

Variabilité génétique dans les populations naturelles

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.

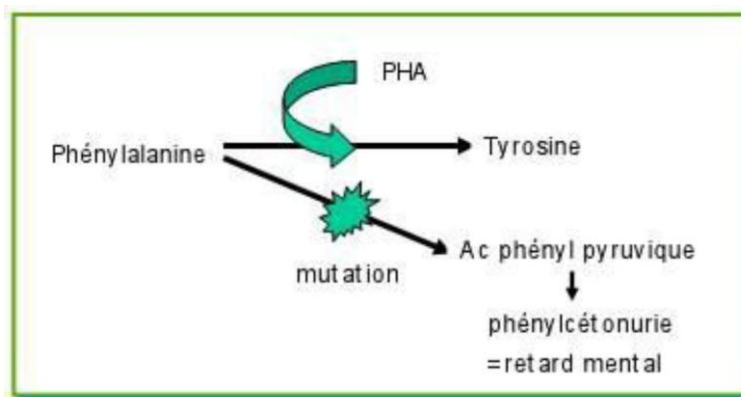
Lorsque les mutations ont des effets sur le phénotype des individus, elles peuvent modifier des caractères biochimiques, physiologiques, anatomiques, morphologiques ou comportementaux. Les mécanismes mis en jeu dans leur expression phénotypique sont divers et complexes. L'expression phénotypique d'un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

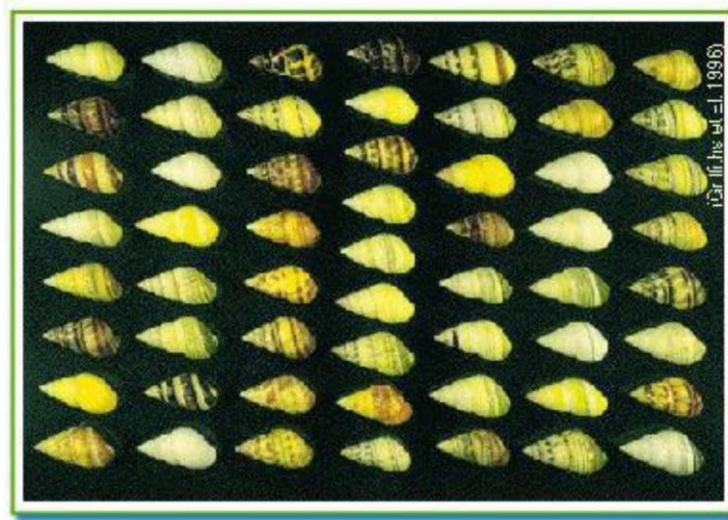
Un exemple célèbre est la variation de la couleur et de l'ornementation de la coquille de l'escargot du genre *Cepaea*. En un même endroit coexistent plusieurs formes phénotypiques déterminées par plusieurs gènes polymorphes : des escargots à coquille rose, jaune ou brune, et des escargots sans bande 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

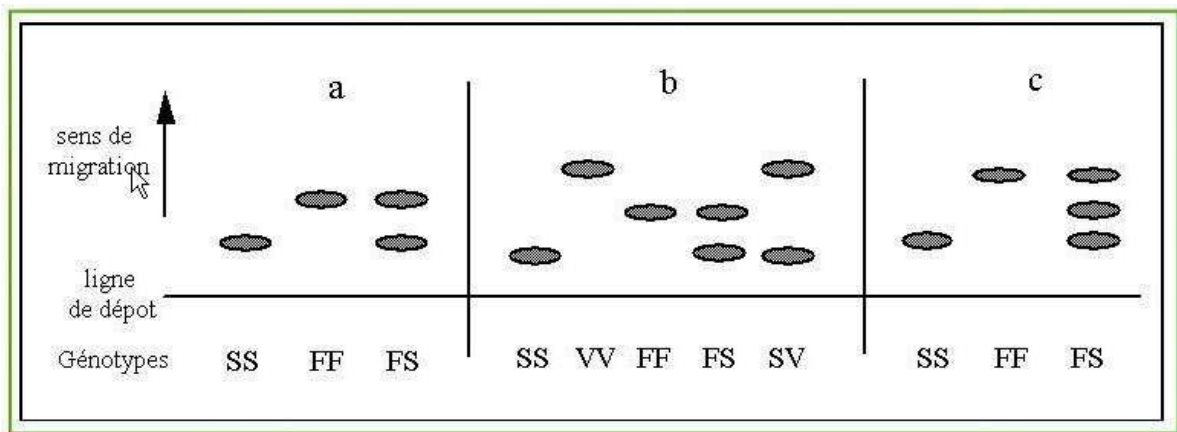
2.2.1. Polymorphisme enzymatique

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 *Drosophilapseudoobscura*. 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. pseudoobscuras* 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-coupage 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**).

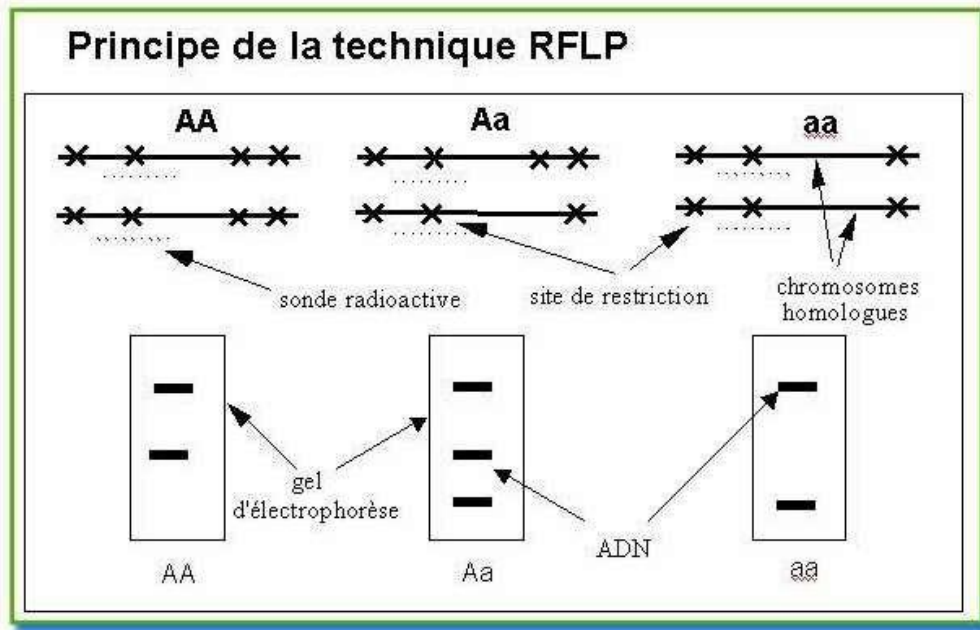
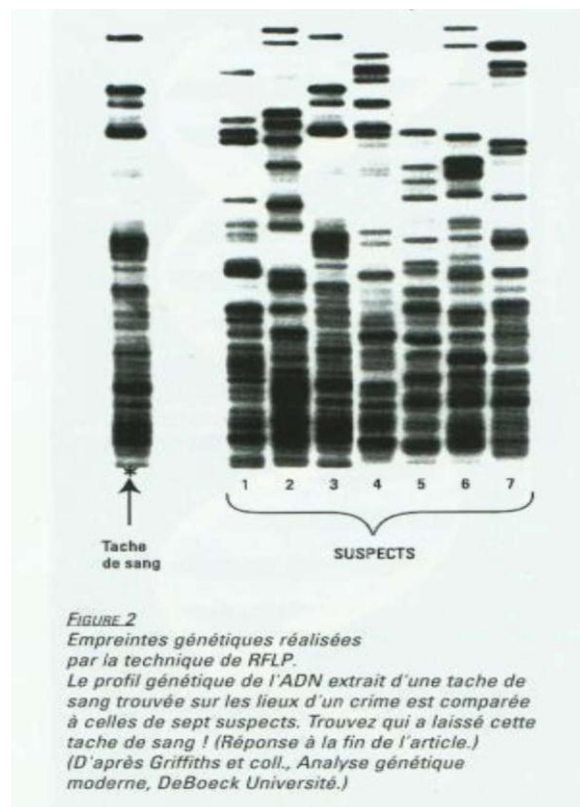
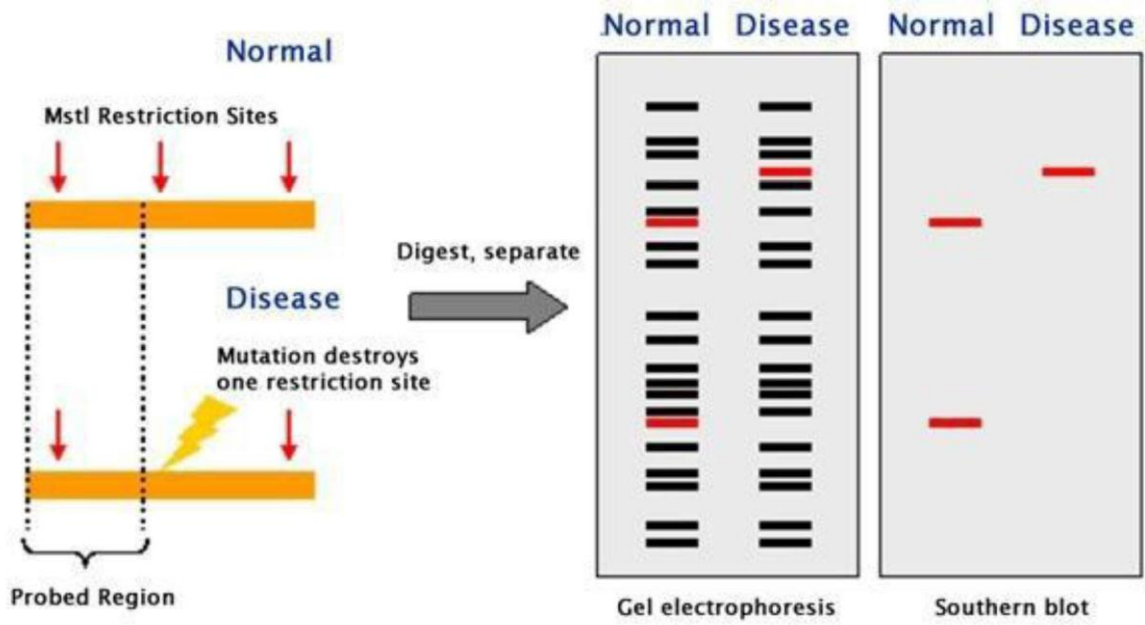


Figure 3: Principe de la technique RFLP.

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).

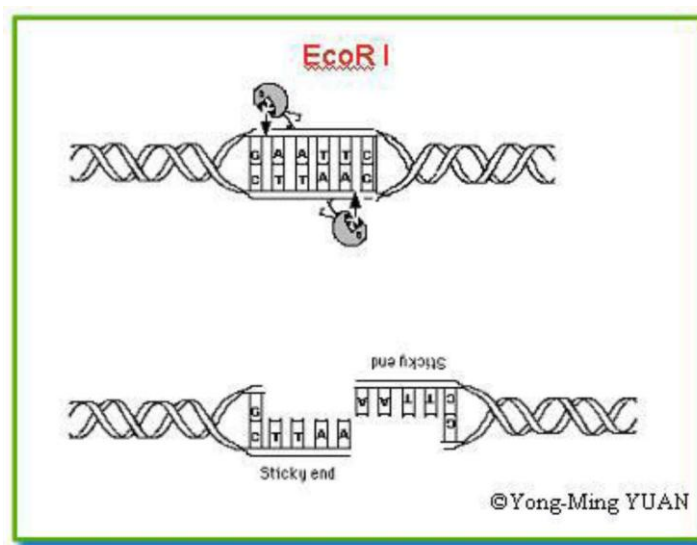
d'abord un fragment de restriction est un morceau de DNA qui a été coupé par un enzyme de restriction (enzyme coupant le DNA).

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.

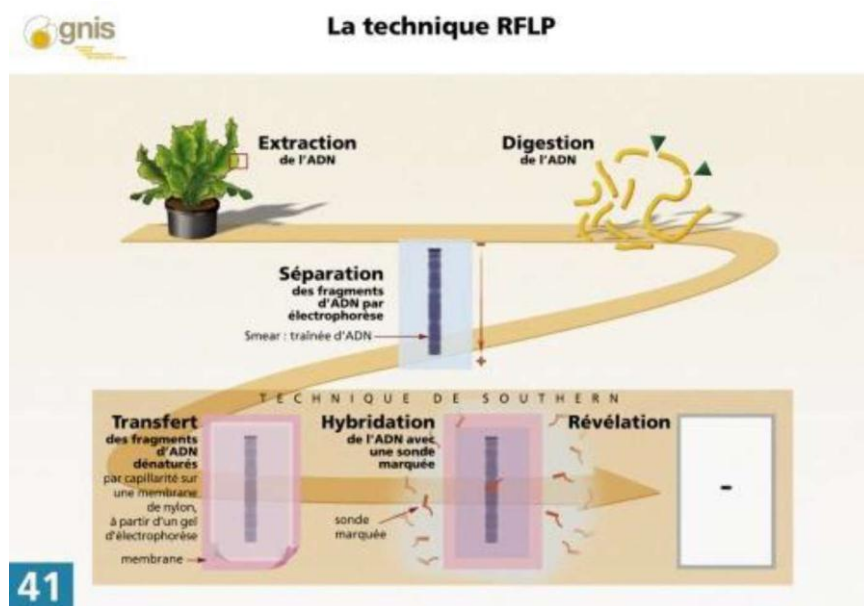
Si le gène A possède 2 fois la séquence GATC il donnera 3 fragments et si le gène B en possède 4 il donnera 5 fragments (et ainsi de suite).

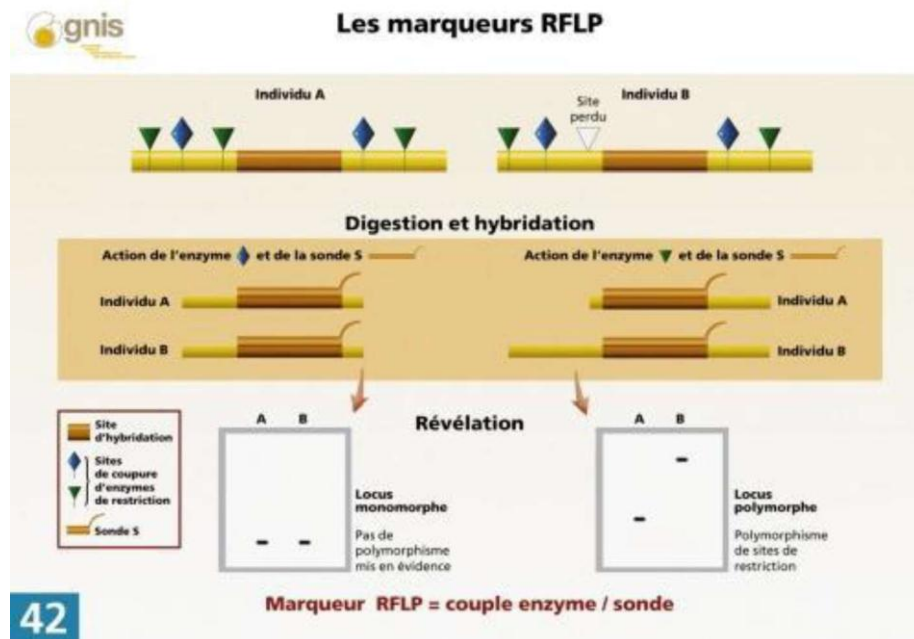
En utilisant un enzyme particulier on sait qu'on doit s'attendre à trouver une série précise de fragments de tailles déterminées. C'est la méthode du finger printing qui permet notamment de reconnaître un gène.

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 fragments 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.



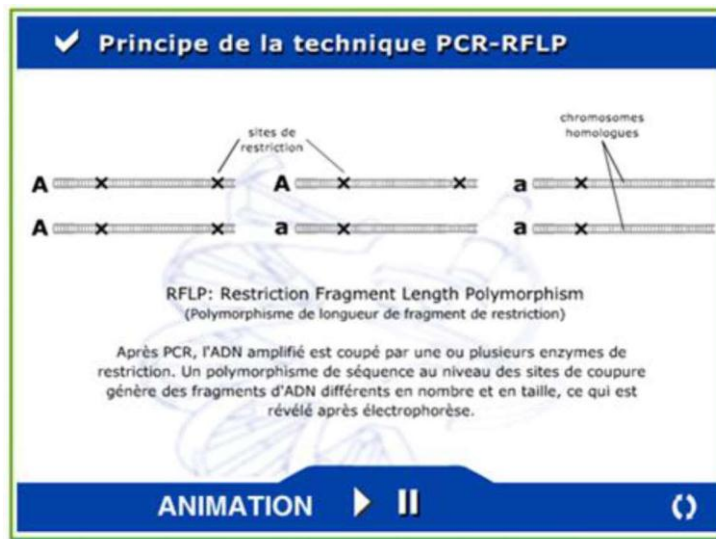


Comparons deux individus A et B. Leur ADN sera digéré séparément par des enzymes de restriction données (triangle et losange), puis hybridé par une sonde S.

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'éthydiuM 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

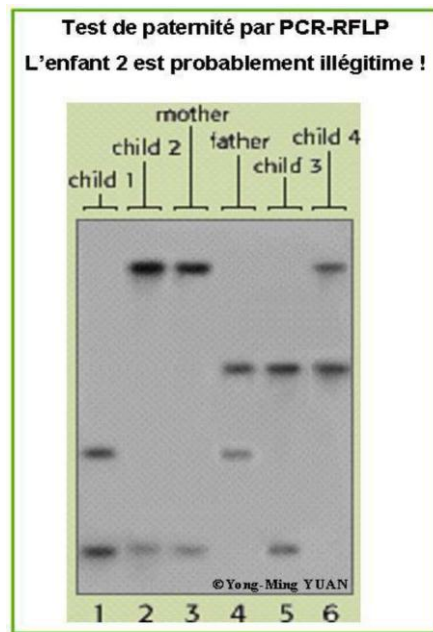
Les microsatellites sont des courtes séquences d'ADN des Chromosomes formées de la répétition d'un motif lui-même constitué de une à quatre bases. Les plus étudiés sont les répétitions (CA)_n. Très polymorphes, ce sont de bons marqueurs pour la cartographie génétique. Il existe dans le génome de très 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** autosomiaux 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

$$2NAA + NAa .$$

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

$$f(A) = p = (2NAA + NAa)/2N \quad f(a) = q = (2Naa + NAa)/2N \quad \text{avec } p + q = 1$$

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),
Cours de Génétique et Dynamique des Populations Dr : bouamou A
~~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 :

$$H_t = 1 - (f_1^2 + f_2^2 + \dots + f_k^2) = 1 - \sum f_i^2$$

C'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 :

$$A_e = 1 / (f_1^2 + f_2^2 + \dots + f_k^2) = 1 / \sum f_i^2$$

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