de ne pas muter)
- $q_{i+1} = (1-v)q_i + up_i$ ($1-v$ correspond à la probabilité pour a de ne pas muter)

Si on cherche les fréquences à l'équilibre p_e et q_e , on obtient après développement des formules précédentes :

- $p_e = v / (u+v)$
- $q_e = u / (u+v)$

À l'équilibre, les fréquences ne dépendent que des taux de mutation dans les deux sens.

Deux cas extrêmes :

- Si les taux de mutations sont identiques dans les deux sens, on se retrouvera à la fin avec deux fréquences égales à 0,5.
- Si la probabilité de muter de a vers A est nulle, à terme, la fréquence de A à l'équilibre sera nulle tandis que celle de a sera égale à 1.

E- Différentes expressions du taux de mutation

Taux de mutation = nombre de mutations attendus sur un segment d'ADN pendant une période de temps donnée.

Pour chaque taux de mutation, la taille du segment étudié peut être mesuré en nucléotides ou en loci. Pour la durée, elle peut être en années ou en génération. Par exemple, sachant que le locus considéré mesure 1500 pb et qu'une génération dure 25 ans :

- 15 mutations par locus par génération
- = 0,6 mutation par locus par an
- = 0,0004 mutation par nucléotide par an
- = 0,01 mutation par nucléotide par génération

F- Estimation du taux de mutation

i. Méthodes directes

a) Modèle d'évolution expérimentale (sur *Escherichia Coli*, bactériophage T4, *Saccharomyces cerevisiae*)

On va utiliser dans ces expériences des lignées d'accumulation. On part d'abord d'une certaine souche à la génération g_0 puis on crée des répliquats de cette génération mère possédant le même matériel génétique. On obtient donc différentes lignées identiques pour le moment qui forment la population ancestrale. Ensuite, on les laisse évoluer pendant un certain nombre de générations. Au fil des divisions et des répliquations, ces lignées vont accumuler des mutations. À un moment donné, on arrête les divisions et on étudie le contenu génétique de ces bactéries. On fait donc le séquençage d'un segment d'ADN à différents temps, à g_0 , au bout de 50, 100, 150 générations et on identifie pour chaque lignée les différentes mutations qui sont apparues. On estime le taux de mutation avec la formule :

$$u = m / (L \times n \times g)$$

- u le taux de mutation par nucléotide par génération
- m le nombre de mutations observées
- L le nombre de lignées d'accumulation
- n le nombre de sites nucléotidiques analysées
- g le nombre de générations

b) Cas particulier des maladies/phénotypes dominants : dénombrement des individus malades nés de parents sains

Peut se réaliser sur des maladies à transmission autosomique dominante. On utilise ici l'exemple de l'achondroplasie sporadique, maladie qui se développe à la suite d'une mutation *de novo* dans 90% des cas, les parents ne sont donc pas atteints et ne portent pas la mutation.

Cette méthode consiste à observer des trios composés d'un enfant et de ses parents. On compare des trios avec des enfants non-atteints et des enfants atteints. Le nombre de naissance est connu grâce à l'état civil et le nombre d'enfants atteints est connu grâce aux réseaux de laboratoires avec des dépistages systématiques de la pathologie. On détermine ainsi le nombre de nouvelles apparitions *de novo* chez le nouveau-né.

Si on a 242 257 nouveaux-nés dont 7 sont atteints d'achondroplasie :

- Prévalence de la maladie = $7 / 242\,257 = 2,9 \times 10^{-5}$
- Taux de mutation = $7 / (2 \times 242\,250) = 1,4 \times 10^{-5}$ (division par 2 car deux allèles)

Le taux de mutation est 10 000 fois plus élevé que le taux de mutation classique du génome humain.

La situation idéale pour faire ces estimations est de se trouver dans le cas d'une pathologie monogénique à pénétrance complète sans hétérogénéité génétique (un seul locus impliqué) et sans hétérogénéité allélique (un seul allèle impliqué).

ii. Méthode historique de Haldane

Basé sur le principe de l'équilibre mutation-sélection, non traitée dans ce cours.

iii. Méthode moléculaire pour les mutations neutres

On compare des séquences d'ADN non-fonctionnel et donc à priori neutres. Avant on prenait n'importe quelle séquence non-codantes mais on s'est rendu compte que certaines séquences non-codantes pouvaient avoir un rôle important dans la modulation de l'expression des gènes. On les compare entre deux espèces dont le temps de divergence est connu et on s'appuie sur la notion d'horloge moléculaire qui dit que les mutations s'accumulent dans différentes espèces à un rythme régulier au cours du temps et similaire entre les espèces (attention ce n'est pas toujours le cas !!)

On peut calculer le taux de mutation u selon la formule :

$$u = k / (2t + 4Ne)$$

- u le taux de mutation par nucléotide par génération.
- k le taux de divergence nucléotidique (nombre de différences nucléotidiques entre deux séquences) entre les deux espèces étudiées.
- t le temps (générations) depuis la divergence entre les 2 espèces.
- Ne la taille efficace de la population ancestrale (abordée dans la partie sur la dérive génétique).

Pour évaluer le taux de mutation des séquences neutres, on prend une séquence donnée du génome humain que l'on peut retrouver chez le chimpanzé, on les compare et on compte les nucléotides qui diffèrent puis on divise ce nombre par le nombre de nucléotides total dans la séquence. Cela nous permet d'obtenir le taux de divergence k de la séquence qui est d'environ 1,3%. Le temps de divergence de ces deux espèces est d'environ 5 millions d'années et si l'on considère qu'une génération dure 20 ans, cela fait $2,5 \times 10^5$ générations. La taille efficace humaine est d'environ 10 000 individus. En remplaçant dans l'équation ces paramètres, on obtient un taux de mutation de $2,5 \times 10^{-8}$ mutation par nucléotide par génération.

Grâce à cette méthode, on a pu remarquer que les substitutions de bases sont 10 fois plus présentes que les insertions/délétions tandis que les transitions sont 10 fois plus nombreuses dans les îlots CpG qu'en dehors de ces îlots et que les transitions sont 2 fois plus nombreuses que les transversions.

Conclusion : Les mutations sont très importantes car elles permettent de créer de la diversité au sein d'une espèce ce qui permet l'adaptation des populations à leur environnement. Ce taux de mutation est très faible mais est non négligeable à l'échelle de l'espèce. Cependant, il est trop faible pour modifier substantiellement les fréquences alléliques.

II. Dérive génétique

A- Définition

Ça représente le hasard dans l'évolution, c'est un processus stochastique.

C'est la fluctuation aléatoire des fréquences alléliques dans une population d'effectif limité, liée à l'effet d'échantillonnage.

Cette dérive génétique n'existe pas chez les populations infinies mais comme ces dernières sont purement théoriques, nous n'en tiendrons pas compte. Dans ce cas, les fréquences alléliques restent stables au cours du temps si les autres forces évolutives n'interviennent pas.

B- Échantillonnage

Dans une population de taille finie, les fréquences alléliques vont varier de façon aléatoire avec une amplitude dépendant de la taille de la population.

Pour schématiser cet échantillonnage aléatoire, on considère un pool génétique d'une population i avec une certaine fréquence allélique dont on place tous les allèles dans une urne allélique. On réalise un tirage aléatoire avec remise de $2N$ allèles pour former ma génération $i+1$ avec N individus. Théoriquement, si une personne a un nombre infini d'enfants, ses deux allèles seraient répartis de façon équitable entre chacun de ses enfants et la fréquence allélique de ces allèles chez les enfants serait de 50% chacun. Sauf qu'en pratique, nous n'avons en moyenne que deux enfants. Il est donc tout à fait possible de transmettre deux fois l'allèle 1 sans transmettre l'allèle 2. Sur un petit nombre de tirage, les variations peuvent être importantes. *C'est un peu comme quand on tire à pile ou face, plus le nombre de tirage est grand, plus on a de chances de finir à 50% pour pile et 50% pour face.*

Ainsi, d'une génération à l'autre, par le simple fait du hasard au niveau de la transmission des allèles, il est possible d'observer une fluctuation des fréquences alléliques. Cette fluctuation est d'autant plus importante que la population est faible.

Prenons l'exemple d'un groupe de 12 adultes possédant deux allèles pour un gène donné, l'allèle A et l'allèle a, les deux ayant une fréquence allélique de 0,5. On prend tous les gamètes de ces adultes pour former une urne gamétique avec à l'intérieur les mêmes fréquences alléliques. Pour constituer les 12 individus de la génération suivante, on doit tirer 24 gamètes au hasard. Du simple fait de ce hasard, on peut tirer une majorité de A et aboutir à une fréquence allélique de 0,75 pour A, ou alors en tirer une minorité et finir avec une fréquence allélique de 0,33 pour A dans cette génération suivante.

L'échantillonnage des allèles d'une génération à l'autre aboutit inexorablement à une réduction de la diversité génétique par fixation d'un allèle (fréquence allélique à 100%) ou par perte d'un allèle (fréquence allélique à 0%) => **état absorbant**.

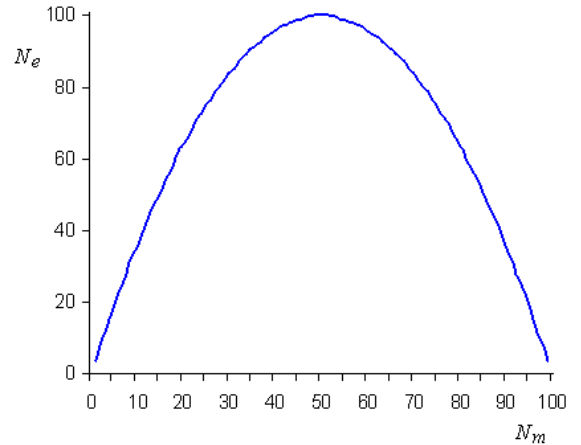
Plus la population est petite, plus les états absorbants seront atteints rapidement.

Si la population est grande, les fluctuations alléliques seront de faible amplitude et les états absorbants seront atteints plus lentement.

C- Notion d'effectif efficace

Ce n'est pas de la population réelle en réalité que dépend la vitesse de fluctuation des allèles mais de l'effectif efficace. Il s'agit de la fraction de la population qui participe réellement à la reproduction. C'est la taille d'une population idéale qui présente le même niveau de dérive génétique que la population donnée.

L'effectif efficace est dans la grande majorité des cas inférieure, très rarement égale et jamais supérieure à la population réelle.



D- Sex ratio et effectif efficace

L'effectif efficace d'une population dépend du sex ratio de cette dernière. Pour illustrer cette variation, prenons l'exemple d'un troupeau de 100 zèbres. Parmi eux, on trouve :

- 50 zèbres non reproducteurs
- 10 zèbres mâles (N_m)
- 40 zèbres femelles (N_f)

L'effectif efficace ici se calcule de la manière suivante, car $N_{\text{mâles}} \neq N_{\text{femelles}}$:

$$N_e = (4 \times N_f \times N_m) / (N_m + N_f)$$

Ici $N_e = 32$ (soit $1/3$ de la population au lieu de $1/2$ si on n'avait pas pris en compte le sex ratio de la population de zèbres)

Mais pourquoi le sex ratio est-il si important pour l'effectif efficace d'une population ?

Si on observe ce graphique on peut en effet voir qu'un équilibre est nécessaire entre les deux sexes pour un effectif efficace maximal.

- Si la population est constituée à 100% de mâles, alors le N_e sera égal à 0, pareil s'il est constitué uniquement de femelles.
- Le sex ratio idéal pour un effectif efficace maximal serait de 1 (soit autant de femelles que de mâles)

E- Calcul de l'effectif efficace

L'effectif efficace varie aussi selon l'histoire de la population étudiée. Pour un même nombre d'individus pouvant se reproduire, et un sex ratio identique, deux populations peuvent en effet avoir des effectifs efficaces différents selon leur histoire. Ainsi, une population de 10 000 individus aujourd'hui n'aura pas le même N_e en fonction de son histoire démographique (il sera plus faible s'il y a eu à un moment une baisse de l'effectif).

Pour déterminer le N_e en fonction de l'histoire de la population, on fait une estimation de la moyenne géométrique des effectifs réels N au fil des générations.

$$N_e = t / \sum (1/N_i) \text{ (moyenne géométrique des } N_i)$$

Soit, $N_e = t / (1/N_1 + 1/N_2 + 1/N_3 + \dots + 1/N_t)$

La caractéristique de cette moyenne géométrique est qu'elle met en avant les petits effectifs. Ainsi, si à un moment de l'histoire la population s'est retrouvé en sous-nombre, le N_e s'en retrouvera réduit.

Ex : l'effectif d'Homo Sapiens a longtemps été très faible. L'expansion démographique ne date que de 10 000ans (peu de temps). N_e augmente très lentement, beaucoup moins vite que l'effectif réel, dû au poids de toutes les générations qu'il y a eu avec un effectif plus faible.

F- Fixation et élimination d'un nouvel allèle

1) Fixation d'un nouvel allèle :

Cette situation s'observe dans le cas où un nouvel allèle a apparaît au sein d'une population possédant un allèle A.

La probabilité de fixation de l'allèle a sera de : $f(a) = p = 1/2N$

Le temps de fixation : $t = 4N_e$

Ici, la fréquence allélique dépendra uniquement du hasard et non de la sélection naturelle. On observe ainsi que le temps de fixation du nouvel allèle dépendra de N_e . Plus l'effectif efficace sera grand et plus il faudra du temps avant que l'allèle ne se fixe.

2) Élimination d'un nouvel allèle :

Probabilité d'élimination d'un nouvel allèle : $1 - f(a) = 1 - 1/2N$

G- Conséquence de la dérive génétique

1) Au sein de la population

Au sein de la population, la dérive génétique aura pour conséquence une baisse de la diversité génétique avec une fixation de certains allèles au détriment d'autres.

On s'attend à voir une diversité génétique plus importante dans une grande population (avec un grand N_e) que dans une petite population.

2) Entre les populations

La dérive génétique permettra d'augmenter les différences génétiques existant entre deux populations séparées. Ainsi, si deux populations possèdent les mêmes fréquences alléliques à t_0 , on verra au fil des générations des divergences entre elles.

La dérive génétique a donc joué un rôle dans notre histoire et a conduit à la diminution de la diversité génétique au sein des populations, mais à l'augmentation de la diversité génétique entre ces dernières.

H- Effet fondateur

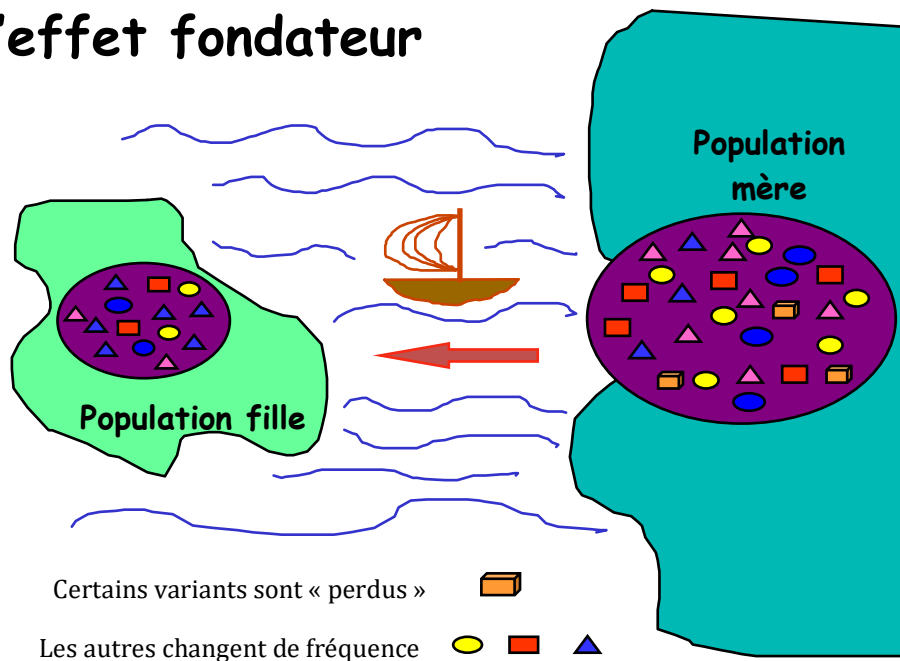
L'effet fondateur est un phénomène que l'on peut observer lors de la constitution d'une nouvelle population à partir d'une population mère. Il correspond à une forme extrême de la dérive génétique (car plus brusque et plus importante que la dérive génétique) et peut s'observer dans le cas de migration (sans être pour autant considéré comme un phénomène de migration (cf ci-dessous)), ou de forte baisse de la population suite à une épidémie ou une guerre. Il s'agit d'un échantillonnage ayant pour conséquence une baisse de la diversité allélique.

Les générations qui vont suivre vont alors subir une forte dérive génétique. D'ailleurs, on peut observer dès la première génération une variation des fréquences alléliques par rapport à celles de la population dite mère. Certains allèles ne seront plus présents dans la population et d'autres auront des fréquences différentes. Plus le nombre de « fondateurs » est petit et plus la dérive sera importante.

Remarque : on peut trouver à cause de ce phénomène une prévalence plus élevée de maladies dans ces populations. En effet, un allèle prédisposant à des maladies peut par hasard avoir une plus grande fréquence dans la nouvelle population que dans la population mère.

Pour résumer, on part d'une grande diversité allélique dans la population mère, et certains allèles vont être perdus, ou gardés et les fréquences alléliques vont être modifiées.

L'effet fondateur



III. Migration

Les migrations sont définies comme des échanges d'individus entre des populations. Ces mouvements sont à l'origine de flux géniques importants entre les populations : les personnes bougent et leurs gènes avec. Il existe différents modèles pour étudier ces phénomènes.

A- Modèle de l'île

Dans ce modèle, on considère deux populations :

- Une population habitant sur le « continent »
- Une population habitant sur une « île »

Les migrations dans ce modèle s'effectuent de manière unidirectionnelle, cad qu'une petite proportion de migrants en provenance du continent se rendent sur l'île, et ce, sur plusieurs générations.

On définit :

- m le taux de migration ($0 < m < 1$) ou encore la proportion d'individu originaire du continent et se reproduisant avec la population insulaire
- p_i la fréquence de l'allèle A dans la population insulaire
- P_i la fréquence de l'allèle A dans la population continentale

On peut ainsi définir p_{i+1} , la fréquence de l'allèle A dans la génération $i+1$ sur l'île comme étant de :

$$p_{i+1} = mP_i + (1-m)p_i$$

Si on considère p_e comme étant la fréquence de l'allèle A au sein de la population insulaire à l'équilibre, on aura alors :

$$p_e = mP_e + (1-m)p_e$$

Après un temps, on verra que les fréquences alléliques insulaires tendront vers celles présentes sur le continent.

B- Migrations multidirectionnelles

Dans cette configuration, nous avons plusieurs populations de même effectif mais avec une fréquence différente d'un allèle selon la population. Ses populations vont alors échanger un taux de migrants m (le même entre toutes les populations) d'une génération à l'autre.

Ce phénomène va conduire à l'uniformisation des différentes fréquences alléliques qui vont donc converger vers une même proportion p dans toutes les populations.

Ce modèle est plus réaliste que celui précédemment vu. Il prend en compte le fait que les migrations peuvent se faire dans les deux sens (bien qu'ici les flux entrants et sortants doivent être égaux). Ces migrations vont permettre l'enrichissement allélique d'une population donnée avec l'arrivée de nouveaux individus possédant des allèles particuliers mais limite fortement la différenciation génétique entre populations (en cela, ce modèle de migrations multidirectionnelles s'opposent au phénomène de dérive génétique).

C- Modèle d'isolement par la distance

Ce modèle de migration est le plus réaliste des trois que nous voyons dans ce cours. En effet, en plus de prendre en compte le fait que les flux migratoires se fassent dans les deux sens, il permet aussi de tenir compte de la distance qui sépare les deux populations. Ainsi, la probabilité d'échange entre population est d'autant plus grand que la distance géographique qui les sépare est faible et vice-versa.

Ce modèle permet ainsi de comprendre :

- L'uniformisation des fréquences alléliques entre des populations interconnectées.
- Les gradients de fréquences alléliques qui sont d'autant plus grands que la distance géographique est importante entre les deux populations.

Conclusion :

- Contrairement aux taux de mutations (qui sont très faibles), les taux de migrations peuvent être très élevés et leurs effets sont souvent importants et rapides.
- L'intensité des échanges migratoires est directement corrélée à la distance géographique entre les populations.
- Les migrations tendent à uniformiser les fréquences alléliques des populations concernées, et s'opposent aux facteurs qui tendent au contraire à les diversifier (dérive génétique)
- Les migrations permettent d'introduire de nouveaux allèles dans une population
- Dans une population de faible effectif, il y a beaucoup de dérive génétique (faible diversité génétique). Si en plus elle est isolée, cette diversité réduite va persister et être amplifiée.

IV. Sélection naturelle

La sélection naturelle est un processus par lequel des variations favorables vont tendre à augmenter en fréquence au sein d'une population, et inversement pour des variations défavorables. Force évolutive proposée par Darwin et qui constitue le principal moteur de l'évolution.

« Pouvons-nous douter... que les individus ayant quelque avantage, aussi léger soit-il, sur les autres, auraient une meilleure chance de survie et de procréer leurs caractères? D'autre part, nous pouvons être sûrs que n'importe quelle variation du moindre degré nuisible serait indéniablement détruite. C'est cette conservation des variations favorables, et le rejet des variations nuisibles, que j'appelle "Sélection Naturelle". Charles Darwin (1859) »

On peut définir de manière plus précise la sélection comme étant la survie ou la reproduction différentielle (capacité à avoir une descendance) des différents phénotypes dans une population, résultant de l'interaction des organismes avec leur environnement. /!\ La sélection naturelle agit sur les phénotypes. Mais puisque les phénotypes sont déterminés par les génotypes, ce phénomène entraîne donc des variations de fréquences alléliques au sein de la population.

Contrairement à la dérive génétique, la sélection naturelle est une force dite déterministe. Il sera donc possible de prévoir l'évolution des fréquences alléliques dans la population.

Remarque : en génétique épidémiologique, on utilise la génétique des populations. Ainsi, si on observe qu'un variant est important sur le plan phénotypique, on en déduit qu'il a dû être ciblé par la sélection naturelle.

A- Fitness : valeur sélective (absolue)

La fitness (ou valeur sélective) est une valeur que l'on attribue à un génotype, à partir du moment où on observe une viabilité et/ou une fécondation différente pour des génotypes différents. La fitness se définit donc comme étant la capacité à pouvoir participer à la création de la génération suivante, et à transmettre son phénotype à la descendance.

On la calcule de la manière suivante :

$$W \text{ (fitness absolue)} = \text{viabilité} \times \text{fécondité}$$

Avec :

- Viabilité : paramètre qui varie de 0 à 1. Taux de survie d'un génotype de la naissance à sa reproduction
- Fécondité : nombre moyen de descendant que produit un individu portant un génotype donné

Ex : On prend une population composée d'individus HMZ AA, aa et HTZ Aa. On décide de calculer les fitness des différents génotypes.

Pour cela on détermine déjà la viabilité des différents types d'individus.

- AA : $v=3/4$
- Aa : $v= 2/3$
- aa : $v = 1/2$

Puis on détermine la fécondité de chaque types d'individus

- AA : $f = 1,5$ (ce qui signifie qu'un individu possédant ce génotype aura en moyenne 1,5 descendants)
- Aa : $f = 1,5$
- aa : $f = 1$

On peut alors calculer la fitness de chaque génotype. On observe alors que $W(AA)$ est plus élevée que les autres fitness. On en déduit donc que la fréquence de cet allèle va augmenter dans les générations suivantes, pour évoluer vers une fixation de l'allèle A.

B- Fitness relative (w)

Il s'agit d'un paramètre qu'on associe aussi aux génotypes, mais qui est relatif. En le calculant la comparaison entre différentes fitness absolues est simplifiée.

Pour la calculer on fait :

$$w \text{ (fitness relative)} = W_{\text{génotype}} / W_{\text{max}}$$

Avec :

- $W_{\text{génotype}}$: fitness absolu d'un génotype donné (ex : Aa, aa ou AA)
 - W_{max} : fitness absolu maximal trouvé parmi l'ensemble des génotypes étudiés
- Ainsi en reprenant l'exemple précédent, on voit que le W_{max} correspond au W_{AA}
 - On peut alors trouver les différentes fitness relatives en calculant W_{AA}/W_{max} , W_{Aa}/W_{max} et W_{aa}/W_{max}

Remarque : la fitness relative du meilleur fitness absolu sera égale à 1

C- Coefficient de sélection (s)

On peut le trouver à partir des valeurs de fitness relatives et absolues et représente la différence entre w et 1.

$$s = 1 - w$$

Il mesure le taux de réduction de la fitness de chaque génotype par rapport à la meilleure dans la population considérée.

Chez l'Homme, les coefficients de sélection sont en général très faibles (rarement supérieur à 1% sauf pour les gènes codant la couleur de la peau ou de la lactase et qui ont permis des évolutions importantes en l'espace de quelques milliers d'années), ce qui signifie que notre génome est « adapté » à notre environnement et qu'il ne nous confère pas de désavantages majeurs pour la survie et la reproduction.

Il faut avoir des coefficients de sélection importants pour observer des évolutions sur de petits laps de temps (quelques milliers d'années).

D- Différents types de sélection

Il existe 3 types de sélections :

- Directionnelle : il s'agit de la sélection proposée par Darwin. Elle a lieu quand un allèle apporte un avantage ou un désavantage. Il est alors facile de prévoir l'évolution des fréquences. L'allèle avantageux va augmenter en fréquence et le désavantageux va tendre à disparaître.
- Balancée : recouvre plusieurs types de sélection, la plus connue étant l'avantage de l'Hétérozygote. Elle s'observe quand la fitness associée au phénotype de l'hétérozygote est supérieure aux fitness des homozygotes. Ce n'est pas un allèle qui est avantageux par rapport à l'autre, mais un génotype par rapport aux autres.
- Diversifiée : ici, ce sont les Hétérozygotes qui représentent un désavantage par rapport aux homozygotes. C'est le contraire de la sélection balancée.

E- Sélection directionnelle

1) Définition

Si l'allèle est avantageux, il va augmenter en fréquence et va se fixer dans la population. Elle conduit à l'augmentation de la fréquence de l'allèle ayant une valeur de fitness absolue importante.

Si on observe un locus neutre (non soumis à la sélection naturelle), on remarque que la sélection directionnelle va aboutir à une réduction de la diversité génétique. On se retrouve donc un peu dans le même cas de la dérive génétique, où on arrive à la fixation d'un allèle plutôt qu'un autre.

Il existe tout de même des différences :

- Dans la sélection, on sait dès le départ quel allèle va se fixer alors que dans la dérive génétique, la fixation d'un allèle se fera au hasard
- Dans la sélection, le processus de fixation sera beaucoup plus rapide que dans la dérive génétique (le processus de fixation est d'autant plus élevé que le coefficient de sélection est important)

2) Exemples

Exemple de CCR5 et de la résistance au VIH :

CCR5 est un récepteur au niveau de la membrane de chacune de nos cellules et utilisé par le VIH pour entrer dans nos cellules. Sans ce récepteur, VIH ne peut pas entrer dans les cellules et l'individu ne peut donc pas être infecté. En pratique, tous le monde a ce récepteur mais il existe des mutations, notamment une délétion de 32pb (mutation $\Delta 32$) aboutissant à une protéine tronquée et non fonctionnelle. Cette protéine tronquée va alors empêcher le virus d'entrer dans les cellules.

On a réalisé différentes études pour observer les effets de cette mutation :

- Un suivi de cohorte au sein d'une population fortement exposée (prostituées) dans lequel on regarde la proportion de personnes séropositives en fonction de leur génotype. On observe alors que les personnes possédant le génotype WT/WT ou WT/ $\Delta 32$ ne sont pas avantagés et qu'aucun homozygotes $\Delta 32$ n'ont été infectés par le VIH. Dans la réalité, on remarque malheureusement que ce n'est pas tout à fait vrai et que des Homozygotes mutants peuvent malgré tout être infectés.
- Une seconde étude s'intéresse ensuite à la durée que prend la maladie à se présenter au sein d'une population infectées par le VIH. (En effet, l'infection par le VIH ne conduit pas immédiatement à l'apparition de la maladie, il faut pour cela plusieurs années en général, le temps pour le virus d'amoindrir le système immunitaire). On remarque alors que la mutation $\Delta 32$ (à l'état Homozygote ou Hétérozygote) confère un avantage aux infectés car la maladie prend plus de temps à apparaître.

On conclut donc que la mutation $\Delta 32$ permet une résistance au VIH (chez les homozygotes) et un retard de développement du virus (chez les porteurs hétérozygotes et homozygotes).

Si on observe la fréquence de cette mutation dans le monde, on remarque qu'elle n'est pas homogène. Elle est en effet beaucoup plus présente en Europe et plus particulièrement au nord-ouest du continent. La fréquence reste malgré tout faible (1% d'homozygotie et 17% d'hétérozygotie).

⇒ Comment expliquer cette différence de répartition ?

- Il est probable que la population mutante ait été exposée à un agent sélectif pathogène autre que le VIH mais utilisant également le CCR5 comme porte d'entrée (la variole ou la peste). Le VIH ayant été introduit très récemment, il est beaucoup trop récent pour induire une variation de la séquence des allèles.
- Ce serait alors à cette époque que la sélection aurait eu lieu, quand la mutation $\Delta 32$ conférait un avantage.

Fiche Récapitulative

I - Mutations

Mutation = force évolutive créant de la variabilité allélique, pour une meilleure adaptation à l'environnement permet la création d'un potentiel adaptatif.

Les mutations peuvent apparaître de façon **spontanées** (erreurs de réplication...) ou **provoquées** (rayons UV...), et peuvent être **neutres** (aucun avantage sélectif), **avantageuses** ou **désavantageuses**.

- Mutation ponctuelles : variation de la séquence nucléotidique de type SNP (insertion/délétion etc..)
- Variations du nombre de copies (moins fréquents) : micro-satellites (répétition de motifs très courts), mini-satellites (motifs plus longs, proche du centromère)
- Remaniements chromosomiques : translocation (échange de matériel génétique entre chromosome non-homologues) modification du nombre de chromosomes (délétion)

Le taux de mutation est faible chez l'Homme. Mais grand potentiel adaptatif du fait de la taille de la population humaine (7 Mds)

Mutations héritées : présentes chez la mère ou le père

Mutation de novo : si elle touche la lignée **germinale** = transmission à la descendance ≠ mutation **somatique** = pas de transmission à la descendance

Théorie de l'évolution des espèces progressive (Lamarck) : les conditions environnementales influent et génèrent l'apparition de nouveaux variants pour s'adapter au mieux à l'environnement et à de nouvelles conditions.

Théorie de Darwin : l'adaptation évolutive arrive **APRES** l'action de la **sélection naturelle** sur ces variations hérissables.

Taux de mutation : nombre de mutations attendues sur un segment d'ADN pendant une période de temps donnée (différentes unités possibles).

Estimation du taux de mutation : méthodes directes

- Évolution expérimentale : accumulation de mutations au fil des générations à partir d'une population g_0 possédant le même matériel génétique

Le taux de mutation est estimé avec la formule : $u = m / (L \times n \times g)$ par nucléotide par génération

u : taux de mutations, m : nb total de mutations observées, L : nb de lignées d'accumulation, n : nb de sites nucléotidiques examinés, g : nb de générations

Cas particulier des Maladies/phénotypes dominants à pénétrance complète : Situation idéale pour l'estimation du taux de mutation : **pathologie monogénique** à **pénétrance complète** sans **hétérogénéité génétique** (un seul locus impliqué) et sans **hétérogénéité allélique** (un seul allèle impliqué)

Estimation du taux de mutation : méthode moléculaire pour les mutations neutres

Principe : comparaison de séquences d'ADN non-fonctionnelles (neutres) entre 2 espèces avec un temps de divergence connue ; repose sur l'horloge moléculaire = accumulation régulière des mutations dans le temps

II – Dérive Génétique

Définition = C'est la **fluctuation aléatoire** des fréquences alléliques dans une population d'effectif limité, liée à l'effet d'échantillonnage.

- Si l'échantillon est de **taille infinie** : fréquences alléliques stables > dans les grandes populations, la **fluctuation des fréquences alléliques est faible**.

- Si l'échantillon est de **taille finie** : fréquences alléliques aléatoires > dans les petites populations, la **fluctuation des fréquences alléliques est grande**

Soit il y a fixation => fréquence = 1 / Soit il y a perte => fréquence = 0 (état absorbant).

L'effectif efficace (N_e) est la **fraction de la population** qui participe réellement à la reproduction ou la taille d'une population idéale qui représente le même niveau de dérive génétique que la population donnée.

Calcul de N_e : estimation par la moyenne géométrique des effectifs réels, de génération en génération

Probabilité de fixation d'un allèle est $f(a) = p = 1/2N$

Le temps de fixation de l'allèle est $t = 4N_e$. Il est plus grand dans une population de grande taille.

Élimination d'un nouvel allèle : probabilité = $1 - f(a)$

Conséquences de la dérive génétique : fluctuation aléatoire des fréquences alléliques au cours du temps => réduction de la diversité génétique en l'absence de mutation, migration ou sélection. On s'attend à une plus grande diversité génétique dans les grandes populations.

Entre des populations, la **dérive génétique** entraîne une divergence progressive (différence des fréquences alléliques de populations différentes, sans migrations).

Effet fondateur : forme extrême de la dérive génétique => on part d'une grande diversité allélique dans la population mère : certains allèles vont être perdus, d'autres gardés et changer de fréquence.

III – Migrations

Modèle de l'île : **migrations unidirectionnelles** d'une population continentale (de grande taille, à composition génétique fixe), vers l'île (petite population, intégrant m migrants à chaque génération)

=> Après un temps, on verra que les fréquences alléliques insulaires tendront vers celles présentes sur le continent.

Migration multidirectionnelle : Plusieurs populations de même effectif avec fréquence d'allèle différente > Échange d'un taux de migrant m d'une génération à l'autre > convergence des fréquences vers une même proportion p dans toutes les populations

- Conséquences : Enrichissement allélique d'une population mais forte limitation de la différenciation génétique entre les populations

Isolement par la distance (modèle le plus réaliste) : la probabilité d'échange entre population est d'autant plus grand que la distance géographique qui les sépare est faible et vice-versa.

IV – Sélection naturelle SN

Définition : la **survie** ou la **reproduction différentielle** (capacité à avoir une descendance) des différents **phénotypes** dans une population, résultant de **l'interaction** des organismes avec leur environnement.

NB : La SN agit sur les phénotypes, déterminés par les génotypes : la SN entraîne donc des variations de fréquences alléliques.

La sélection naturelle est une force **déterministe**

Fitness (=valeur sélective) = capacité à pouvoir participer à la création de la génération suivante, et à transmettre son phénotype à la descendance.

W (fitness absolue) = viabilité x fécondité

Avec :

- Viabilité : paramètre qui varie de 0 à 1. Taux de survie d'un génotype de la naissance à sa reproduction
- Fécondité : nombre moyen de descendant que produit un individu portant un génotype donné

Fitness relative d'un génotype (w) : rapport entre sa valeur de fitness W et la plus forte valeur de fitness (W_{max}) observée dans la population

Coefficient de sélection s = w (fitness relative) - 1

- **Directionnelle** = L'allèle avantageux va augmenter en fréquence et le désavantageux va tendre à disparaître. Aboutit à une réduction de la diversité génétique.
- **Diversifiée** = l'hétérozygote est désavantagé par rapport aux 2 types d'homozygotes
- **Balancée** = la valeur de fitness associée à l'hétérozygote est supérieure aux valeurs des 2 types d'homozygotes

Exemple de CCR5 et de la résistance au VIH : aucun des individus homozygotes pour la mutation du gène CCR5 (dont la protéine est utilisée comme porte d'entrée du VIH pour infecter les cellules) n'est porteur du VIH.

Chez les hétérozygotes porteurs de la mutation et infectés par le VIH : retard du développement du VIH