

Cours de Génétique et dynamique de populations

Lien: meet.google.com/foo-qpnt-gqd

9h à 11h

Génétique et Dynamique des Populations Master I (Protection des Ecosystème)

La génétique des populations s'intéresse aux conséquences de la transmission de l'information, de génération en génération pour la structure d'une population.

La génétique des populations est l'étude 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, recombinaison, mutations, et migration). Les changements de fréquence des allèles sont un aspect majeur de l'évolution, la fixation de certains allèles .

Définitions

1/GENETIQUE MENDELIIENNE

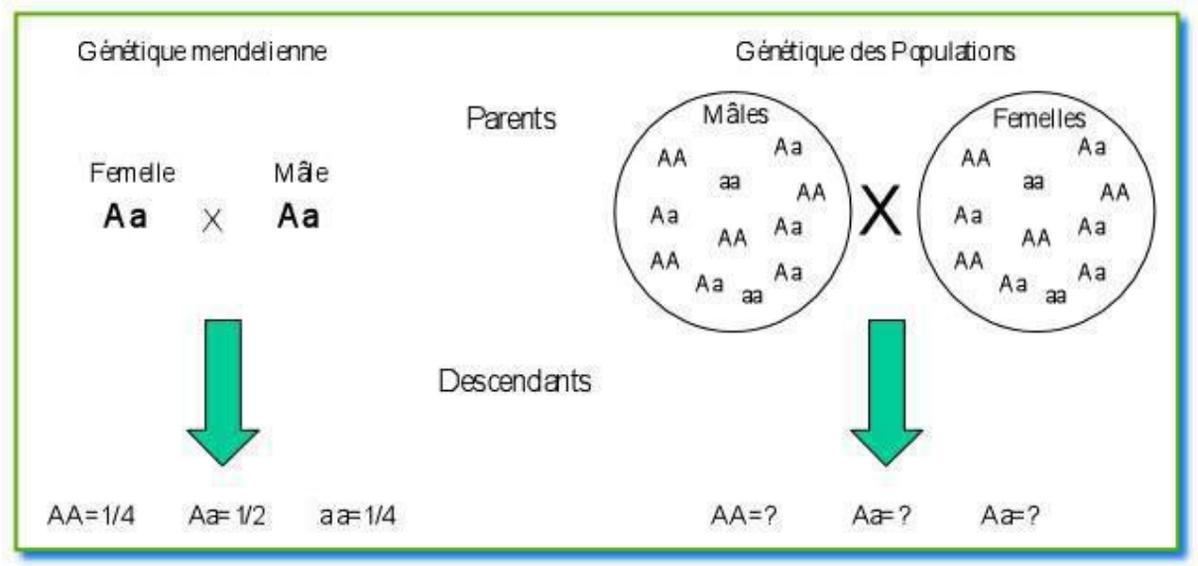
Comprendre le déterminisme et la transmission des caractères par l'analyse de la descendance, à la suite d'un croisement contrôlé par l'expérimentateur , entre individus de génotypes différents.

2/ GENETIQUE MOLÉCULAIRE

La recherche des mécanismes du déterminisme, de l'expression et de la transmission des caractères. Génomique (séquençage e identification des gènes). Protéomique (inventaire et fonctions des protéines).

3/GÉNÉTIQUE DES POPULATIONS

Les individus en interaction avec leur environnement. Une application des principes de base de la génétique mendélienne à l'échelle des populations.



Génétique mendélienne et génétique des populations

Le gène

Définition du gène a évolué avec le développement des connaissances en génétique. La biologie moléculaire permet de définir aujourd'hui le gène comme toute séquence d'ADN transcrite en ARN traduite ou non en peptide. Un gène est un fragment de chromosome qui contient l'information pour l'apparition d'un caractère. L'exemple typique est celui du gène "groupe sanguin ABO". Nous possédons tous ce gène puisque nous avons tous un groupe sanguin ABO (A, B, O ou AB). Mais ce gène existe sous trois versions différentes, appelées allèles : A, B et O.

Allèles

Un gène peut exister dans la population sous différentes formes. Deux gènes sont dits des allèles quand:

- Ils occupent des locus identiques mais sont exclusifs l'un de l'autre. Ils peuvent naître les uns des autres par mutations.

- Ils ont la même fonction (avec une efficacité qui peut être différente).

Allèles d'un gène

Les allèles d'un gène correspondent aux différentes versions de ce gène (leurs séquences de Nucléotides diffèrent).

Il y a quelques nucléotides qui diffèrent entre chacun de ces allèles, mais leur fonction est toujours de permettre l'apparition d'un groupe sanguin de type ABO.

Chaque fois qu'un caractère est soumis à variations (taille, couleurs des yeux ou des cheveux ou de la peau), il y a un gène qui existe sous la forme de plusieurs allèles.

Locus

Les gènes sont disposés de façon linéaire sur les chromosomes. On appelle locus l'emplacement spécifique qu'occupe un gène sur un chromosome. Un même locus peut être occupé par des allèles différents. Locus: emplacement physique précis et invariable sur un chromosome.

Homozygote, hétérozygote

Chez l'homme (organisme diploïde à 46 chromosomes), deux allèles occupent les deux loci correspondants sur les deux chromosomes homologues.

Si les deux allèles sont identiques, on dit que le sujet est homozygote pour le locus considéré.

Si les deux allèles sont différents, le sujet est dit hétérozygote. L'état hémizygotique désigne tous les gènes portés par le chromosome X chez le sexe masculin (présent en un seul exemplaire).

Pour un locus donné avec deux allèles **A** et **a**, 3 génotypes sont possibles:

- **AA** et **aa** Homozygotes
- **Aa** Hétérozygote.

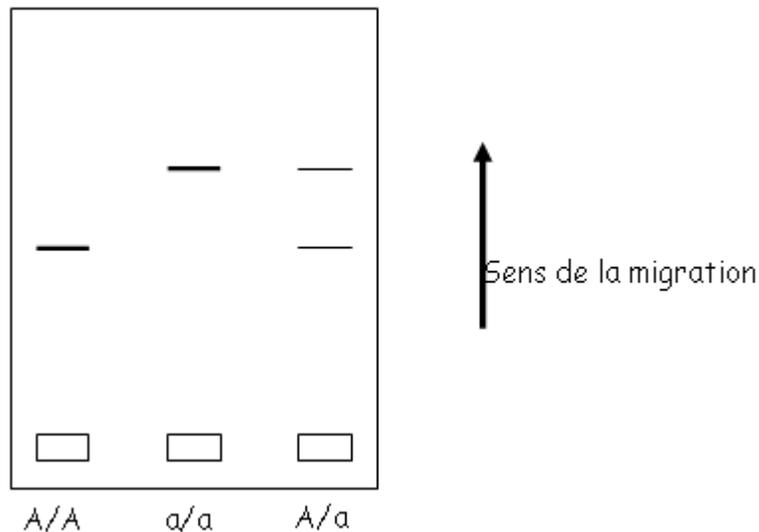
Si dans une population un gène existe sous forme de **n allèles**, le nombre de génotypes possibles au niveau de ce locus est **$n(n+1) / 2$** .

S'il y a une relation de dominance (**A**) /récessivité (**a**), on a deux génotypes. **A/A** et **A/a** donnent le phénotype [**A**] et **a/a** donne le phénotype [**a**].

Si, pour un gène, il y a de nombreux allèles différents, on va parler de série allélique.

Si dans une population un gène existe sous forme de **n allèles**, le nombre de génotypes possibles au niveau de ce locus est **$n(n+1) / 2$** .

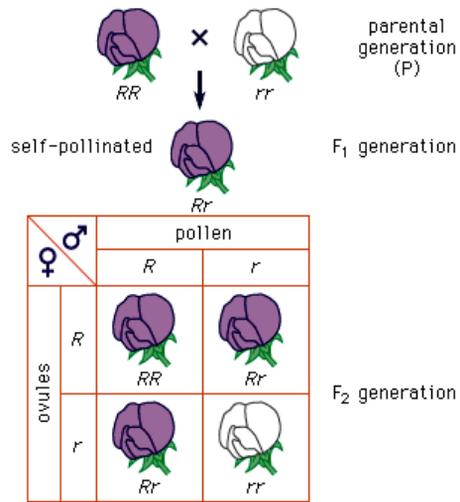
Si dans une population un gène existe sous forme de **n allèles**, le nombre de génotypes possibles au niveau de ce locus est **$n(n+1) / 2$** .



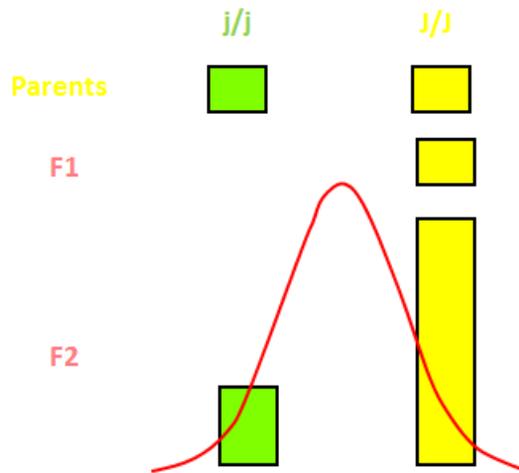
Charles Darwin 1809-1882

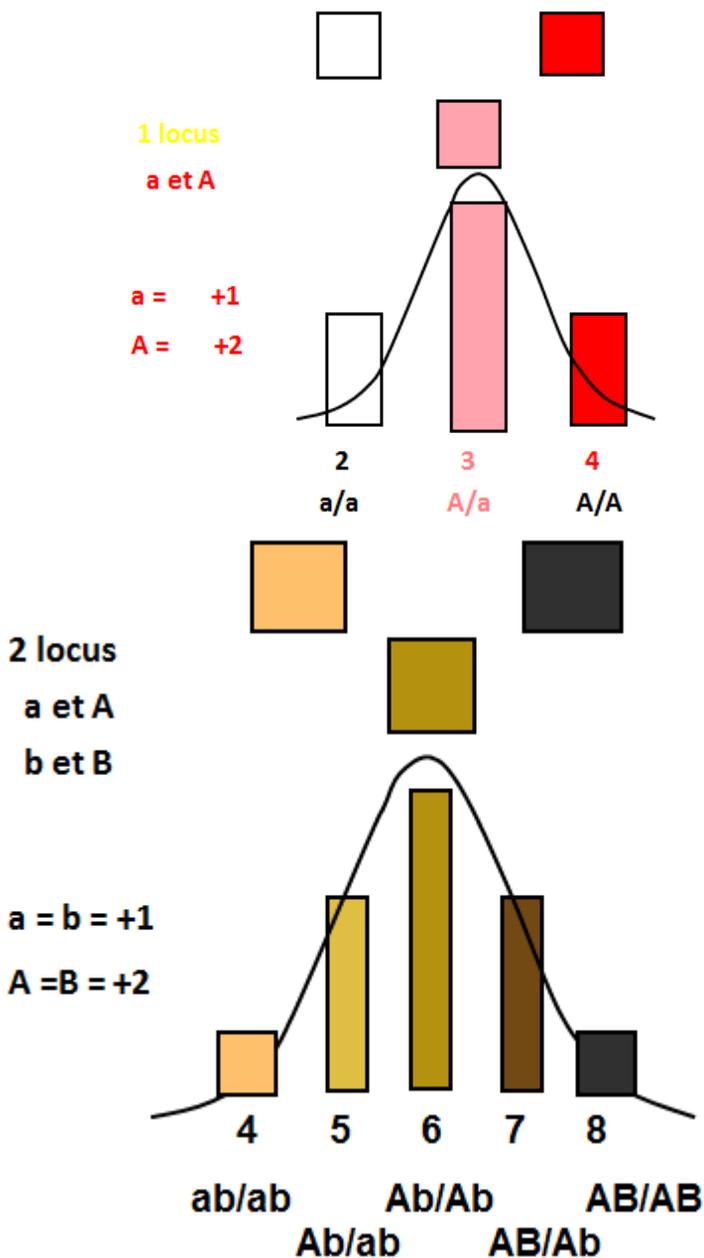
Variation >> Hérité
Sélection >> *in natura*
Espèces >> Spéciation
Hasard >> Mutations
Temps >> Durée
paléontologique
Embryologie
Comportement
(altruisme, émotions...)

Gregor Mendel 1809-1882



© 2006 Encyclopædia Britannica, Inc.





Une modification génétique de la population, et l'accumulation de tels changements dans différentes populations peut conduire au processus de spéciation

Discipline initiée dans les années 1920 à 1940 par Ronald Fisher, J.B.S.Haldane et Sewall Wright, la génétique des populations est une application des principes fondamentaux de la génétique mendélienne à l'échelle des populations.

Cette application a permis de faire la synthèse entre la génétique mendélienne et la théorie de l'évolution, donnant ainsi naissance au néo-darwinisme (théorie synthétique de l'évolution) et à la génétique quantitative.

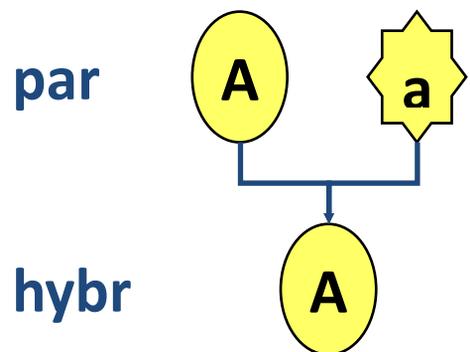
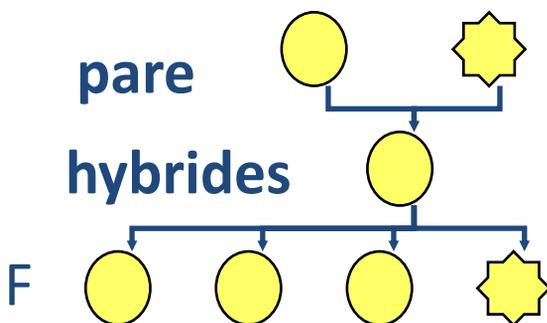
La génétique des populations a des applications en épidémiologie où elle permet de comprendre la transmission des maladies génétiques, mais aussi en agronomie, où des programmes de sélection modifient le patrimoine génétique de certains organismes pour créer des races ou variétés plus performantes, ou plus résistantes à des maladies.

Elle permet également de comprendre les mécanismes de conservation et de disparition des populations et des espèces (Génétique de la conservation). C'est une discipline des sciences de la vie faisant un fort usage d'outils mathématiques.

Principe de base

La valeur phénotypique (**P**) d'un individu, c'est-à-dire le résultat de la mesure effectuée sur un individu, est l'effet combiné de la valeur génotypique (**G**) et de l'effet environnemental (**E**)

La génétique initiée par Gregor Mendel, appelée classiquement génétique mendélienne, a pour objectif de comprendre le déterminisme et la transmission des caractères par l'analyse de la descendance d'un croisement contrôlé entre individus de génotype différent (proportions des diverses catégories de descendants).



Après la découverte du support de l'information génétique (ADN), la génétique moléculaire continue à rechercher les mécanismes fins du déterminisme, de l'expression et de la transmission des caractères. Elle trouve aujourd'hui de nombreuses extensions avec les programmes de génomique (séquençage des génomes et identification des gènes) et de protéomique (inventaire et fonction des protéines d'un organisme).

La compréhension du déterminisme et de la transmission des caractères doit aussi étudier les individus dans les conditions naturelles où ils sont génétiquement uniques et libres de se reproduire avec n'importe quel autre individu de la même espèce. Cette partie de la génétique, qui considère les individus en interactions avec leur environnement, est la génétique des populations.

2. Objectifs de la génétique des populations

La génétique des populations étudie la variabilité génétique présente dans et entre les populations avec 3 principaux objectifs :

- 1) Mesurer la variabilité génétique, appelé aussi diversité génétique, par la fréquence des différents allèles d'un même gène.
- 2) Comprendre comment la variabilité génétique se transmet d'une génération à l'autre.
- 3) Comprendre comment et pourquoi la variabilité génétique évolue au fil des générations.

A la différence de la génétique mendélienne, la génétique des populations étudie les proportions des génotypes au sein d'un ensemble d'individus issus de croisements non contrôlés entre de nombreux parents.

3. Définition d'une population

Une population est l'ensemble des individus de la même espèce qui ont la possibilité d'interagir entre eux au moment de la reproduction. La notion de population fait donc appel à des critères d'ordre spatiaux, temporels et génétiques et résulte du fait que les individus d'une même espèce n'ont pas tous la possibilité de se rencontrer et de se croiser à cause de l'éloignement géographique et de l'hétérogénéité de l'habitat. La population représente une communauté génétique constituée par l'ensemble des génotypes des individus qui la composent. La population se caractérise donc par un génome collectif ou patrimoine

génétique, appelé aussi pool génétique qui est la somme des génotypes individuels pour chacun des gènes. Si chaque génotype individuel est fixé définitivement à la naissance et cesse d'exister à la mort de l'individu, le pool génétique d'une population présente une continuité à travers les générations, et peut varier au cours du temps. C'est cette évolution que la génétique des populations cherche à comprendre. La population est à distinguer de la notion d'espèce qui rassemble tous les individus interfertiles même si ceux-ci n'ont jamais la possibilité de se croiser. C'est l'unité d'étude dans de nombreux domaines des sciences de la vie (épidémiologie, évolution, écologie, biogéographie, biologie de la conservation).

Simple au plan théorique, cette définition est souvent difficile à appliquer aux situations naturelles.

Les limites d'une population sont incertaines et dépendent des caractéristiques intrinsèques des espèces (répartition spatiale et temporelle des individus, mobilité, mode de reproduction, durée de vie, socialité, etc).Lorsqu'une espèce présente de très grands effectifs et occupe un vaste territoire apparemment homogène, seule l'étude détaillée de la distribution des individus, de leurs comportements, de leurs déplacements et de leurs génotypes peut permettre de déceler d'éventuelles discontinuités correspondant à des limites de populations.

Chapitre II. Variabilité génétique dans les populations naturelles.

1. Introduction

Une particularité du monde vivant est la variabilité des phénotypes individuels. A l'intérieur d'une espèce, il n'existe pas 2 individus ayant exactement les mêmes caractéristiques phénotypiques : l'individu est unique. Si pour une espèce donnée on peut noter l'absence de variations pour certains caractères essentiels, il existe toujours de nombreux autres caractères pour lesquels des variations entre individus sont observées.

Certaines de ces variations s'expriment au niveau phénotypique (morphologie, physiologie, comportement, etc) mais les autres restent "**cachées**" et leur mise en évidence nécessite l'utilisation de techniques adaptées (**variabilité des protéines** ou des **séquences d'ADN**).

Les variations du phénotype sont dues pour partie à des facteurs environnementaux (alimentation, climat, interactions avec les autres espèces, etc) et pour partie à des différences entre les génotypes individuels, transmissibles à la descendance. Dans la plupart des cas, ces deux causes de variation interagissent fortement (= interactions génotype-environnement), et il est difficile de mesurer leur part relative dans la variation phénotypique globale.

La mise en évidence du déterminisme génétique des variations nécessite des études faisant appel soit à des expériences de croisements, soit à des analyses de généalogie, soit, pour les caractères complexes déterminés par plusieurs gènes, des comparaisons entre individus

apparentés et non apparentés à l'aide de méthodes statistiques qui sont du domaine de la génétique quantitative.

2. Déterminisme des variations et notion de polymorphisme

La variabilité d'un caractère est déterminée génétiquement lorsqu'elle est due, au moins en partie, à la présence de plusieurs formes alléliques dans la population.

Dans certains cas, la variabilité phénotypique est due à la variation d'un seul gène = déterminisme **monogénique**, cela ne veut pas dire que le caractère est contrôlé par un seul

gène mais que la variation d'un seul de ces gènes est suffisante pour entraîner une variation phénotypique. On parle alors de caractères mendéliens.

2.1 Déterminisme génétique

Le caractère présente alors une plasticité phénotypique. De telles variations épigénétiques sont très fréquentes dans les populations animales et végétales. Chez la sagittaire *Sagittaria sagittifolia*, la forme des feuilles varie en fonction du degré d'immersion de la plante.

Hors de l'eau, les feuilles ont la forme d'un fer de lance, elles sont arrondies à la surface de l'eau et prennent sous l'eau l'aspect de longues lanières.



Chez d'autres plantes, c'est la nature du sol qui peut être à l'origine d'une variation de la couleur des fleurs comme chez le mourron *Anagallis arvensis* (Primulacée).



Il est probable que dans l'avenir cette forme d'hérédité prenne de plus en plus d'importance comme l'illustre l'exemple du polymorphisme de la symétrie des fleurs chez la Linaire (*Linaria vulgaris*). Chez cette espèce, deux formes avaient été décrites par Linné (1707-1778) : une forme à symétrie bilatérale et une forme "peloric" à symétrie radiale.



L'analyse génétique de ces caractères relève de la génétique quantitative qui sépare les effets des gènes en effets additifs A, effets de dominance D, effet d'épistasie ou d'interaction entre gènes I : $G = A + D + I$

3. Les mutations source de variabilité

La variabilité génétique est le résultat des mutations qui font apparaître de nouveaux allèles, auxquelles il faut ajouter les phénomènes de recombinaison (notamment pour les caractères quantitatifs). Les mutations peuvent affecter une portion plus ou moins grande d'ADN et, en fonction de leur localisation dans le génome, peuvent avoir ou non des effets phénotypiques.

Il existe ainsi tous les intermédiaires entre les mutations neutres qui n'ont aucun effet sur l'organisme et les mutations létales, qui réduisent l'espérance de vie des individus.

Il existe différents types moléculaires de mutations qui n'ont pas les mêmes conséquences phénotypiques:

3.1. Les mutations ponctuelles

Sont des modifications d'un nucléotide (ou d'un faible nombre de nucléotides) qui créent de nouveaux allèles. Il faut distinguer :

- 1) Les insertions de nucléotides qui, lorsqu'elles se produisent dans une portion codante de l'ADN, décalent le cadre de lecture et conduisent à une protéine anormale.
- 2) Les délétions de nucléotides qui ont les mêmes effets que les insertions.
- 3) Les substitutions d'une base par une autre qui peuvent être des transitions (remplacement purine/purine de **A** avec **G** ou pyrimidine/pyrimidine de **C** avec **T**) ou plus rarement des transversions (remplacement **purine/pyrimidine**).

Les substitutions en 3^{ème} position des codons sont silencieuses ou synonymes alors que la plupart des substitutions en position 1 et 2 des codons se traduisent par un remplacement d'acide aminé (**non synonymes**).

3.2. Les remaniements chromosomiques

Sont des modifications dans la structure des chromosomes. Les changements concernent un fragment chromosomique dont la taille peut correspondre à une partie ou plusieurs gènes et donc qui sont souvent très défavorables. Les différents types sont :

- 1) Les duplications défavorables lorsqu'elles se produisent à l'intérieur d'un gène mais qui peuvent augmenter le nombre de copies d'un gène lorsqu'elles concernent un plus grand segment chromosomique.
- 2) Les inversions qui correspondent à un changement d'orientation d'un fragment chromosomique et qui modifient l'ordre des gènes.
- 3) Les délétions qui sont des pertes d'un fragment chromosomique ayant le plus souvent des effets létaux car elles peuvent concerner un ou plusieurs gènes.
- 4) Les translocations qui correspondent à des échanges de fragments entre chromosomes.

3.3. Les changements du nombre de chromosomes sont de deux types :

- 1) L'aneuploïdie : perte ou ajout d'un ou plusieurs chromosomes (Par exemple la trisomie = $2N+1$).
- 2) La *polyploïdie* : changement du nombre d'exemplaire du lot haploïde (Passage **diploïde** = $2N$ à tétraploïde = $4N$).

4. Du génotype aux phénotypes

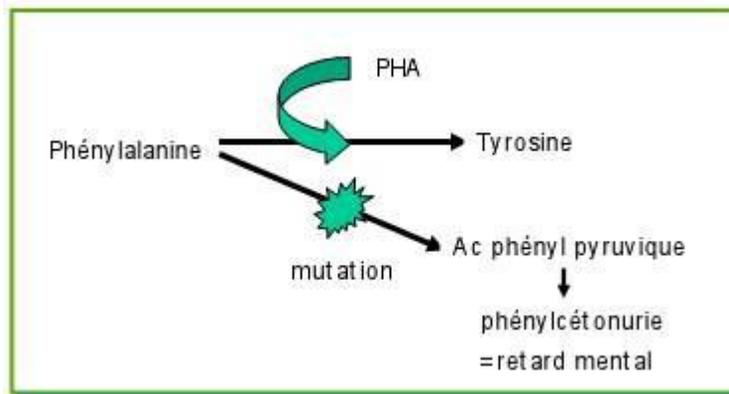
Parmi l'ensemble des mutations qui affectent le génome d'un organisme, seule une partie ont des conséquences phénotypiques. L'absence d'effet sur le phénotype peut être la conséquence de mutations dans une région non codante de l'ADN ou de mutations dans des gènes qui sont présents en plusieurs exemplaires dans le génome (redondance des gènes). Ces mutations sont qualifiées de neutres.

Lorsque les mutations ont des effets sur le phénotype des individus, elles peuvent modifier des caractères biochimiques, physiologiques, anatomiques, morphologiques ou comportementaux. Les mécanismes mis en jeu dans leur expression phénotypique sont divers et complexes. L'expression phénotypique d'un génotype dépend des conditions

environnementales dans lesquelles se sont développés les individus. Pour la plupart des caractères, le phénotype résulte des effets conjoints de **3 composantes** :

- Le génotype **G**.
- L'environnement **E** qui contribue toujours pour une part au phénotype.
- L'interaction entre le génotype et l'environnement **IGxE** ceci est résumé dans une formulation additive: **P = G + E + IGxE**

Cette interaction entre le génotype et l'environnement est très importante car elle signifie que l'expression d'un gène n'est pas indépendante du milieu dans lequel ce gène s'exprime. Une même mutation peut donc avoir des effets phénotypiques différents. L'effet de l'environnement sur l'expression phénotypique d'un caractère peut être illustré par l'exemple de la phénylcétonurie chez l'homme, maladie récessive due à une mutation du gène codant pour la phénylalanine hydroxylase (**PHA**). Chez les homozygotes récessifs, cette enzyme ne dégrade plus la phénylalanine en tyrosine et il se produit une accumulation d'acidephényl pyruvique, toxique, qui affecte le développement du système nerveux des jeunes enfants. Les individus atteints présentent alors un grave retard Mental (idiotie phénylpyruvique).



Le diagnostic précoce des individus homozygotes récessifs et la mise en place d'un régime alimentaire adapté, pauvre en phénylalanine, permet le développement normal du système nerveux des jeunes enfants qui deviennent insensibles à la phénylalanine à l'âge de l'adolescence. Ce test est effectué chez tous les nouveau-nés = **test de Guthrie**.

Déterminisme des variations :

Notion de polymorphisme

La génétique des populations s'intéresse principalement à la variabilité d'origine génétique présente dans les populations et que l'on désigne sous le nom de polymorphisme. Dans sa définition historique (Ford années 1940), le polymorphisme concernait les caractéristiques phénotypiques accessibles aux observations de cette époque (couleur, forme, etc). Cette définition du polymorphisme peut être résumée de la façon suivante:

Il y a polymorphisme si dans une même population coexistent pour un caractère donné plusieurs formes phénotypiques discontinues, déterminées génétiquement, et dont la plus fréquente ne représente pas plus d'une certaine fraction de la population totale, fixée à 95 ou 99%. La population est alors qualifiée de polymorphe.

L'utilisation de plus en plus répandue des techniques de biologie moléculaire permettant d'étudier la variabilité non exprimée au niveau phénotypique (portion non codante de d'ADN) a nécessité une définition plus large du polymorphisme qui peut être la suivante :

Il y a polymorphisme si dans une même population une portion codante ou non codante d'ADN présente une variation de séquence correspondant à plusieurs formes alléliques dont la plus fréquente ne représente pas plus d'une certaine fraction de la population totale, fixée à 95 ou 99%.

Dans ces deux définitions, le seuil de 1% ou 5% permet de distinguer les gènes polymorphes, pour lesquels les variations alléliques sont fréquentes, et les gènes pour lesquels les variations alléliques ont un caractère exceptionnel avec un allèle très majoritaire et une ou plusieurs formes alléliques rares (inférieure à 1%). On parle dans ce cas de cryptopolymorphisme qui résulte le plus souvent de mutations désavantageuses qui seraient éliminées par la sélection naturelle. La plupart des maladies génétiques chez l'homme relèvent du Cryptopolymorphisme. Par opposition, on appelle monomorphes les gènes qui ne présentent pas de variabilité (un seul allèle présent dans la population).

L'état polymorphe ou monomorphe est une caractéristique d'un gène (ou portion non codante d'ADN) d'une population. Ainsi, une même population peut être polymorphe pour un

caractère donné et monomorphe pour un autre caractère. De la même façon, un caractère monomorphe dans une population peut être polymorphe dans une autre population.

2.1. Types de Polymorphisme

2.1.1. Polymorphisme morphologique

C'est le polymorphisme de taille, de forme, de couleur etc. La variabilité génétique de la couleur de certaines espèces, appelée polychromatisme, est certainement l'un des polymorphismes qui a été le plus étudié.

Un exemple célèbre est la variation de la couleur et de l'ornementation de la coquille de l'escargot du genre *Cepaea*. En un même endroit coexistent plusieurs formes phénotypiques déterminées par plusieurs gènes polymorphes : des escargots à coquille rose, jaune ou brune, et des escargots sans bande et avec bandes dont le nombre varie entre 1 et 5. Ces variations sont sous le contrôle de quatre gènes principaux entre lesquels existent des relations d'épistasie :

- Le gène C, multi-allélique, détermine la couleur. Par exemple, l'allèle CR (couleur rose) est dominant sur l'allèle CJ (couleur jaune).
- Le gène B détermine la présence ou l'absence des 5 bandes : l'allèle B0 (absence de bandes) est dominant sur l'allèle Bb (présence des 5 bandes).
- Le gène U supprimeur des bandes 1, 2, 4 et 5. Cette inhibition est due à un allèle U3 dominant sur l'allèle U.
- Le gène T supprimeur des bandes 1 et 2. Cette inhibition est due à un allèle T345 dominant sur l'allèle T.

Les gènes B, U et T sont en interaction par les relations d'épistasie suivantes : le gène B est épistatique sur le gène U qui est lui-même épistatique sur le gène T ($B > U > T$).



Cette variation de la couleur de la coquille se retrouve chez un très grand nombre d'espèces de mollusques avec parfois une très grande diversité génétique comme c'est le cas chez *Liguus fasciatus*.



Un autre exemple de polychromatisme, qui a fait l'objet de très nombreuses études de génétique des populations, est celui observé chez le papillon *Biston betularia*. Ces papillons de nuit sont normalement de couleur claire, légèrement tachetée, ce qui les rend mimétiques le jour lorsqu'ils se reposent sur le tronc des arbres. Dans certaines régions où le tronc des arbres est plus sombre, les populations sont caractérisées par une forte fréquence de papillons

sombres presque noirs, appelés "melanica", due à un allèle D, dominant sur l'allèle d.



Chez l'homme, un grand nombre de caractères morphologiques sont polymorphes, avec des fréquences élevées des différentes formes. C'est le cas de la couleur des yeux ou de la peau, de la forme des oreilles.

2.2. Polymorphisme des protéines

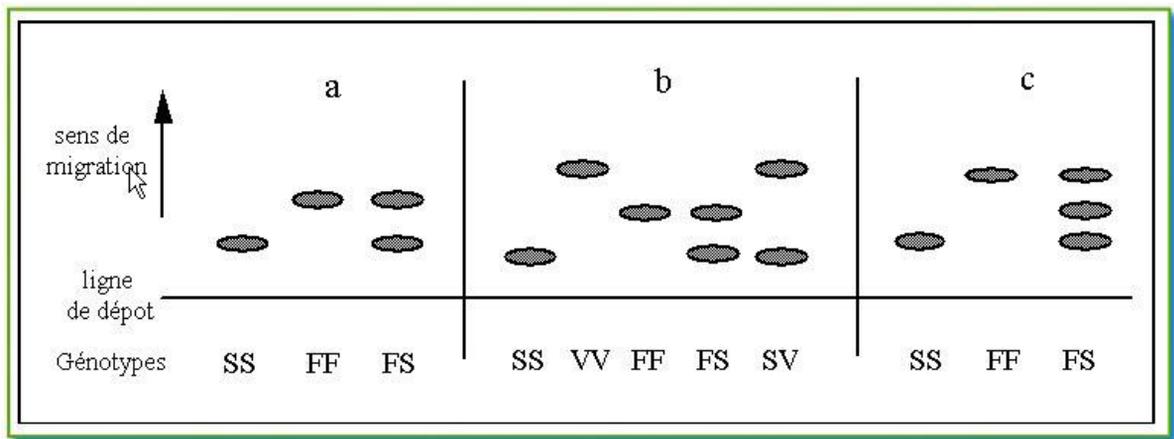
2.2.1. Polymorphisme enzymatique

Depuis les années 1960, la variabilité des protéines est étudiée par électrophorèse. Les protéines sont des molécules chargées qui se déplacent dans un support poreux (gel d'agarose, d'amidon, de polyacrylamide, d'acétate de cellulose) lorsqu'elle celui-ci est soumis à un champ électrique.

La vitesse de migration dépend de la charge globale de la protéine, de sa taille et de sa conformation. Toute mutation dans la séquence d'un gène codant pour une protéine peut modifier le sens d'un codon, altérer la séquence d'acides aminés donc la charge électrique de la protéine et sa vitesse de migration. Ce changement de structure primaire peut être détecté par électrophorèse qui sépare les variants protéiques ayant des vitesses de migration différentes appelées souvent F (fast) et S (slow).



La mise en évidence de différents allèles d'un même gène est possible pour les enzymes grâce à la spécificité de la réaction enzyme-substrat visualisée par une réaction colorée. L'existence de variations génétiques à un locus donné est détectée par la présence de différents niveaux de migration dans le gel d'électrophorèse, qui sont associés à des allèles différents appelés allozymes.



Représentation schématique d'un gel d'électrophorèse pour différents systèmes génétiques

- a) cas d'une protéine monomérique codée par un gène à deux allèles (F = Fast et S = Slow);
- b) cas d'une protéine monomérique codée par un gène à trois allèles (V = very Fast, F = Fast et S = Slow);
- c) cas d'une protéine dimérique codée par un gène à deux allèles (F = Fast et S = Slow) ou les hétérozygotes sont représentés par 3 bandes.

2.2.2. Polymorphisme immunologique

La variabilité de certaines protéines peut être étudiée par des techniques d'immunologie. Classiquement, il s'agit de mesurer la spécificité et l'affinité des réactions antigènes-anticorps lorsque l'on fait réagir un anticorps, produit contre un antigène défini, avec des antigènes d'origines variées (hétérologues).

Chez l'homme, le polymorphisme immunologique le plus étudié est celui des antigènes présents à la surface des globules rouges dont les plus connus sont le système ABO, le système rhésus (allèle Rh⁺ dominant sur Rh⁻), le système MN (M et N codominants). Pour le système ABO, les allèles A et B sont codominants entre eux et tous les deux dominants sur l'allèle O ce qui donne la typologie antigènes/anticorps suivante

Génotype	Antigène	Anticorps
IA IA IA IO	Groupe A	Anti B
IB IB IB IO	Groupe B	Anti A
IA IB	Groupe AB	Ni Anti A, Anti B Receveur universel
IO IO	Groupe OO	Anti B, anti A Donneur universel

De fortes variations géographiques existent pour les fréquences des allèles du système ABO à l'échelle des continents.

Un autre polymorphisme immunologique bien connu chez l'homme est celui du système HLA (Human Leucocyte Antigen), appelé aussi complexe majeur d'histocompatibilité (CMH), mis en évidence au niveau des leucocytes (Globules blancs) et des plaquettes sanguines. Ce polymorphisme implique 6 gènes étroitement liés, portés par le chromosome 6.

Chaque gène comporte de très nombreux allèles, ce qui conduit à une diversité quasi infinie des combinaisons ce qui assure l'identité immunitaire de chaque individu.

2.2.3. Polymorphisme chromosomique

Ce polymorphisme peut être dû soit à une variation du nombre des chromosomes (euploïdie, aneuploïdie) soit à un changement de leur structure (délétion, duplication, inversion, translocation). Par exemple, chez une graminée *Dactylis glomerata*, il existe plusieurs catégories d'individus, certains étant diploïdes c'est-à-dire ayant $2N$ chromosomes, d'autre tétraploïdes à $4N$ chromosomes. Un autre exemple de polymorphisme chromosomique bien connu est celui des inversions chromosomiques observées chez la drosophile américaine *Drosophila pseudoobscura*. De très nombreuses inversions différentes ont été observées chez cette espèce de drosophile. Des études menées par Dobzhansky T ont montré que les populations de *D. pseudoobscura* sont extrêmement polymorphes pour certaines de ces inversions et de fortes différences de fréquence entre populations d'origine géographique différente sont observées avec, semble-t-il, une corrélation avec les facteurs climatiques (température).

2.2.4. Polymorphisme de l'ADN

Les techniques issues de la biologie moléculaire permettent de rechercher des variations dans les séquences nucléotidiques de l'ADN et sont de plus en plus utilisées pour étudier le fonctionnement génétique des populations. Cette variabilité peut être recherchée dans des régions codantes de l'ADN mais de très nombreuses techniques permettent d'étudier le polymorphisme des régions non codantes qui composent la grande majorité des génomes. Cette variabilité, qui n'est généralement pas exprimée au niveau phénotypique, est utilisée pour définir des marqueurs permettant soit de caractériser des individus = empreinte génétique (ou fingerprint), soit de caractériser des populations, soit de cartographier des gènes.

Parmi l'ensemble des techniques disponibles, il faut distinguer celles qui permettent de mettre en évidence une variabilité dispersée dans tout le génome (les marqueurs révélés sont alors multilocus et dominants) de celles qui permettent de révéler une variabilité à des endroits plus limités du génome (les marqueurs sont souvent monolocus et codominants).

Il faut également distinguer parmi ces techniques celles qui nécessitent uniquement une extraction de l'ADN des individus étudiés de celles qui nécessitent une amplification in vitro d'une portion définie d'ADN par PCR (Polymerase Chain Reaction).

Sans être exhaustif, les principaux marqueurs moléculaires utilisés en génétique des populations sont les suivants.

2.2.4. Polymorphisme RFLP (Restriction Fragment Length Polymorphism) et PCR-RFPL

Après extraction, l'ADN est soumis à une (ou des) enzyme de restriction qui coupe la molécule à des endroits précis, définis par une séquence de bases, appelé sites de restriction. Toute modification par mutation dans la séquence du site de restriction empêche l'action de l'enzyme.

Cette non-coupeure de l'ADN est détectée par une variation du nombre et de la longueur des fragments d'ADN (fragments de restriction) obtenus après digestion enzymatique puis séparation par électrophorèse et visualisation par hybridation avec une sonde radioactive ou fluorescente (**Figure 3**).

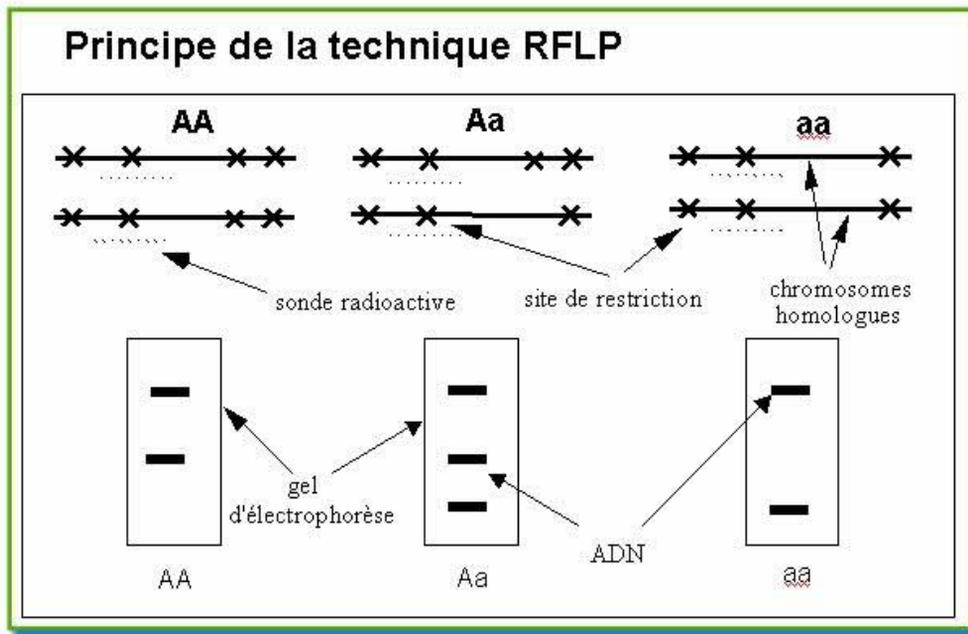
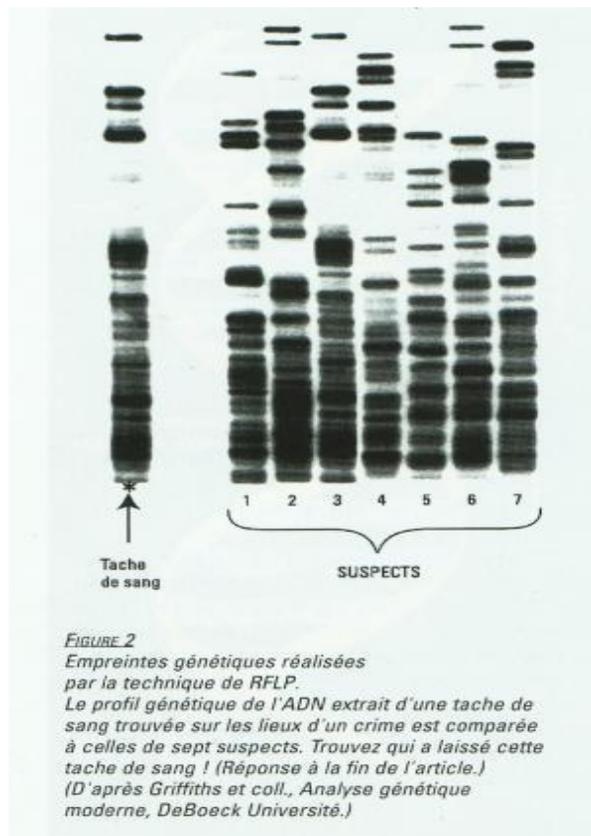
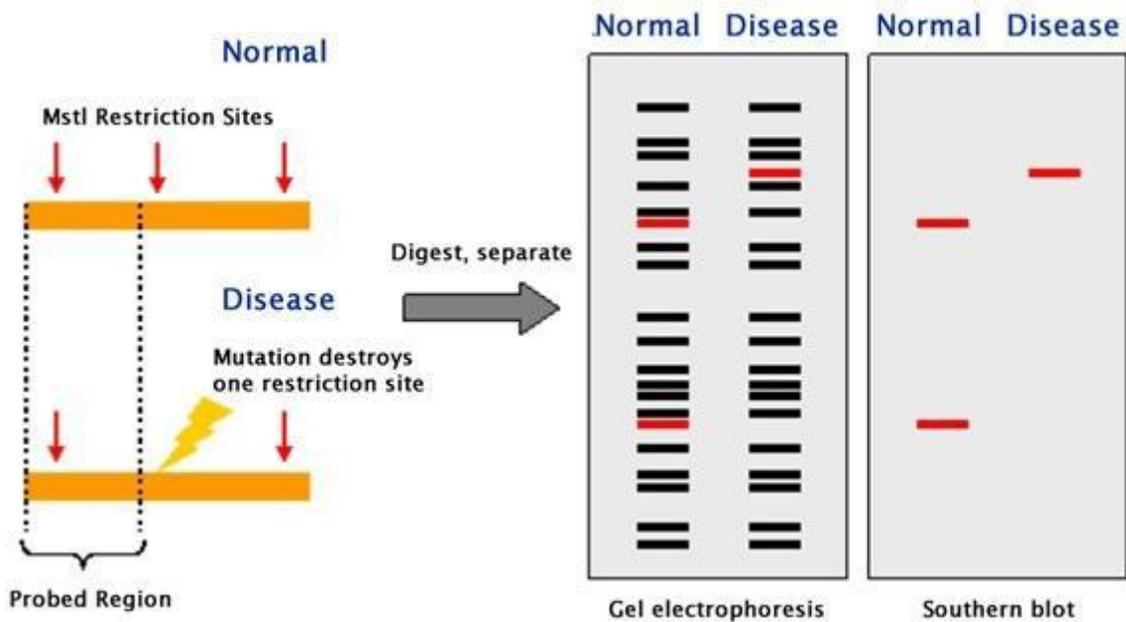


Figure 3: Principe de la technique RFLP.

Principe

Les endonucléases de Restriction sont des enzymes qui coupent l'ADN prolongé en pièces courtes. Chaque endonucléase de restriction vise différentes séquences de nucléotides dans un Brin d'ADN et coupe pour cette raison à différents sites. La distance entre les sites de clivage d'une certaine endonucléase de restriction diffère entre les personnes. Par Conséquent, la longueur des Fragments d'ADN produits par une endonucléase de restriction différera en travers des deux différents organismes et substances.



La majorité de ces enzymes coupent le DNA une fois qu'ils ont reconnu une séquence particulière dans celui ci (par exemple l'enzyme EcoR1 coupe dès qu'il repère la séquence GATC).

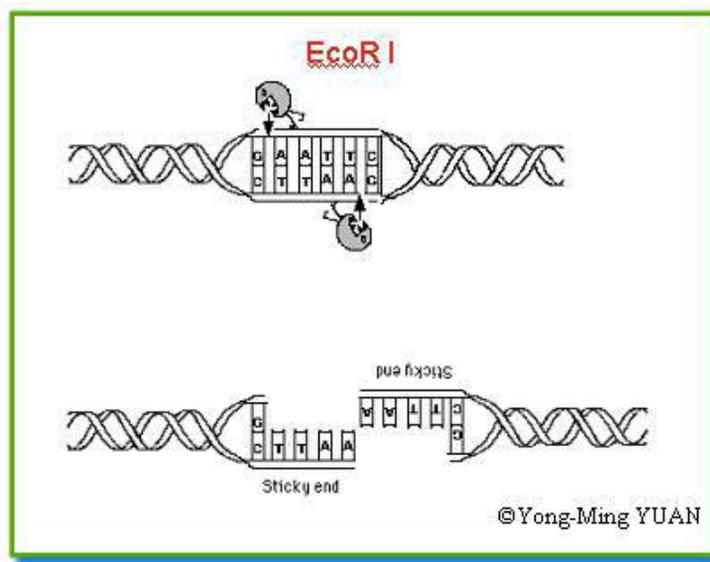
d'abord un fragment de restriction est un morceau de DNA qui a été coupé par un enzyme de restriction (enzyme coupant le DNA).

Grâce à cette méthode on peut caractériser des morceaux de DNA. Chaque gène, ayant un contenu en base spécifique, aura un nombre de fragments spécifique si on décide de le "découper" à l'aide de l'enzyme.

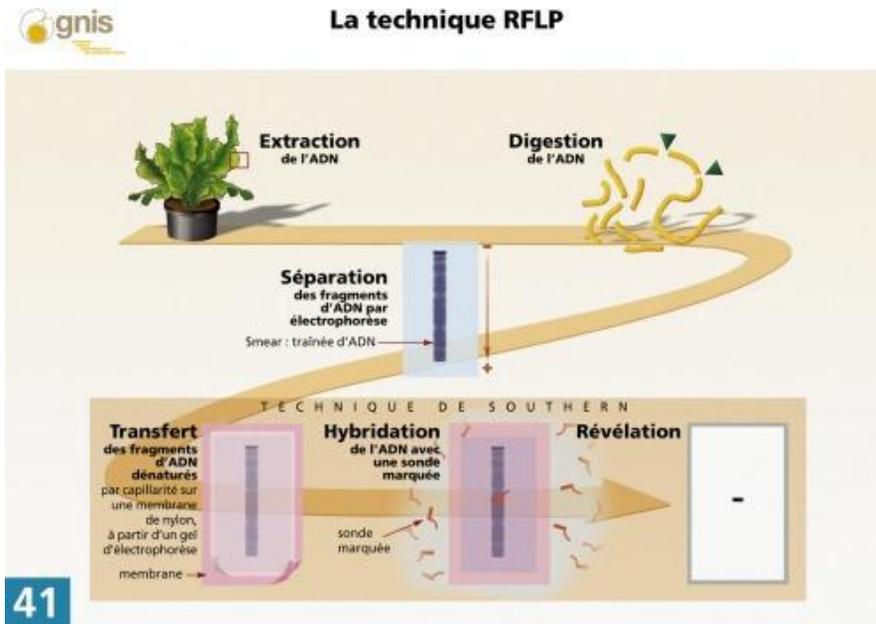
Si le gène A possède 2 fois la séquence GATC il donnera 3 fragments et si le gène B en possède 4 il donnera 5 fragments (et ainsi de suite).

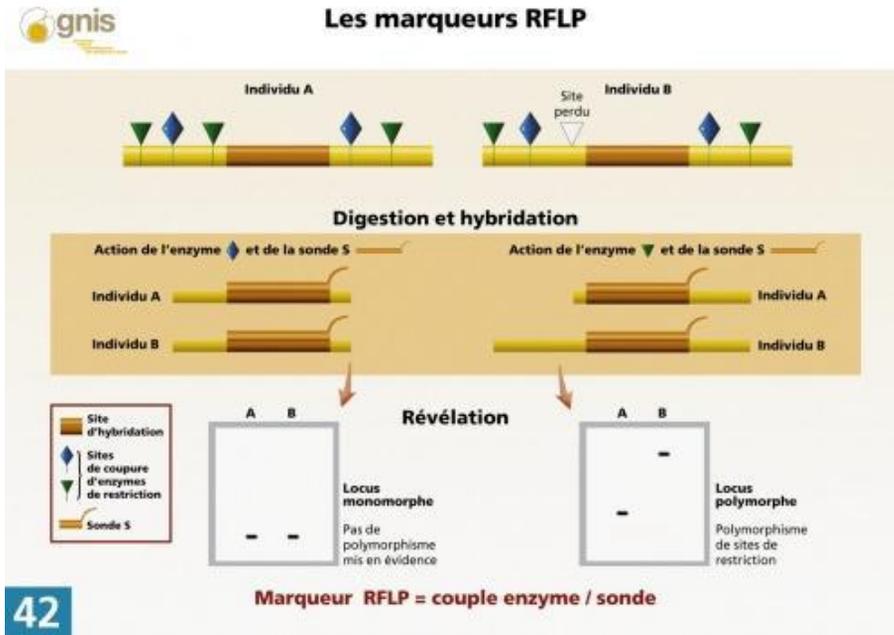
En utilisant un enzyme particulier on sait qu'on doit s'attendre à trouver une série précise de fragments de tailles déterminée. C'est la méthode du finger printing qui permet notamment de reconnaître un gène.

Si une mutation intervient à cet endroit précis de reconnaissance de l'enzyme (par exemple un GTTC au lieu de GATC), il n'y aura plus de coupure, le gène B par exemple ne donnera que 4 fragment dont un sera beaucoup plus long. C'est ce qu'on appelle un RFLP ("restriction fragment length polymorphism" ou en français polymorphisme de longueur des fragments de restriction)



Cette méthode permet de repérer des mutations. Dans une cellule saine on "découpe" un gène sain on obtiendra 4 fragments, si dans une cellule malade on "découpe" le même gène en obtenant cette fois ci 3 fragments on pourra se dire qu'il y a eu mutation, et que celle ci est peut être responsables de la maladie.



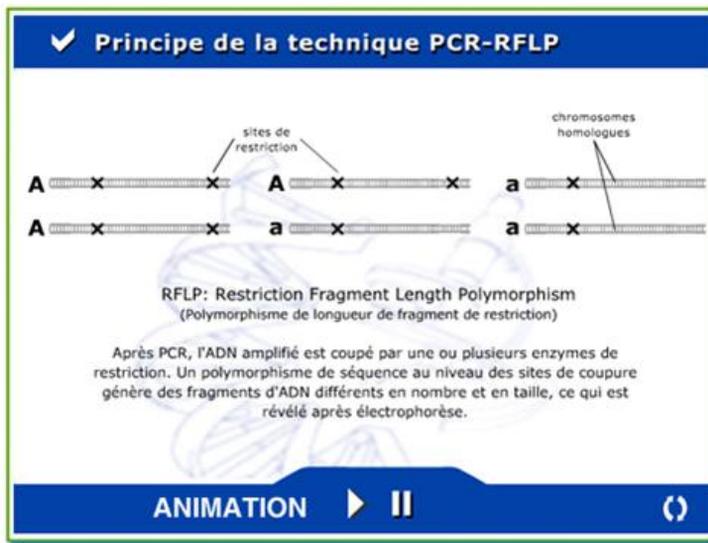


Comparons deux individus A et B. Leur ADN sera digéré séparément par des enzymes de restriction données (triangle et losange), puis hybridé par une sonde S.

La digestion avec l'enzyme –losange- donne des fragments de restriction identiques pour les deux individus. Les deux profils révélés après électrophorèse ne permettent pas de les distinguer. Avec ce couple enzyme - sonde S, aucun polymorphisme n'est mis en évidence.

En revanche, pour l'enzyme –triangle-, l'individu B présente une mutation au niveau d'un site de restriction, entraînant la perte de ce site. Ainsi, par digestion, l'individu A donne un fragment plus petit que celui de l'individu B. On révèle le polymorphisme entre les deux individus : un fragment rapide pour A et un plus lent pour B. Ce couple enzyme - sonde S révèle un polymorphisme. -Ce type de marqueur est codominant si le nombre d'enzymes utilisées est faible et s'il y a peu de sites d'hybridation de la sonde dans le génome. En faisant agir simultanément plusieurs enzymes de restriction, on étudie le polymorphisme à autant de sites particuliers qui se répètent tout au long de la molécule d'ADN. L'utilisation de la PCR permet d'amplifier une région définie du génome puis d'appliquer la technique RFLP au produit PCR. Cette méthode appelée PCR-RFLP permet d'obtenir facilement des marqueurs codominants et évite l'étape d'hybridation et l'utilisation de sonde radioactive. Le produit

PCR digéré par une (ou des) enzyme de restriction est simplement mis à migrer dans un gel d'agarose et le polymorphisme de la position et du nombre de bande est visualisé par une réaction colorée (Bromure d'éthyidium BET).



2.2.5. Les séquences répétées en tandem ou minisatellites (VNTR)

Il existe dans le génome de très nombreux organismes des séquences nucléotidiques répétées en tandem les unes à la suite des autres.

Définition

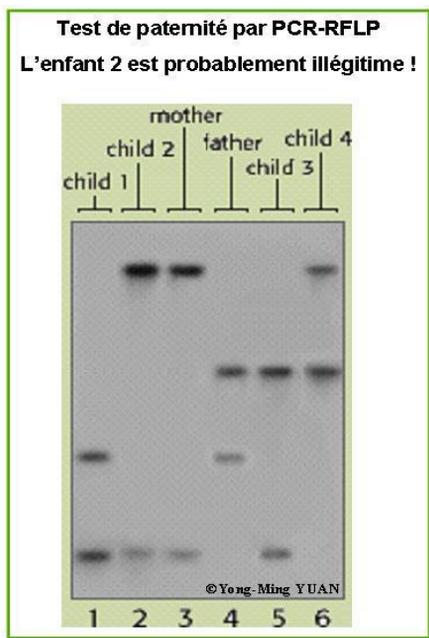
Les microsatellites sont des courtes séquences d'ADN des Chromosomes formées de la répétition d'un motif lui-même constitué de une à quatre bases. Les plus étudiés sont les répétitions (CA)_n. Très polymorphes, ce sont de bons marqueurs pour la cartographie génétique. Il existe dans le génome de très nombreux organismes des séquences nucléotidiques répétées en tandem les unes à la suite des autres. Le nombre de répétition est extrêmement variable entre individus d'où leur nom de VNTR (Variable Number of Tandem Repeat). Cette variation du nombre de répétitions est à l'origine d'un important polymorphisme dans les populations naturelles. On distingue 2 grands types de séquences répétées

a. les séquences minisatellites: des séquences de 15 à 25 pb) sont répétées un grand nombre de fois (1000 à 2000 fois), répétitions de séquences inférieures à 150 paires de bases. Ils sont télomériques, très polymorphes.

b. les microsatellites: des séquences de 1 à 5 paires de bases sont répétées un grand nombre de fois, répétitions longues de plusieurs Kb, il y a plus de 10 000 zones de répétition dans le génome. par exemple ATATATATATATATATAT soit $(AT)_n$ n variable entre individus
CAGACAGACAGACA soit $(CAGA)_n$

Les minisatellites peuvent être détectés par RFLP en utilisant des enzymes de restriction qui coupent un grand nombre de fois le génome mais jamais dans les minisatellites.

Un polymorphisme de longueur de fragment est alors révélé par l'existence d'un nombre de répétitions différent entre individus, ce qui produit des fragments de tailles différentes. Ces marqueurs sont donc multilocus et codominants. Les minisatellites sont révélés après PCR ce qui nécessite la mise au point d'amorces spécifiques. Cette étape est souvent longue et laborieuse mais permet d'obtenir des marqueurs monolocus et codominants.



9.3. Les RAPD (Random Amplified Polymorphism DNA)

Cette technique consiste à réaliser une amplification PCR avec des amorces d'environ 10 pb définies de façon aléatoire.

Cette technique a l'avantage d'être rapide avec peu de mise au point et révèle un polymorphisme important, mais ce marqueur est dominant et les conditions de PCR sont très sensibles.

Principes

Il s'agit d'une réaction de PCR dans laquelle les segments d'ADN amplifiés ne sont pas choisis par l'expérimentateur, mais amplifiés "au hasard".

L'amplification aléatoire d'ADN polymorphe, plus connue sous le nom de *RAPD* (pour *random amplified polymorphic DNA*) est une technique d'analyse de l'ADN utilisée en biologie moléculaire.

Une variabilité dans la séquence des sites entre individus sera détectée par un polymorphisme du nombre et de la longueur des fragments d'ADN amplifiés.

Une variante de la technique RAPD est d'utiliser des microsatellites comme amorces ce qui permet d'amplifier des régions comprises entre deux microsatellites (ISSR = Internal simple sequencerepeat).

Si les 2 sites d'hybridation sont suffisamment proches, l'amplification PCR est possible.

9.4. Les AFLP (Amplified Fragment Length Polymorphism)

Cette technique est une combinaison des RFLP et des RAPD. L'ADN est dans un premier temps digéré par des enzymes de restriction souvent (*EcoR1* et *Mse1*), puis des adaptateurs vont venir se fixer aux deux bouts des produits de digestion. Une amplification PCR est ensuite effectuée à l'aide d'amorces qui s'hybrident avec les adaptateurs mais comportent en plus quelques bases choisies au hasard afin d'amplifier de façon sélective uniquement certains fragments. Une seconde PCR peut ensuite être effectuée pour réaliser une

amplification plus sélective. Cette technique est très efficace pour révéler rapidement et facilement du polymorphisme.

9.5. Séquençage

Le séquençage complet de certaines portions du génome fournit le maximum de renseignements sur l'étendue de la variabilité interindividuelle, mais cette technique est peu compatible avec l'analyse d'un grand nombre d'individus d'une population.

Quoiqu'il en soit, même si une partie seulement du génome a été explorée chez les eucaryotes, le polymorphisme trouvé est suffisant pour entraîner une infinie diversité des génotypes individuels.

Chaque génotype individuel correspond à un certain arrangement des divers allèles présents à chaque locus dans la population. De façon analogue à un jeu de cartes, à partir d'un nombre limité de gènes polymorphes, la méiose redistribue les allèles par le jeu des recombinaisons intergéniques inter- ou intra-chromosomique, responsable de l'infinie diversité des génotypes individuels

10. Mesure de la diversité génétique

10.1. Fréquences alléliques et fréquences génotypiques

Lorsqu'une population est polymorphe pour un caractère donné, il est possible de calculer la fréquence des phénotypes observés (Masson et Hove, 2011, 2014).

Par exemple dans une population de N individus dont N_n ont le corps noir et N_b le corps blanc, les fréquences phénotypiques de la population pour le caractère couleur du corps sont les suivantes :

Fréquence du phénotype noir $f[n] = N_n/N$

Fréquence du phénotype blanc $f[b] = N_b/N$

Si ce caractère est gouverné par un gène à deux allèles A et a autosomaux, avec a récessif responsable de la couleur blanche, les génotypes AA et Aa correspondent au phénotype noir et le génotype aa au phénotype blanc.

Les fréquences phénotypiques permettent alors uniquement de connaître la fréquence du génotype *aa* puisque parmi les individus noirs, on ne peut pas distinguer les génotypes *AA* des génotypes *Aa*.

La fréquence des allèles *A* et *a* ne peut également pas être calculée.

Si la couleur du corps des individus présente non plus 2 mais 3 phénotypes (noir, jaune et blanc) gouvernés par un couple d'allèles **A1** et **A2** autosomaux et codominants, les trois génotypes possibles **A1A1**, **A1A2** et **A2A2** peuvent être distingués puisqu'ils correspondent à des phénotypes différents (respectivement noir, jaune et blanc).

La composition phénotypique de la population correspond alors à sa composition génotypique et si on appelle N_n , N_j et N_b les nombres d'individus présentant les phénotypes noir, jaune et blanc, on peut facilement calculer les fréquences génotypiques dans cette population :

$$f(A1A1) = N_n/N = D$$

$$f(A1A2) = N_j/N = H$$

$$f(A2A2) = N_b/N = R$$

Ainsi, pour un locus donné, une population est complètement décrite si l'on connaît la fréquence de chacune des catégories génétiques.

Dans le cas d'un système diallélique **A** et **a**, la structure d'une population d'effectif **N** est complètement connue si l'on connaît les effectifs

$$N_{AA} \text{ de } AA, N_{Aa} \text{ de } Aa \text{ et } N_{aa} \text{ de } aa \text{ avec } N = N_{AA} + N_{Aa} + N_{aa}$$

à partir desquels on calcule les fréquences relatives des trois génotypes.

Cette caractérisation génétique de la population n'est possible que si l'on sait lire les génotypes individuels, c'est à dire s'il y a codominance.

A partir des fréquences génotypiques, il est facile de calculer les fréquences alléliques dans la population, c'est à dire les fréquences des différents états alléliques du locus considéré.

Dans le cas d'un gène autosomal à deux allèles *A* et *a*, la fréquence de l'allèle *A* est le rapport du nombre d'allèles *A* au nombre total d'allèles à ce locus, soit $2N$ pour une population de N individus diploïdes:

- les N_{AA} individus **AA** sont porteurs de deux allèles **A**

- les **NAa** individus **Aa** d'un allèle **A** et d'un allèle **a**
- les **Naa** individus **aa** de deux allèles **a**.

Le nombre d'allèles **A** dans la population est donc

$$2N_{AA} + N_{Aa} .$$

Les fréquences **p** et **q** des allèles **A** et **a** sont alors les suivantes:

$$f(A) = p = (2 N_{AA} + N_{Aa})/2N \quad f(a) = q = (2N_{aa} + N_{Aa})/2N \quad \text{avec } p + q = 1$$

Autrement dit si **D** et **R** sont les fréquences des homozygotes **AA** et **aa**, **H** la fréquence des hétérozygotes **Aa**, les fréquences alléliques peuvent aussi être calculées à partir des fréquences génotypiques :

$$f(A) = p = D + H/2 \quad f(a) = q = R + H/2$$

Ces fréquences **p** et **q** représentent également une estimation de la probabilité qu'un gamète mâle ou femelle porte l'allèle **A** ou l'allèle **a**.

Il est important de noter que les fréquences alléliques comportent moins d'information que les fréquences génotypiques car on perd la manière dont les allèles sont associés 2 à 2 dans les génotypes individuels.

Exemple du groupe sanguin MN chez l'homme L'examen de 730 aborigènes australiens a donné les résultats suivants :

Groupe sanguin	Génotype	Nombre	Fréquence
[M]	MM	22	0.03
[MN]	MN	216	0.30
[N]	NN	492	0.67

$$f(M) = \frac{2 \times 22 + 216}{2 \times 730} = 0,03 + \frac{1}{2} \times 0,3 = 0,18 = p$$

$$f(N) = \frac{2 \times 492 + 216}{2 \times 730} = 0,67 + \frac{1}{2} \times 0,3 = 0,82 = q$$

10.2. Taux de polymorphisme et taux d'hétérozygotie

Pour quantifier la variabilité d'une population étudiée sur plusieurs gènes, différents paramètres peuvent être calculés.

a. Taux de polymorphisme P

C'est la proportion des gènes polymorphes parmi l'ensemble des gènes étudiés

$$P = \frac{\text{Nbre de gènes polymorphes}}{\text{Nbre total de gènes étudiés}}$$

Par exemple, si 30 loci enzymatiques ont été étudiés par la méthode d'électrophorèse avec 12 loci monomorphes et 18 polymorphes, le taux de polymorphisme P est $18/30 = 0,6$.

Ce paramètre présente cependant l'inconvénient de ne pas prendre en compte le nombre d'allèles rencontrés à chacun des loci polymorphes, ni leurs fréquences.

Il est évident par exemple qu'un locus possédant 10 allèles de fréquences voisines apporte plus de variation génétique à la population qu'un locus n'ayant que deux allèles dont un faiblement représenté.

b. Le taux d'hétérozygotie H

C'est la moyenne des fréquences des hétérozygotes observées à chacun des locus étudiés :

$$H_o = 1/N \sum H_i$$

N étant le nombre total de loci étudiés qu'ils soient monomorphes ou polymorphes.

H_i hétérozygotie au locus i

Le taux d'hétérozygotie fournit une bonne estimation de la variabilité génétique de la population, à condition toutefois que les individus de cette population se reproduisent au hasard.

Des modes de reproduction différents (homogamie, consanguinité, autogamie), conduisent à des situations où H_o ne donne plus une bonne estimation de la variabilité génétique.

Les modes de reproduction n'étant pas toujours connus, on calcule également un autre paramètre qui est l'hétérozygotie théorique attendue (H_t). Pour un locus A à k allèles A_1, A_2, \dots, A_k de fréquences f_1, f_2, \dots, f_k , l'hétérozygotie attendue est la suivante :

$$H_{tA} = 1 - (f_1^2 + f_2^2 + \dots + f_k^2) = 1 - S f^2$$

C'est est une estimation de la fréquence des hétérozygotes si les allèles sont associés au hasard pour former les génotypes.

L'hétérozygotie théorique globale est la moyenne des hétérozygoties attendues à chacun des loci étudiés:

$$H_t = 1/N \sum H_{ti}$$

N étant le nombre total de loci étudiés qu'ils soient monomorphes ou polymorphes H_{ti} hétérozygotie théorique au locus i

Il est alors possible de comparer la variabilité génétique des populations qui présentent des modes de reproduction différents. Diversité allélique Une autre mesure de la variabilité est la moyenne du nombre d'allèles par locus appelée diversité allélique :

$$A = \text{nbre total d'allèles} / \text{nbre de loci}$$

Pour prendre en compte la fréquence de ces allèles, on peut calculer le nombre d'allèles efficaces A_e . Pour un locus A à k allèles A_1, A_2, \dots, A_k de fréquences f_1, f_2, \dots, f_k , le nombre d'allèles efficaces est :

$$A_e = 1 / (f_1^2 + f_2^2 + \dots + f_k^2) = 1 / S f^2$$

On calcule alors la moyenne du nombre d'allèles efficaces par Locus

Chapitre III. Structure génétique de population idéale.

Population théorique idéale :

Définition :

Le devenir de la variabilité génétique d'une population au cours des générations (la transmission des différents allèles et leurs fréquences) est au premier abord très difficile à prévoir. Outre la difficulté à identifier une population, c'est-à-dire les limites du groupe d'individus sur lequel calculer les fréquences alléliques, de très nombreux facteurs peuvent modifier la fréquence de ces allèles (mutations, migrations, différence de survie ou fécondité entre individus). De plus, il faut considérer la transmission simultanée de très nombreux gènes polymorphes qui peuvent interagir entre eux et ne sont donc pas indépendants.

Une première étape pour contourner ces difficultés est d'aborder la transmission des caractères dans un cas simple, appelé population théorique idéale, qui se définit par les caractéristiques suivantes:

- 1) Population d'organismes diploïdes à reproduction sexuée et à générations non chevauchantes (aucun croisement entre individus de générations différentes).
- 2) Population d'effectif infini où les croisements sont entièrement aléatoires.
- 3) Population close génétiquement (absence de flux migratoires).
- 4) Tous les individus, quel que soit leur génotype, ont la même capacité à se reproduire et à engendrer une descendance viable (absence de sélection).
- 5) Absence de mutation et de distorsion de ségrégation méiotique (un individu Aa produira toujours 50% de gamètes A et 50% de gamètes a).

Parmi toutes ces caractéristiques, le croisement au hasard des individus, appelé système de reproduction panmictique, est l'hypothèse la plus importante.

Cette hypothèse suppose que les individus ne choisissent pas leur partenaire sexuel ni en fonction de leur génotype, ni en fonction de leur phénotype = panmixie et que la rencontre des gamètes se fait au hasard = pangamie.

2. L'équilibre de Hardy-Weinberg

Dans une population théorique idéale, les fréquences des allèles et des génotypes au cours des générations suivent une loi simple appelée loi de Hardy-Weinberg qui constitue le modèle de référence en génétique des populations. Cette loi doit son nom à Hardy, mathématicien anglais et Weinberg, médecin allemand, qui l'ont établie indépendamment en 1908. La loi de Hardy-Weinberg stipule que les fréquences alléliques et les fréquences génotypiques (c'est-à-dire la structure génétique de la population) reste stable de génération en génération. On dit alors que la population est à l'équilibre et il existe une relation simple entre les fréquences alléliques et les fréquences génotypiques (Masson et Hove, 2011, 2014).

3. Transmission d'un gène à 2 allèles

Supposons qu'une population soit polymorphe pour un caractère gouverné par un locus à deux allèles A et a et que les fréquences des génotypes AA, Aa et aa soient les mêmes dans les deux sexes, respectivement D, H et R (avec $D+H+R = 1$).

Les fréquences alléliques à la génération G_0 seront :

Pour l'allèle A $p_0 = (D_0 + H_0)/2$

Pour l'allèle a $q_0 = (R_0 + H_0)/2$ avec $p_0 + q_0 = 1$

Dans une population théorique idéale, ces fréquences seront également les fréquences des différentes catégories de gamètes (identiques dans les deux sexes) soit p_0 pour les gamètes qui portent l'allèle A et q_0 pour les gamètes qui portent l'allèle a (absence de mutation, de sélection et de distorsion méiotique).

La formation d'un nouvel individu de la génération suivante G_1 est alors le résultat de deux tirages au sort indépendants, l'un parmi les gamètes mâles, l'autre parmi les gamètes femelles (croisement au hasard = panmixie).

Les fréquences des différents génotypes de la génération suivante G_1 résultent alors de la répétition de ce simple tirage au sort qui donnera les fréquences génotypiques suivantes :

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

Dans une population théorique idéale, ces fréquences seront également celles des adultes reproducteurs de la génération G_1 (absence de sélection), pour lesquels les fréquences

alléliques seront pour A $p_1 = p_0^2 + p_0q_0 = p_0(p_0+q_0) = p_0 = p_0$ pour a $q_1 = q_0^2 + p_0q_0 = q_0(p_0+q_0) = q_0 = q_0$

Les fréquences alléliques n'ont donc pas changé, ce qui donnera à la génération suivante G2 les mêmes fréquences génotypiques qu'à la génération précédente soit p^2 , AA, $2pq$ Aa, et q^2 aa.

Le système est donc stable aussi bien en ce qui concerne les fréquences alléliques que les fréquences génotypiques.

On dit qu'on est à l'équilibre de Hardy-Weinberg dont la loi peut se s'énoncer de la façon suivante : Dans une population théorique idéale, les fréquences alléliques et les fréquences génotypiques restent stables de génération en génération. Les fréquences génotypiques sont déterminées à partir des fréquences alléliques par une relation simple qui correspond au développement du binôme $(p+q)^2$ dans le cas d'un locus à deux allèles A de fréquence p et a de fréquence q, soit les fréquences génotypiques p^2 AA, $2pq$ Aa et q^2 aa.

Fréquences génotypiques constantes

Fréquences génotypiques

Gamètes mâles ou femelles	A (p0)	a (q0)
A (p0) a (q0)	AA p0 ² Aa (p0q0)	Aa (p0q0) aa (q0 ²)

Fréquences géniques:

<p>pour l'allèle A</p> $p1 = p0^2 + p0q0$ $= p0 (p0+q0)$ $= p0 \times 1 \quad p1 = p0$	<p>Pour l'allèle a</p> $q1 = q0^2 + p0q0$ $= q0 (p0+q0)$ $= q0 \times 1$ $q1 = q0$
---	---

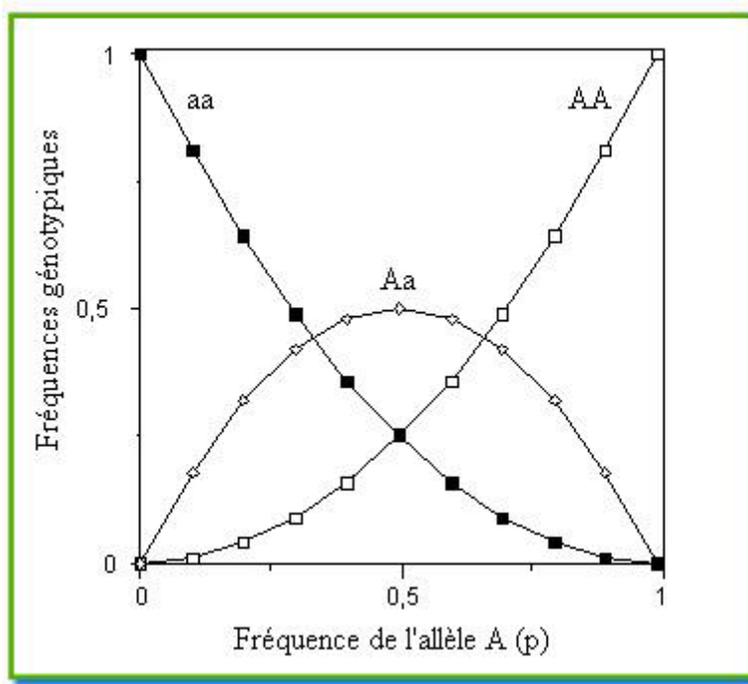
Conclusion:

D'une génération à la suivante: Fréquences alléliques constantes

Fréquences génotypiques constantes

p^2 AA ,, $2pq$ Aa ,, q^2 aa.

Cette relation entre fréquences alléliques et fréquences génotypiques peut être visualisée par la figure suivante:



On peut remarquer que les proportions mendéliennes 1/4, 1/2, 1/4 que l'on trouve lorsque l'on croise deux hétérozygotes est un cas particulier de la loi de Hardy-Weinberg lorsque $p = q = 0,5$. Chaque hétérozygote Aa a en effet comme fréquence allélique $f(A)=1/2$ et $f(a)=1/2$.

4. Systèmes multialléliques

La loi de Hardy-Weinberg s'applique également à des gènes qui existent sous plus de 2 états alléliques.

L'équilibre correspond alors à l'association aléatoire des différents allèles pour former les génotypes dont la fréquence reste stable de génération en génération.

Pour un locus à k allèles A1, A2, A3, ..., Ak, il y aura en théorie $(k(k+1))/2$ génotypes différents dans la population.

Si les fréquences de ces différents allèles sont respectivement $p_1, p_2, p_3, \dots, p_k$, les fréquences des différents génotypes seront données par le développement de $(p_1 + p_2 + p_3 + \dots + p_k)^2$

soit $p_1^2 A_1A_1$ $p_2^2 A_2A_2$ $p_3^2 A_3A_3$ $p_k^2 A_kA_k$

$2p_1p_2$	$2p_1p_3$	$2p_1p_k$	$2p_2p_3$	$2p_2p_k$	etc
A1A2	A1A3	A1Ak	A2A3	A2Ak	

Explication

Si les k allèles ont même fréquence ($p_1=p_2= p_3= \dots= p_k$), chaque allèle a une fréquence de $1/k$ dans la population.

Il y a donc : $k \times (1/k)^2 = 1/k$ homozygotes - $1- 1/k$ hétérozygotes

Par exemple, une population à l'équilibre peut comporter de plus de 90% d'hétérozygotes à un locus donné lorsqu'il existe plus de 10 allèles ayant les mêmes fréquences.

5. Application et utilisation du modèle de Hardy-Weinberg

5.1. Test de l'équilibre

Une question centrale est de savoir si la loi de Hardy-Weinberg établie pour une population théorique idéale s'applique également aux populations naturelles.

Cette loi s'appuie en effet sur un raisonnement probabiliste, ne s'applique en théorie qu'à des populations d'effectif infini, et suppose remplies toute une série de conditions qui ne sont jamais respectées dans la nature (absence de mutation, migration, sélection).

L'application de la loi de Hardy-Weinberg, peut être vérifiée pour des caractères codominants, pour lesquels le calcul des fréquences alléliques est possible. C'est le test de l'équilibre

Le principe du test est simple et peut être résumé en 3 étapes:

1- échantillonnage d'une population, dénombrement des effectifs génotypiques réels (possible grâce à la codominance) et calcul des fréquences alléliques réelles parmi les N individus échantillonnés soit $p= f(A)$ et $q = f(a)$

2- calcul des effectifs génotypiques attendus dans une population théorique idéale qui aurait le même effectif et les mêmes fréquences alléliques que la population étudiée soit

$AA = p^2 \times N$ $Aa = 2pq \times N$ $aa = q^2 \times N$

3- comparaison des effectifs observés et des effectifs attendus (comparaison des deux distributions) par un test statistique du Chi Deux (ou d'autres tests).

$$\chi^2 = \sum (\text{effectifs observés} - \text{effectifs théorique})^2 / \text{effectifs théoriques}$$

La somme est effectuée sur tous les génotypes et la valeur X2 est comparée à une valeur seuil, lue dans une table χ^2 , en fonction de 2 paramètres :

-un risque α choisi par l'utilisateur qui est en général 5% et un nombre de degrés de liberté ddl égale à la différence entre le nombre de génotypes et le nombre d'allèles du système génétique étudié.

- Si χ^2 calculé est inférieur à χ^2 seuil, H0 est acceptée et on conclut que la population suit la loi de Hardy-Weinberg donc est à l'équilibre.

- si X2 calculé est supérieur à X2 seuil, H0 est rejetée et on conclut que la population ne suit pas la loi de Hardy-Weinberg avec un risque $\alpha = 5\%$ de se tromper

Table de χ^2

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du degré de liberté (d.d.l.).

α d.d.l.	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,610	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,790	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	10,865	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	11,651	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	12,443	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	13,240	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797
22	14,041	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	14,848	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	15,659	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	16,473	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	17,292	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	18,114	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476
28	18,939	27,336	31,391	34,027	37,916	41,337	45,419	48,278	56,893
29	19,768	28,336	32,461	35,139	39,087	42,557	46,693	49,588	58,302
30	20,599	29,336	33,530	36,250	40,256	43,773	47,962	50,892	59,703

Exemple : avec d.d.l. = 3, pour $\chi^2 = 0,584$ la probabilité est $\alpha = 0,90$.

Exemple :

Chez l'homme, le groupe sanguin MN est déterminé par un gène à deux allèles codominants M et N, ce qui permet d'attribuer un génotype à chaque individu échantillonné, puis d'estimer les fréquences alléliques dans la population.

Une étude portant sur 730 aborigènes australiens a donné les résultats suivants :

22 MM, 216 MN, 492 NN

1- Calcul des fréquences p et q des allèles M et N :

$$p = (22 + 1/2 \times 216) / 730 = 0.178 \text{ pour l'allèle M.}$$

$$q = 492 + 1/2 \times 216) / 730 = 0.822 \text{ pour l'allèle N.}$$

2- Calcul des effectifs théoriques attendus des différentes catégories génotypiques :

$$MM = p^2 \times 730 = (0,178)^2 \times 730 = 23,1$$

$$MN = 2pq \times 730 = (2 \times 0.178 \times 0.822) \times 730 = 213,6$$

$$NN = q^2 \times 730 = (0.822)^2 \times 730 = 493,22$$

3- Test du Chi deux

$$\chi^2 = (22-23,1)^2/23,1 + (216-213,6)^2/213,6 + (492-493,2)^2/493 = 0,083$$

La valeur seuil pour 3-2 = 1 degré de liberté et un risque de 5% est 3.84. La valeur de la statistique χ^2 étant très inférieure à la valeur seuil, on conclut qu'il n'y a pas de différence significative entre la distribution observée et la distribution théorique.

On admet donc que la population d'aborigènes australiens est à l'équilibre de HardyWeinberg. Le fait qu'une population soit considérée à l'équilibre de Hardy-Weinberg après un test statistique n'implique pas que toutes les conditions d'application de cette loi soient respectées (effectif infini, absence de mutation, absence de sélection, etc).

L'hypothèse la plus importante qui doit être respectée est la panmixie. Un équilibre génétique instantané apparent peut donc être observé même si la population est soumise à une forte sélection. On en conclut que le fonctionnement génétique global est "proche" du fonctionnement théorique et ce n'est qu'en prenant en compte la dimension temporelle que l'on peut apprécier l'état génétique d'une population et prévoir son évolution.

En conclusion, dans la plupart des cas, le modèle de Hardy-Weinberg constitue un bon descripteur de la structure génétique des populations naturelles car l'hypothèse de panmixie est souvent respectée alors que les effets des mutations, migration, sélection ne sont pas assez forts pour faire diverger les fréquences Génotypiques des proportions du modèle de Hardy-Weinberg. Cette loi peut alors être utilisée pour faire des prévisions notamment dans le domaine médical. Cet équilibre de Hardy-Weinberg ne s'applique pas obligatoirement à tous les gènes d'une même population et peut ou non être rejeté en fonction du système génétique considéré.

Exercice d'application

L'analyse du polymorphisme de l'enzyme estérase 1 dans un échantillon de 300 personnes a révélé l'existence de 3 niveaux de migration (E1, E2 et E3) à l'origine de 6 génotypes dont les effectifs sont les suivants :

$$E1E1 = 72 \quad E2E2 = 24 \quad E3E3 = 15 \quad E1E2 = 99 \quad E1E3 = 57 \quad E2E3 = 33$$

Solution

Fréquences alléliques : $E1 = 0,5 = p$; $E2 = 0,3 = q$; $E3 = 0,2 = r$

Test de l'équilibre par un test Chi 2 (X2) : Hypothèse testée : "la population étudiée suit la loi de Hardy Weinberg

Effectifs observés Effectifs théoriques

$$15 r^2 \times N = 12$$

$$72 p^2 \times N = 75$$

$$24 q^2 \times N = 27$$

$$99 2pq \times N = 90$$

$$57 2pr \times N = 60$$

$$33 2qr \times N = 36$$

$$\chi^2 = \sum \frac{(\text{effectifs observés} - \text{effectifs théoriques})^2}{\text{effectifs théoriques}}$$

d'où $X^2 = 2,5$ X^2 seuil (5% ; 3ddl) = 7,81

3 ddl car 6 classes génotypiques - 3 (2 fréquences alléliques et l'effectif total)

Donc on accepte H_0 : la population suit la loi de Hardy Weinberg.

5.2. Estimation des fréquences alléliques

Lorsque la variabilité d'un caractère est dûe à un gène à 2 allèles avec une forme allélique totalement dominante sur l'autre (A dominant sur a), seuls 2 phénotypes peuvent être distingués dans la population :

- le phénotype [A] correspondant aux génotypes AA et Aa
- le phénotype [a] correspondant au génotype aa

C'est le cas par exemple de nombreuses maladies génétiques chez l'homme qui sont dues à un allèle récessif (mucoviscidose, phénylcétonurie).

Contrairement à un système codominant, il n'est pas possible de calculer les fréquences alléliques dans la population car les proportions respectives des génotypes AA et Aa ne sont pas connues

Le modèle de Hardy-Weinberg va permettre de donner une estimation de ces fréquences à partir de la fréquence du phénotype homozygote récessif qui est égale à q^2 si la population est conforme au modèle

La résolution de cette équation à une inconnue permet d'estimer la fréquence q de l'allèle récessif dans la population en calculant la racine carrée de la fréquence des homozygotes récessifs.

La fréquence p de l'allèle dominant est obtenue par différence à 1.

A partir de cette estimation des fréquences alléliques, et toujours sous l'hypothèse de conformité au modèle de Hardy-Weinberg, on obtient une estimation de la fréquence des homozygotes AA et des hétérozygotes Aa parmi les individus de phénotype [A] c'est-à-dire la probabilité qu'un individu de phénotype [A] soit homo- ou hétérozygote :

-fréquence des homozygotes parmi les individus

$$[A] = \frac{p^2}{(p^2+2pq)} \text{ ou } \frac{p^2}{(1-q^2)}$$

-fréquence des hétérozygotes parmi les individus

$$[A] = 2pq/(1-q^2).$$

Exemple :

Chez l'homme, une étude portant sur le système Rhésus a recensé 14% d'individus Rhésus négatif. Sachant que l'allèle Rh⁺ est dominant sur l'allèle Rh⁻, l'estimation de la fréquence de l'allèle Rh⁻ est $q = 0.37$ (racine carrée de 0.14) sous l'hypothèse que la population suit la loi de Hardy-Weinberg. On peut en déduire la fréquence des individus Rh⁺Rh⁺ et Rh⁺Rh⁻ parmi les individus Rhésus positif, respectivement $p^2/(p^2+2pq) = 0.45$ et $2pq/(p^2+2pq) = 0.55$

5.3. Diagnostic et conseil génétique

La loi de Hardy-Weinberg permet de faire des prévisions sur le génotype d'un individu lorsque l'on connaît la population dont il est issu. Ce calcul est utilisé en génétique humaine pour calculer la probabilité qu'un individu soit atteint d'une anomalie génétique.

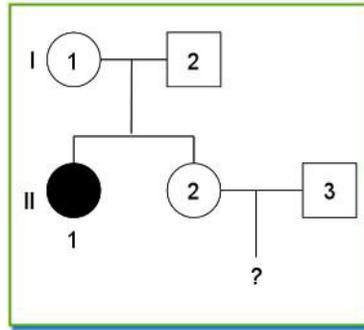
C'est le conseil génétique.

Le calcul du risque d'apparition d'une anomalie génétique chez un individu donné dépend de plusieurs paramètres :

- du déterminisme du caractère et des relations de dominance
- entre les allèles
- de la fréquence du gène responsable de la maladie dans la population
- de la généalogie de l'individu notamment des phénotypes
- des ascendants, descendants et collatéraux

Pour une maladie autosomique récessive déterminée par un allèle a de fréquence q, la probabilité qu'un individu dont on ne connaît ni la généalogie ni le phénotype soit atteint par cette maladie correspond à la fréquence de ce phénotype dans la population soit q^2 .

Le diagnostic s'affine considérablement lorsque l'on dispose de plus d'informations par exemple dans la généalogie suivante où il s'agit de calculer la probabilité que le couple formé des individus sains II₂ et II₃ donne naissance à un enfant atteint de l'anomalie ce qui nécessite que les deux parents, tous les deux sains, soient hétérozygotes



Pour l'individu II3, aucune information n'est disponible, excepté son propre phénotype.

La probabilité que cet individu soit porteur de l'allèle a est $2pq/(p^2+2pq)$ c'est à dire la fréquence des individus hétérozygotes parmi les sains dans la population.

L'individu II2 ayant une soeur atteinte, leurs parents sont obligatoirement tous les deux hétérozygotes et la probabilité est alors de $2/3$ pour que II2 soit hétérozygote sachant qu'il est lui-même non atteint. ($2/3$ et non $1/2$ car le phénotype de l'individu II2 est connu). La probabilité pour que le couple II2 x II3 donne naissance à un enfant atteint de l'anomalie est alors la suivante:

$$2pq/(p^2 + 2pq) \times 2/3 \times 1/4$$

Soit : proba (père Aa) x proba (mère Aa) x proba (enfant aa sachant que les parents Aa).

6. Transmission des gènes liés au sexe

L'étude des gènes liés au sexe n'est pas anecdotique car chez certains organismes, ils constituent une grande partie de l'ADN codant. C'est par exemple le cas chez la drosophile dont plus d'un tiers des gènes sont porté par le chromosome X.

Pour les caractères portés par les chromosomes sexuels, les deux sexes ont des constitutions génétiques différentes et il faut distinguer : - Le sexe homogamétique qui porte les deux mêmes chromosomes sexuels (femme XX chez les mammifères, certains insectes dont la drosophile ; mâle ZZ chez certains crustacés et papillons). Ce sexe est donc diploïde pour ce chromosome. Le sexe hétérogamétique, qui porte deux chromosomes sexuels différents (ou un seul) donc haploïde ou hémizygote (mâles XY chez les mammifères, femelles WZ chez les crustacés et papillons). Ce sexe est haploïde (hémizygoïde) pour ce chromosome. Les

deux sexes ont donc des contributions génétiques différentes et s'ils sont en fréquence égale (sex-ratio équilibrée), le sexe homogamétique détient pour les gènes concernés les 2/3 du pool génétique de la population, et le sexe hétérogamétique 1/3 seulement.

a/Lorsque les fréquences alléliques sont les mêmes chez les mâles et les femelles, $p_m = p_f = p$ et $q_m = q_f = q$, la loi de Hardy-Weinberg s'applique au sexe homogamétique et les fréquences génotypiques dans le sexe hétérogamétique sont directement déduites des fréquences alléliques

Chez l'homme, pour un gène porté par X et présentant 2 allèles A et a de fréquences p et q, les fréquences génotypiques dans chacun des deux sexes pour une population à l'équilibre seront :

	Femme (XX)	Homme (XY)
Fréquences génotypiques à l'équilibre	$X^A X^A = p^2$	$X^A Y = p$
	$X^A X^a = 2pq$	$X^a Y = q$
	$X^a X^a = q^2$	

Cette différence de structure génétique des sous-populations mâle et femelle a un effet spectaculaire dans le cas où un allèle est récessif. Le phénotype récessif est beaucoup plus fréquent dans le sexe hétérogamétique que dans le sexe homogamétique, où il peut être exceptionnel si la fréquence de l'allèle est faible.

C'est l'exemple classique du daltonisme chez l'homme qui est dû à une mutation récessive sur un gène porté par X.

Maladie récessive liée au chromosome X					
	Garçon		Filles		
Phénotype	Sain	Malade	Saine	Saine	(Malade)
Génotype	A	A (l)	AA	Aa(H)	aa
Fréquence	P	q	P ²	2Pq	2Pq

On connaît I, l'incidence de la maladie dans la population masculine = q

La fréquence des filles hétérozygotes, H est donc égale à :

$$2 \times q \times (1 - q)$$

Si q est très petit : H = 2I (il y a deux fois plus de femmes hétérozygotes que de garçon atteints).

En Europe:

- la fréquence de l'allèle récessif est de l'ordre de (q = 0,04).
 - l'anomalie est fréquente chez les hommes (q = 4%) mais très rare chez les femmes (q² = 0.16%).
 - la fréquence des hétérozygotes est élevée chez les femmes (de l'ordre de (2pq = 8%).
- Rappelons que ces femmes ont une chance sur deux de transmettre l'anomalie à chacun de leurs descendants mâles.

b/ Lorsque les fréquences alléliques sont différentes chez les mâles et les femelles, $p_m \neq p_f$ et $q_m \neq q_f$,

la contribution différentielle des deux sexes à la descendance maintient cette différence pendant plusieurs générations et l'égalité des fréquences alléliques n'est obtenue que progressivement (contrairement aux caractères autosomaux):

Exemple: Daltonisme chez l'homme qui est dû à une mutation récessive sur un gène porté par X.

- chaque mâle XY reçoit un chromosome X de sa mère, donc les fréquences alléliques chez les mâles à la génération t correspondent aux fréquences alléliques chez les femelles de la génération précédente t-1:

$$f(A) : p_{mt} = p_{ft-1}$$

$$f(a) : q_{mt} = q_{ft-1}$$

- chaque femelle XX reçoit un chromosome X de sa mère et un chromosome X de son père donc les fréquences alléliques chez les femelles à la génération t correspondent à la moyenne des fréquences alléliques des deux sexes de la génération précédente t-1 :

$$f(A): p_{ft} = (p_{mt-1} + p_{ft-1})/2$$

$$f(a): q_{ft} = (q_{mt-1} + q_{ft-1})/2$$

Dans l'ensemble de la population, les fréquences alléliques globales des allèles A et a sont les moyennes de la fréquence de ces allèles dans les deux sexes pondérées par leurs contributions relatives soit les coefficients 1/3 et 2/3 lorsqu'il y a autant de mâles que de femelles :

population :

$$f(A): p = 2/3 p_f + 1/3 p_m.$$

$$f(a): q = 2/3 q_f + 1/3 q_m.$$

Si la proportion des sexes est inégale, la pondération tient compte des effectifs N_m des mâles et N_f des femelles :

$$\text{population : } f(A): p = (p_m.N_m + 2.p_f.N_f) / (N_m + 2N_f)$$

Au cours des générations, l'évolution des fréquences alléliques dans chacun des deux sexes fluctue autour de cette valeur d'équilibre avec une différence qui s'inverse et diminue de moitié à chaque génération. Cette fluctuation conduit à l'égalité des fréquences alléliques dans les deux sexes après plusieurs générations de croisements panmictiques. Fluctuation

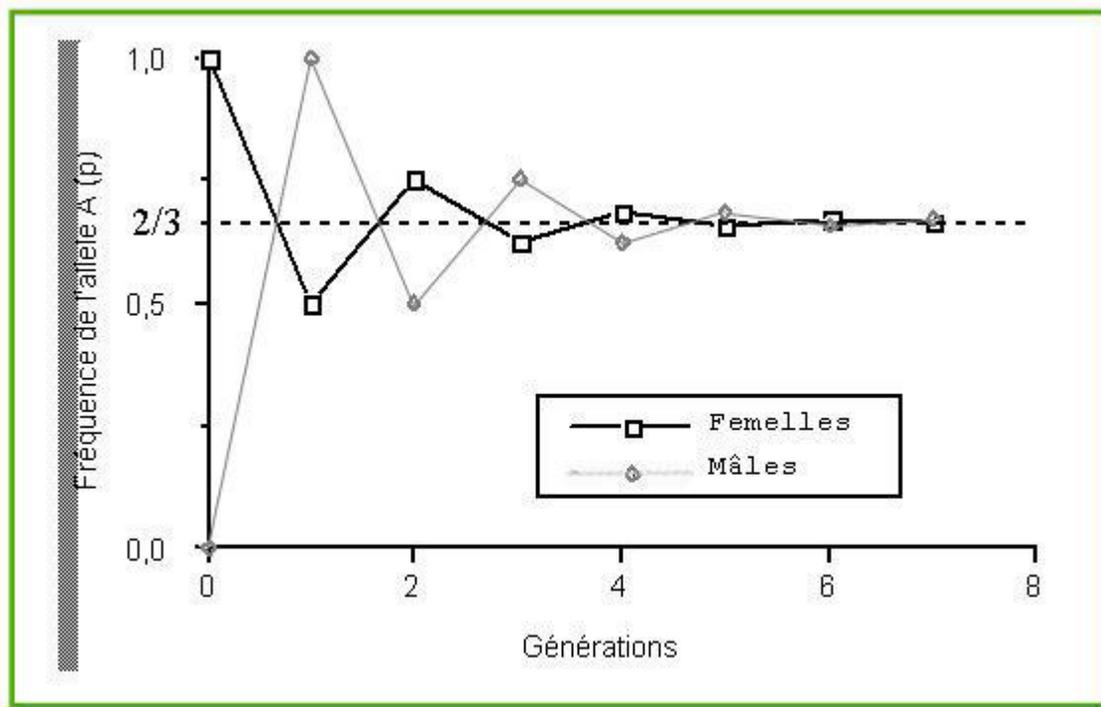
Si à la génération G_0 tous les mâles sont X_aY ($p_{m0} = 0$) et toutes les femelles sont homozygotes $XAXA$ ($p_{f0} = 1$), la fréquence de l'allèle A dans la population est $p_0 = 2/3$.

A la première génération, la fréquence allélique chez les mâles est $p_{m1} = p_{f0} = 1$ et chez les femelles : $p_{f1} = (p_{m0} + p_{f0})/2 = 0,5$.

A la génération suivante, on a $p_{m2} = 0,5$ et $p_{f2} = 0,75$, etc

Il se produit donc des oscillations des fréquences alléliques dans les deux sexes, en opposition de phase, qui s'amortissent progressivement comme le montre la figure ci-dessous.

Les fréquences alléliques dans l'ensemble de la population restent invariables (ici $2/3$ pour A) et c'est vers cette valeur d'équilibre que convergent les fréquences alléliques des deux sexes. A l'équilibre, $q_m = q_f = 2/3$. La structure génétique est alors la suivante : - chez les mâles : p individus XAY ($2/3$) et q individus XaY (ici $1/3$); - chez les femelles, p^2 individus $XAXA$ ($4/9$), q^2 individus $XaXa$ ($1/9$), $2pq$ individus $XAXa$ ($4/9$).



G0 tous les mâles sont XaY ($p_m0 = 0$)

G0 toutes les femelles sont homozygotes $XAXA$ ($p_f0 = 1$)

La fréquence de l'allèle A dans la population est $p_0 = 2/3$.

G1 La fréquence allélique chez les mâles est $p_m1 = p_f0 = 1$

G1 les femelles : $p_f1 = (p_m0 + p_f0)/2 = 0,5$.

G2 A la génération suivante, on a $p_m2 = 0,5$ et $p_f2 = 0,75$, etc

7. Transmission de plusieurs gènes et déséquilibre gamétique

Dans les populations naturelles, un très grand nombre de caractères sont génétiquement variables. Il est donc nécessaire d'étudier la transmission simultanée de plusieurs gènes polymorphes. C'est en effet l'ensemble du pool génétique qui se transmet d'une génération à l'autre, c'est à dire pour une population de N individus diploïdes l'ensemble des $2N$ exemplaires de chacun des gènes dont certains sont présents sous plusieurs formes alléliques. L'étude de la transmission simultanée des caractères polymorphes se complique rapidement car elle concerne un grand nombre de gènes pouvant ou non être portés par les mêmes chromosomes (gènes physiquement liés) ou interagissant entre eux par des relations d'épistasie. Seule l'étude du cas de deux loci alléliques peut être abordée de façon simple



L'analyse de la composition génétique d'une population à plusieurs loci permet de mieux comprendre son fonctionnement à la fois au niveau des effectifs, des flux migratoires et des pressions de sélections qui s'exercent sur les caractères concernés.

D'un point de vu appliqué, l'étude des l'association alléliques à des loci étroitement liés

permet de cartographier les gènes et de diagnostiquer des maladies génétiques à partir de marqueurs moléculaires non impliqués dans la maladie.

7.1. Equilibre gamétique à 2 loci

Dans une population théorique idéale, les allèles des différents loci sont associés au hasard donc sont statistiquement indépendants.

Cette indépendance statistique résulte du processus de recombinaison qui est maximum pour des gènes indépendants et d'autant plus faible que les gènes sont proches les uns des autres sur le même chromosome.

La notion d'équilibre intègre donc non seulement l'association au hasard des allèles d'un même gène pour former les génotypes (loi de Hardy-Weinberg) mais également l'association au hasard des allèles de différents gènes.

Cet équilibre à plusieurs loci peut être facilement formalisé dans le cas de 2 gènes à 2 allèles avec :

- Au locus A les allèles A et a de fréquences p et q.
- Au locus B les allèles B et b de fréquences r et s.

7.2. L'équilibre gamétique

Correspond à l'association au hasard de tous les allèles au niveau des gamètes, c'est-à-dire des haplotypes. Les fréquences des 4 catégories de gamètes correspondent alors au produit des fréquences des allèles qui forment ces gamètes soit :

Fréquence des gamètes AB = pr

Fréquence des gamètes Ab = ps

Fréquence des gamètes aB = qr

Fréquence des gamètes ab = qs On distingue souvent les associations de type couplage (ou cis), correspondant aux associations AB et ab, des associations de type répulsion (ou trans) Ab et aB. Cette distinction n'est cependant possible que lorsque des relations de dominance existent entre les allèles d'un même locus.

A l'équilibre, le produit des associations de type couplage est égal au produit des associations de type répulsion :

$$f(AB) \times f(ab) = f(Ab) \times f(aB)$$

$$pr \times qs = ps \times qr$$

$$pqrs = pqrs$$

Au niveau des génotypes, cet équilibre se traduit à la fois par une association au hasard des allèles à un même locus et une association au hasard des génotypes des différents loci. Les fréquences génotypiques sont alors :

$$\text{Au locus A } p^2 \text{ AA } 2pq \text{ Aa } q^2 \text{ aa}$$

$$\text{Au locus B } r^2 \text{ BB } 2rs \text{ Bb } s^2 \text{ bb}$$

Les fréquences des 9 génotypes possibles correspondent alors aux produits des fréquences génotypiques à chaque locus (indépendance statistique entre génotypes des deux loci) :

$$AABB = p^2 r^2 \quad AaBB = 2pq r^2 \quad aaBB = q^2 r^2$$

$$AABb = 2 p^2 rs \quad AaBb = 4pqrs \quad aaBb = 2 q^2 rs$$

$$AAbb = p^2 s^2 \quad Aabb = 2pq s^2 \quad aabb = q^2 s^2$$

Ces fréquences correspondent également à la fusion au hasard des gamètes qui forment ces génotypes soit par exemple la fusion au hasard de 2 gamètes AB chacun de fréquence pr pour le génotype donc $(pr)^2 = p^2 r^2$

Chapitre IV. Cause de la variabilité de la structure génétique des populations naturelles.

1. Définition de la variation de la fréquence allélique dans les populations

Dans la nature, les populations s'écartent toujours des conditions théoriques idéales. Les effectifs sont limités, fluctuent au cours du temps et les différents génotypes se distinguent par des capacités reproductives différentes.

De plus, les populations ne sont jamais génétiquement closes, car elles subissent l'apport de nouveaux allèles soit par migration soit par mutation.

Ces facteurs ont pour conséquence une variation de la fréquence des allèles au cours des générations successives.

Ces changements, appelés changements micro évolutifs ou microévolution, modifient la structure génétique de la population qui peut ou non évoluer vers une meilleure adaptation à son environnement.

L'analyse des conséquences génétiques de ces différents facteurs nécessite l'emploi de modèles mathématiques ou de simulations qui permettent de déterminer comment et sous quels facteurs les fréquences alléliques varient au cours du temps.

On recherche généralement les conditions pour lesquelles un équilibre peut être atteint, c'est-à-dire où les fréquences alléliques ne varient plus, bien que les populations restent sous l'action de ces différents facteurs (mutations, migrations, sélections).

Ce sont donc les conditions qui permettent un maintien de la variabilité génétique qui sont étudiées. L'étude de l'effet de ces différents facteurs sur la variation des fréquences alléliques et la recherche de cet équilibre se fait généralement de la façon suivante :

a. Mesure de la variation des fréquences alléliques entre 2 générations successives

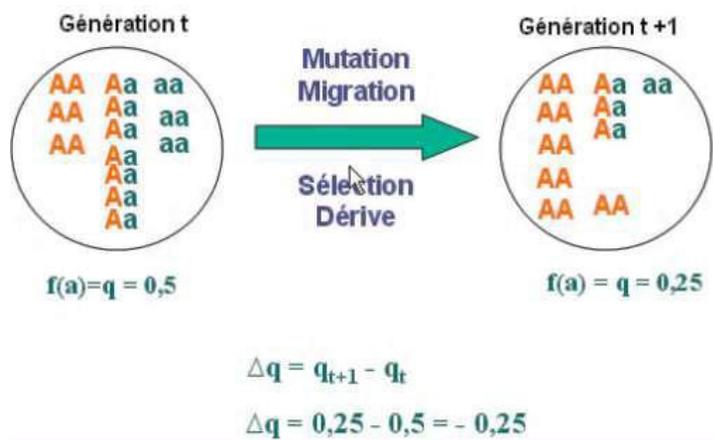
Pour l'allèle A $\Delta p = p_{t+1} - p_t$

pour l'allèle a $\Delta q = q_{t+1} - q_t$ et p_t et p_{t+1} étant respectivement la fréquence de l'allèle A aux générations t et t+1.

b. La recherche d'un équilibre, c'est-à-dire les conditions qui permettent d'annuler Δq ou Δp

Pour l'allèle A $\Delta p = 0$

Pour l'allèle a $\Delta q = 0$



2. Les mutations

Les mutations sont à l'origine des variations génétiques observées dans les populations naturelles. Elles font apparaître de nouveaux allèles ou de nouvelles structures chromosomiques. Les mutations sont dans la plupart des cas récurrentes, car elles se produisent à chaque génération, mais leur fréquence est généralement faible de l'ordre de 10^{-7} (Jean-Pierre, 2008). La fréquence avec laquelle les allèles d'un locus donné subissent une mutation à une génération donnée est appelée taux de mutation (Exemple tableau IV)

Tableaux IV : Les taux de mutation pour différentes espèces et

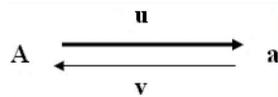
Espèce	Caractères Maladies	Taux de mutation	Mutation réverse
Homme	Achondroplasie	$1 \cdot 10^{-5}$	-
	Achondroplasie	$7 \cdot 10^{-6}$	-
	Hémophilie A	$5.7 \cdot 10^{-5}$	-
	Hémophilie B	$3 \cdot 10^{-6}$	-
	Myopathie de Duchène	$10.5 \cdot 10^{-5}$	-
	Myopathie de Duchène	$4.6 \cdot 10^{-5}$	-
Sourie	Non agouti	$4.5 \cdot 10^{-5}$	$4.2 \cdot 10^{-6}$
	Albinos	$3.3 \cdot 10^{-5}$	-

Lorsque les différents états alléliques produits par mutation ne sont pas soumis à sélection (allèles neutres), l'évolution de la fréquence de ces allèles dépendra uniquement de leurs taux de mutation respectifs.

Lorsqu'un gène subit de manière récurrente une mutation le faisant passer de l'état allélique A vers l'état allélique a sans que la mutation inverse soit possible (mutation de a vers A), l'allèle a va progressivement remplacer l'allèle A dans la population.

Cette évolution sera cependant extrêmement lente, voire indétectable, compte tenu des taux faibles de mutation. L'allèle a sera par contre rapidement éliminé si il est à l'origine d'un désavantage pour l'individu qui le porte. Lorsque les mutations sont récurrentes et que les mutations réverses sont possibles, l'évolution des fréquences alléliques sera fonction des valeurs respective des différents taux de mutations.

Dans le cas d'un gène à deux allèles A et a, les fréquences respectives sont p et q, et les taux de mutation sont respectivement u pour le passage de A vers a et v pour le passage de a vers A.



Si p_0 et q_0 sont les fréquences alléliques de A et a à la génération G_0 , les fréquences p_1 et q_1 de ces allèles à la génération G_1 seront les suivantes :

$$f(A) p_1 = p_0 - u p_0 + v q_0$$

$$f(a) q_1 = q_0 - v q_0 + u p_0$$

La variation Δq de la fréquence de l'allèle a entre les générations G_0 et G_1 est :

$$\Delta q = q_1 - q_0$$

$$\Delta q = -v q_0 + u p_0$$

A l'équilibre, les fréquences des allèles A et a ne varieront plus entre deux générations successives ($p_{n+1} = p_n$ et $q_{n+1} = q_n$). Ces valeurs d'équilibre, appelées p_e pour A et q_e pour a, seront obtenues lorsque $\Delta q = 0$ c'est à dire :

$$\Delta q = -v q_e + u p_e = 0 \quad \Delta q = -v q_e + u (1 - q_e) = 0 \quad \Delta q = -v q_e + u - u q_e = 0 \quad \Delta q = u - q_e (u + v) = 0 \quad q_e = u / (u + v)$$

Et de la même façon, pour l'allèle A :

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

La valeur des fréquences alléliques à l'équilibre dépend donc uniquement des taux de mutation. Si $u = 10^{-6}$ et $v = 10^{-7}$ (taux de mutation réalistes chez les eucaryotes), la fréquence

de l'allèle A à l'équilibre sera $p_e = 0.09$ donc $q_e = 0.91$.

Ainsi, si seul l'allèle A est présent à la première génération ($p_0 = 1$), la fréquence de cet allèle diminuera au cours des générations pour atteindre la fréquence de 0.09 à l'équilibre.

Cependant, cette évolution ne se fera que très lentement puisqu'il faut à peu près 1000 générations pour faire passer p de 1 à 0.99.

Ainsi, le processus de mutation seul n'a pas d'effet important sur la structure génétique des populations car la variation des fréquences alléliques qu'il induit est extrêmement faible voire indétectable à l'échelle de quelques dizaines voire quelques centaines de générations. On peut démontrer que l'évolution des fréquences alléliques au cours du temps est : $p_{n+1} = p_e + (p_0 - p_e)(1 - u - v)^n$

Les mutations sont cependant d'une importance capitale puisqu'elles génèrent des formes alléliques nouvelles dont les fréquences vont pouvoir varier rapidement sous l'action d'autres processus en particulier la sélection naturelle.

En résumé, les mutations font apparaître à chaque génération des allèles nouveaux, et en dernier ressort, ce sont elles les principales responsables de l'existence du polymorphisme. De ce point de vue, leur rôle est capital car elles déterminent le potentiel adaptatif des populations donc leur maintien à long terme. Par contre, les taux de mutation sont très faibles et la répétition des mutations à chaque génération ne modifie que très lentement les fréquences de ces allèles. A une échelle de temps de quelques dizaines à quelques centaines de générations, l'effet des mutations sur l'évolution des fréquences alléliques est donc négligeable

3. Les migrations Dans la nature, les populations d'une même espèce ne sont pas génétiquement isolées. Il se produit à chaque génération des échanges d'adultes, mais aussi d'embryons (graines chez les plantes), de pollen, ou de gamètes (gamètes des animaux aquatiques) qui sont autant d'échanges de gènes, appelés flux géniques.

Ces flux géniques sont généralement d'autant plus importants que les populations sont proches géographiquement. C'est le cas lorsqu'une population est subdivisée en sous-populations à cause d'une discontinuité de l'habitat ou des ressources. Ces échanges de gènes provoqués par la dispersion des individus entre sous populations limitent alors leur divergence génétique (Jean-Pierre, 2008).

Les conséquences génétiques de ces phénomènes migratoires dépendent de l'organisation spatiale des populations, de leur densité en individus et des nombreux facteurs

environnementaux responsables de leur dispersion, l'ensemble devenant rapidement complexe à analyser (Figure 7)

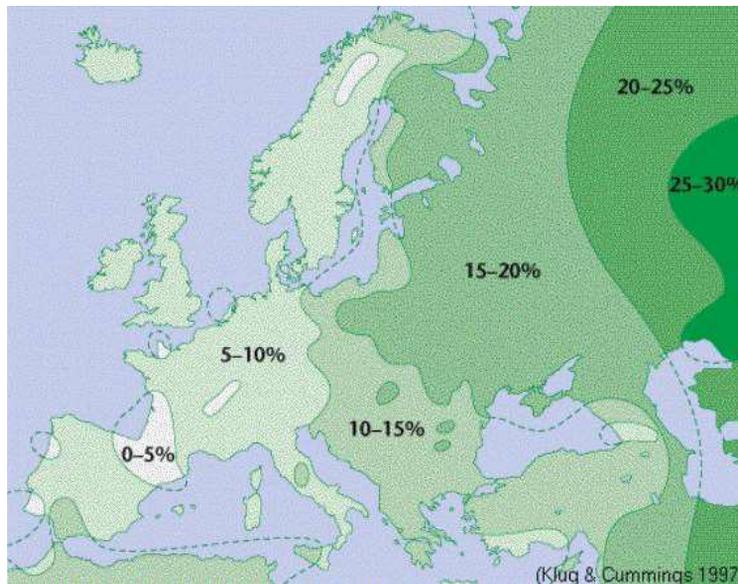


Figure: **Migration chez l'homme. Diffusion de l'allèle B du système ABO en Europe de l'Ouest suite au flux migratoire venu de l'Est.**

3.1. Le modèle insulaire

Le modèle de migration le plus simple, appelé modèle insulaire (Figure 8), est celui où les échanges d'individus s'effectuent dans un seul sens avec un flux de gènes unidirectionnel entre une population 1 de grande taille (le continent) et une population 2 de petit effectif (l'île). C'est le cas d'une île proche d'un continent, d'une étendue d'eau située en aval d'un lac, ou d'une minorité isolée du reste de la population sur une base ethnique ou socio-culturelle.

Dans tous ces cas, la population 2 (l'île) est constituée de résidents dont le taux d'émigration peut être négligé alors que la population 1 (le continent) envoie à chaque génération une proportion donnée de migrants. Si à chaque génération, la population 2 d'effectif N reçoit n individus migrants de la population 1, le flux migratoire m correspond à la proportion de migrants qui arrivent dans la population résidente : $m = n / (N + n)$

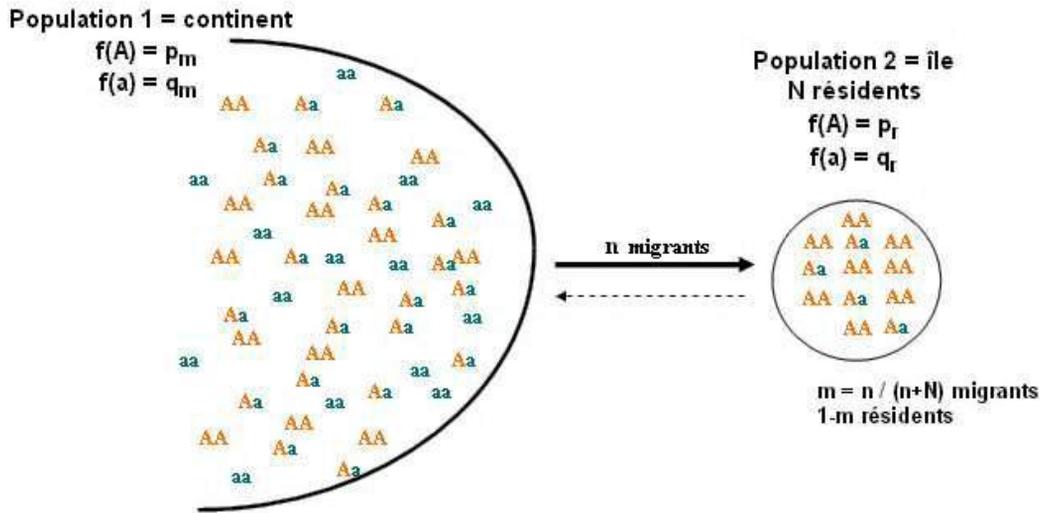


Figure : Modèle insulaire (île-continent).

A chaque génération, il y a une proportion m d'individus migrants et une proportion $1-m$ d'individus résidents.

Si à une génération G_0 donnée les deux populations ont des fréquences différentes pour les allèles d'un même gène, par exemple p_m pour un allèle A dans la population de migrants et p_0 pour ce même allèle dans la population qui reçoit ces individus (île), la nouvelle fréquence de A à la génération suivante G_1 après l'arrivée de m migrants sera:

$$p_1 = (1 - m) p_0 + m p_m$$

La variation de la fréquence allélique Δp de l'allèle A entre ces 2 générations sera alors :

$$p_1 - p_0 = m (p_m - p_0)$$

Explication :

$$p_1 - p_0 = (1 - m) p_0 + m p_m - p_0$$

$$p_1 - p_0 = p_0 - m p_0 + m p_m - p_0$$

$$p_1 - p_0 = m p_m - m p_0$$

$$p_1 - p_0 = m (p_m - p_0)$$

Soit entre les générations t et $t+1$:

$$p_{t+1} - p_t = m (p_m - p_t)$$

La variation des fréquences alléliques est donc proportionnelle au flux migratoire m et à la différence de fréquence entre les deux populations. L'équilibre p_e (ou q_e) sera atteint lorsque $\Delta p = 0$ c'est-à-dire lorsque les fréquences alléliques dans la population qui reçoit de nouveaux individus (l'île) seront les mêmes que celles de la population immigrante (le continent).

$$p_e = p_m$$

$$q_e = q_m$$

Après t générations de migrations, il y a une proportion $(1-m)^t$ d'individus résidents, où la fréquence de A est p_0 , et $1-(1-m)^t$ migrants, où la fréquence de A est p_m . La fréquence allélique à la génération t dans l'ensemble de la population sera :

$$p_t = (1-m)^t p_0 + (1 - (1-m)^t) p_m$$

C'est-à-dire :

$$p_t = (1-m)^t (p_0 - p_m) + p_m$$

Lorsque les deux sens de migrations existent, les fréquences des allèles dans les deux populations seront fonction des fréquences initiales et de leurs taux de migration respectifs.

3.2. Autres modèles de migration

Lorsque plus de deux populations sont partiellement isolées, les échanges de migrants sont multidirectionnels et l'analyse des conséquences des flux géniques se complexifie rapidement. Plusieurs types de modèles de migration ont été proposés.

3.2.1. Le modèle de l'archipel

Correspond à un ensemble d'îles interconnectées (Figure 9) par des échanges de migrants dans toutes les directions possibles avec des conséquences sur l'ensemble des populations.

Sous l'effet de ces flux géniques multidirectionnels, les différences de fréquences alléliques entre populations diminuent progressivement, et les populations convergent vers une fréquence allélique commune qui correspond à la moyenne des fréquences alléliques dans les populations concernées, pondérée par leurs effectifs respectifs.

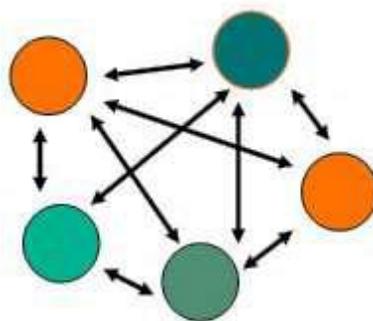


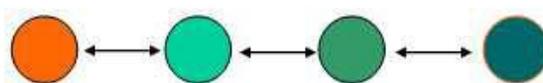
Figure: Modèle archipel.

3.2.2. Le modèle linéaire

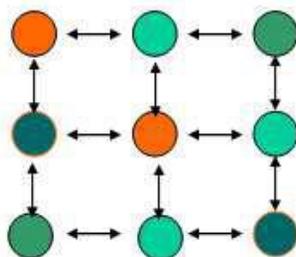
Appelé aussi modèle en pas japonais (stepping stone model), correspond au cas où les échanges entre populations sont « canalisés » par une structure particulière de l’habitat.

C’est ce qui se passe dans une rivière, où les populations de poissons sont échelonnées de l’amont vers l’aval, les poissons de la population amont ne pouvant atteindre la population aval qu’en passant par les populations intermédiaires.

Dans ces conditions, l’intensité des échanges migratoires est directement corrélée à la distance géographique entre les populations



modèle une dimension



modèle 2 dimensions

Figure: Modèle de migration en pas japonais (stepping stone model).

3.2.3. Le modèle en méta-populations

Est certainement le plus réaliste puisqu'il prend en compte une taille inégale des populations, des flux migratoires de différente intensité et surtout la possibilité que certaines populations disparaissent (extinction) alors que d'autres sont nouvellement créées par colonisation.

L'analyse du fonctionnement en métapopulation doit alors incorporer les effets de la variation des effectifs (dérive génétique) (Figure 11).

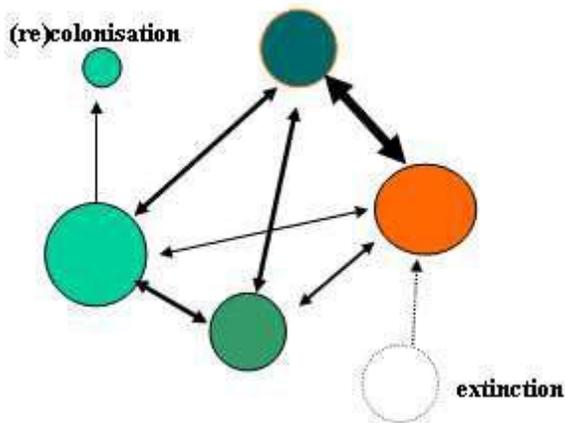


Figure: Modèle en métapopulation.

3.3. Estimation des taux de migrations

Contrairement aux taux de mutation qui sont très faibles, les taux de migration peuvent être élevés et leurs effets sont souvent importants et rapides. Les migrations tendent à uniformiser les fréquences alléliques des populations concernées, et ainsi s'opposent aux phénomènes qui tendent au contraire à les diversifier (dérive génétique ou sélection). La différenciation réelle que l'on observe entre les populations traduit donc le compromis qui s'établit entre ces processus. La mesure directe des flux migratoires n'est possible que par le repérage ou le

marquage des individus, et n'est réalisable que chez des animaux de grande taille (principalement les vertébrés y compris l'homme).

Chez la plupart des organismes, les flux géniques ne peuvent qu'être estimés par des mesures indirectes grâce à l'utilisation de marqueurs génétiques (biochimiques ou moléculaires) dont l'expression phénotypique est la plus faible possible (marqueurs neutres).

Plus les fréquences alléliques de ces marqueurs sont différentes entre populations, moins les flux géniques sont intenses et moins les flux migratoires sont importants.

Des modèles permettent d'estimer les flux migratoires (en nombre d'individus par génération) à partir des structures génétiques des populations (variance des fréquences alléliques entre populations, statistique F_{st}). En multipliant les locus examinés, on peut rechercher et analyser les déséquilibres gamétiques, ce qui constitue une méthode encore plus puissante. C'est grâce à ces méthodes qu'il est possible d'évaluer l'impact génétique des introductions volontaires sur les populations indigènes (alevinages sur les populations de poissons dans les rivières), ou d'étudier les déplacements naturels de populations sur de très grandes distances. En résumé, les migrations sont à l'origine de flux géniques importants entre populations. Quel que soit le modèle considéré (modèle insulaire, archipel, en japonais, métapopulations), ces flux géniques tendent à uniformiser les fréquences alléliques des populations interconnectées. Ils s'opposent donc aux facteurs qui font diverger la composition génétique des populations (dérive génétique). Lorsque le repérage direct des individus n'est pas possible, la mesure des flux migratoires se fait indirectement en utilisant des modèles génétiques qui considèrent que les flux géniques entre populations sont d'autant plus faibles que la différenciation entre ces populations est plus forte.

4. La sélection

Le modèle de Hardy-Weinberg suppose que l'effet de la sélection est négligeable sur quelques générations, une condition évidemment irréaliste sur une grande échelle de temps puisque la sélection, avec d'autres forces y fut responsable de l'évolution du vivant.

Charles Darwin fut le premier des naturalistes à concevoir la sélection naturelle comme une force active de la transformation (l'évolution) des espèces. La pensée darwinienne est populationnelle car l'évolution des espèces est conçue non comme une adaptation ou une transformation des individus sous les contraintes de l'environnement, mais comme le tri, au sein des populations, par la sélection naturelle, des individus qui sont le mieux adaptés à ses contraintes, et qui, de ce fait, laissent un plus grand nombre de descendants. Cette fécondité différentielle induit une modification de la diversité (génétique) des populations vers une augmentation de la fréquence des types les mieux adaptés, et aboutit, après une longue période, à une transformation de l'espèce. À la même époque, un autre naturaliste, Alfred Russel Wallace, vivant aux Indes, avait, indépendamment de Darwin, exprimé les mêmes conceptions. Il s'en était d'ailleurs ouvert à Darwin lui-même qui se dépêcha d'écrire un essai afin de publier ses conceptions conjointement avec celles de Wallace, dans le même numéro du journal of the Linnean society. Darwin poursuivit et approfondit sa réflexion et, par sa position en Angleterre, s'imposa comme le seul fondateur de la théorie de l'évolution par la sélection naturelle (Jean-Pierre, 2008).

Le fait que les concepts développés par Darwin aient pu surgir à la même époque chez un autre naturaliste signifie tout simplement que les esprits étaient mûrs pour reconsidérer complètement la problématique transformiste en dépassant les conceptions économiques, philosophiques et théologiques du XVIIIème siècle. Les conceptions darwiniennes apparaissent, pour tout historien des idées, comme un transfert, dans les sciences naturelles, de la pensée libérale, économique et philosophique, qui accompagna le développement

industriel et commercial du début du XIXème siècle. Que Darwin, à des passages clé de son ouvrage, fit explicitement référence à Malthus n'empêcha nullement Marx et Engels de considérer la théorie darwinienne comme de la plus grande importance sur le plan philosophique, révolutionnaire au sens Hégélien du terme. Cependant la sélection telle que l'a conçue Darwin, ne résout pas vraiment toute la question de l'évolution des espèces. D'une part la formulation mathématique de la sélection, si elle rend compte de l'évolution

darwinienne de la diversité des populations par la « sélection du plus apte », apporte aussi des résultats inattendus dans le cadre strict du darwinisme, dont la possibilité de maintenir cette diversité. D'autre part cette formulation montre qu'une population soumise à une contrainte sélective de l'environnement peut évoluer dans plusieurs directions, parfois opposées, ce qui pose le problème de la force qui « exerce le choix évolutif ».

Enfin le coût de la sélection est incompatible avec l'importance de la diversité observée dans les populations naturelles, ce qui implique qu'une grande part de cette diversité est sélectivement neutre, mais que sa variation participe obligatoirement à la définition des espèces nouvelles ; la question se pose alors de définir les forces qui participent à la variation de la diversité sélectivement neutre (Serre, 2006).

5. La dérive génétique

Est l'évolution d'une population ou d'une espèce causée par des phénomènes aléatoires, impossible à prévoir. Du point de vue génétique, c'est la modification de la fréquence d'un allèle, ou d'un génotype, au sein d'une population, indépendamment des mutations, de la sélection naturelle et des migrations. La dérive génétique est causée par des phénomènes aléatoires et imprévisibles, comme le hasard des rencontres des spermatozoïdes et des ovules, dans le cas d'une reproduction sexuée. Les effets de la dérive génétique sont d'autant plus importants que la population est petite, car les écarts observés par rapport aux fréquences alléliques y seront d'autant plus perceptibles. Cette situation peut se produire au moment de l'apparition d'une espèce, ou après un goulot d'étranglement (quand une grande partie d'une espèce a disparu, à la suite de phénomènes épidémiques ou d'une crise climatique ou d'une catastrophe par exemple). Elle peut aussi survenir dans une situation d'insularisation écologique vraie (insularisation causée par une montée de la mer) ou due à la fragmentation écologique. La dérive génétique concerne tous les allèles même si l'impact sur les allèles neutres (c'est-à-dire qui ne confèrent ni avantages ni inconvénients) est plus conséquent. Malgré tout un allèle favorable peut disparaître ou un allèle défavorable se fixer dans une population par dérive, ce qui est fréquent pour des populations aux tailles très réduites.

5.1. La notion d'effectif efficace

Dans les grandes populations, les variations (liées au hasard) du nombre d'enfants produits par des individus de génotypes différents, n'ont pas d'effet significatif sur la fréquence des gènes. Dans les petites populations, ces variations peuvent avoir un effet considérable :

-Si un gène particulier n'est retrouvé que chez un petit nombre d'individus, si ces individus n'ont pas d'enfants ou, que par chance (hasard), ces enfants n'héritent pas de ce gène, le gène en question va complètement disparaître de la population (éteint : fréquence = 0) et son allèle va devenir fixé (fréquence = 1).

La part de la dérive génique aléatoire dépend de la taille de la population.

Elle est plus grande dans les petites populations où les variations dans la fréquence des gènes peuvent être considérables d'une génération à l'autre.

Démonstration :

Influence de la dérive aléatoire sur les fréquences géniques, soit N_e l'effectif de la population des reproducteurs. La nouvelle génération est formée à partir d'un échantillon de $2N_e$ gamètes. Soient p et q la proportion des 2 allèles à la première génération.

La loi de probabilité de p est une loi binomiale de variance

$V_p = pq/2N_e$ et σ (écart-type) = racine carrée de V_p .

Exemple :

$N_e = 50$ et $p = 0.5$

$\sigma^2 = 0.5 \times 0.5/100 = 1/400$

$\sigma = 1/20$

A la génération suivante, p sera compris entre 0.40 et 0.60 (+/- 2 écart-types) avec une probabilité de 95%.

A la génération suivante, si $N_e = 50$ et $p = 0.4$ $\sigma^2 = 0.4 \times 0.6 / 100 = 0.24 / 100 = 1/400$ $\sigma = 0.05$ p sera compris entre 0.3 et 0.5. Dans un tel système, il n'y a que 2 états stables : $p = 0$ et $p = 1$. Sous l'influence de la dérive, une population évoluera vers l'un de ces deux états.

Dans une petite population, lorsqu'une mutation apporte un nouvel allèle, le destin ultime de celui-ci est :

-Soit de disparaître : probabilité = $1 - 1/2 N_e$

-Soit d'éliminer les autres : probabilité = $\frac{1}{2N}$

Dans une population de 500 individus, un nouvel allèle a 1 chance sur 1000 d'éliminer les autres et 999 chances sur 1000 d'être éliminé.

La durée moyenne est de :

$2 \log(2N)$ génération pour la disparition.

$4N$ générations pour la fixation.

Chapitre V. La Spéciation.

1. Introduction

La notion d'espèce est d'autant plus importante en biologie qu'elle constitue l'unité taxinomique fondamentale évolutive la plus évidente pour tous.

Son utilisation dans plusieurs domaines de la biologie en précise l'intérêt :

- Les taxinomies et les systématiques reposent souvent sur la notion d'espèce.
- Le raisonnement génétique est fondé sur la transmission d'allèles chez des individus d'une même espèce et rarement au niveau du genre ou de la famille.
- L'étude de l'évolution se rapporte à celle de l'espèce.
- L'écologie utilise abondamment la notion d'espèce: reconstitution de réseaux trophiques,prévisions des conséquences de pollutions, lutte biologique ou non contre des ravageurs...
- La recherche médicale, en parasitologie par exemple, repose sur une définition très précise de l'espèce. C'est ainsi que la recherche d'un vaccin antimalaria a permis de reconnaître six espèces jumelles, chez le moustique vecteur (*Anophelesmaculipennis*). Elle a pu alors continuer à progresser en apportant de nouvelles informations épidémiologiques.

2. Les critères spécifiques et leur discussion

La notion d'espèce repose sur plusieurs critères : morphologique, biologique, écologique et cladiste, qui permettent de la cerner.

2.1. Le concept morphologique