

# **POLYMORPHISME GENETIQUE**

# PLAN

I. Introduction

II. Définition et caractéristiques des polymorphismes

III. Les marqueurs génétiques

3.1. RFLP

3.2. Microsatellites

3.3. Minisatellites

3.4. SNP

3.5. CNV

IV. Applications des marqueurs génétiques:

# I- Introduction

- Le développement de la biologie moléculaire du **gène** puis le **séquençage** des génomes, ont permis de mettre en évidence l'existence de **polymorphismes moléculaires de l'ADN**, une diversité génétique qui ne touche pas seulement les séquences des gènes mais se répartit sur l'ensemble du génome, aussi bien dans les séquences codantes que dans les séquences intergéniques.

- De ce fait, ces polymorphismes sont très utiles dans nombre d'applications de la génétique des populations à la biodiversité ou à l'épidémiologie génétique.
- En effet, ils peuvent servir de marqueurs génétiques dans la mesure où on peut:
  - ✓ attribuer à leurs différents états le statut d'« allèles », homozygotes ou hétérozygotes,
  - ✓ suivre la transmission de ces marqueurs dans une généalogie,
  - ✓ estimer leur diversité génétique dans une population ou une espèce et l'analyser.

## II. Définition et caractéristiques des polymorphismes:

- Les polymorphismes génétiques correspondent à des **variations de la séquence de l'ADN** d'un individu à l'autre au sein d'une même espèce.
- Le polymorphisme de l'ADN est généré au cours de l'évolution par différents mécanismes moléculaires produisant soit des **substitutions** nucléotidiques, soit des **insertions** ou **délétions** d'une ou plusieurs bases.
- Ces variations sont **fréquentes** , on les trouve toutes les 100 à 300 bases .

- Elles ne traduisent **aucune conséquence pathologique** => elles ne dénaturent pas le message héréditaire soit:
  - ✓ Parce qu'elles siègent dans des **régions non codantes** ( introns, ou séquences inter géniques).
  - ✓ Parce qu'elles **n'altèrent pas le contenu informationnel des exons** donc n'altèrent pas la signification du codon.

- elles sont **stables** et **transmises de manière mendélienne** traduisant ainsi un polymorphisme génotypique qui est le support moléculaire de la diversité phénotypique intra spécifique des individus.
- Il existe **un nombre potentiellement considérable des sites polymorphes**, il est de l'ordre de 10 millions.
- Généralement, seules les différences observées avec **une fréquence supérieure à 1 %** sont considérées comme des polymorphismes, les autres étant assimilées à des anomalies ou mutations.

- On distingue ainsi trois types de polymorphisme:

Le polymorphisme de séquence, le polymorphisme d'insertion-délétion et le polymorphisme de nombre d'unités de répétitions dans les régions répétées.

### 1. Les polymorphismes de restriction:

- Mis en évidence en 1978 par la méthode de Southern.
- Ce sont des variations ponctuelles de la séquence touchant un site de restriction
- Appelés **RFLP** Restriction Fragment Length polymorphisme, Polymorphisme de longueur des fragments de restriction.



## 2. Les polymorphismes de répétition:

- Dus à des différences dans le nombre de copies de séquences répétées en tandem. On distingue deux catégories:

### ✓ Les mini satellites VNTR:

- Découverts en 1985 par la méthode de Southern
- Le motif de base répété en tandem est d'une longueur  $> 10$  nucléotides.

### ✓ Les microsatellites STR:

- Mis en évidence en 1989 par la méthode d'amplification par PCR
- Le motif de base répété en tandem est plus court (1 à 4 NC).

## • 3. Les polymorphismes d'insertion et de répétition ou CNV:

- Un segment d'ADN de 1 kb ou plus est présent en un nombre variable de copies, comparé à un génome de référence.

### III. Les marqueurs génétiques:

- Les techniques moléculaires qui révèlent le polymorphisme de l'ADN ont connu un développement très important ces dernières décennies.
- Avec l'augmentation du nombre des marqueurs, il est aujourd'hui possible de décrire ces variations génétiques à l'échelle du génome.

- Un **marqueur** est un **trait génotypique** représenté par un **gène** ou une **séquence polymorphe** ( RFLP) permettant de **repérer** un chromosome .
- Il doit correspondre à un **locus** bien individualisable ( c-à-d a une position définie sur le génome) et doit être **informatif**.
- Ils permettent de caractériser un génome de manière fiable, spécifique et rapide
- On peut les utiliser en cartographie génétique pour « **baliser** » le génome et **identifier** des individus ou des espèces.

# L'informativité d'un marqueur

- On définit ainsi **l'informativité** d'un marqueur comme étant **l'hétérozygotie du locus** étudié, et donc **la probabilité** pour un **individu d'être hétérozygote** au locus étudié (deux allèles différents).
- Plus **l'informativité** d'un marqueur est élevée, plus il sera efficace de repérer un locus sur l'un ou l'autre des chromosomes d'une même paire.

- Un marqueur **codominant** est **plus informatif** qu'un marqueur dominant, dont l'allèle récessif n'est observable qu'à l'état homozygote.
- D'autre part, un marqueur est d'autant plus **informatif** que le nombre d'allèles est élevé et que leurs fréquences sont équilibrées.
- Ces considérations expliquent pourquoi les généticiens recherchent des marqueurs codominants très polymorphes

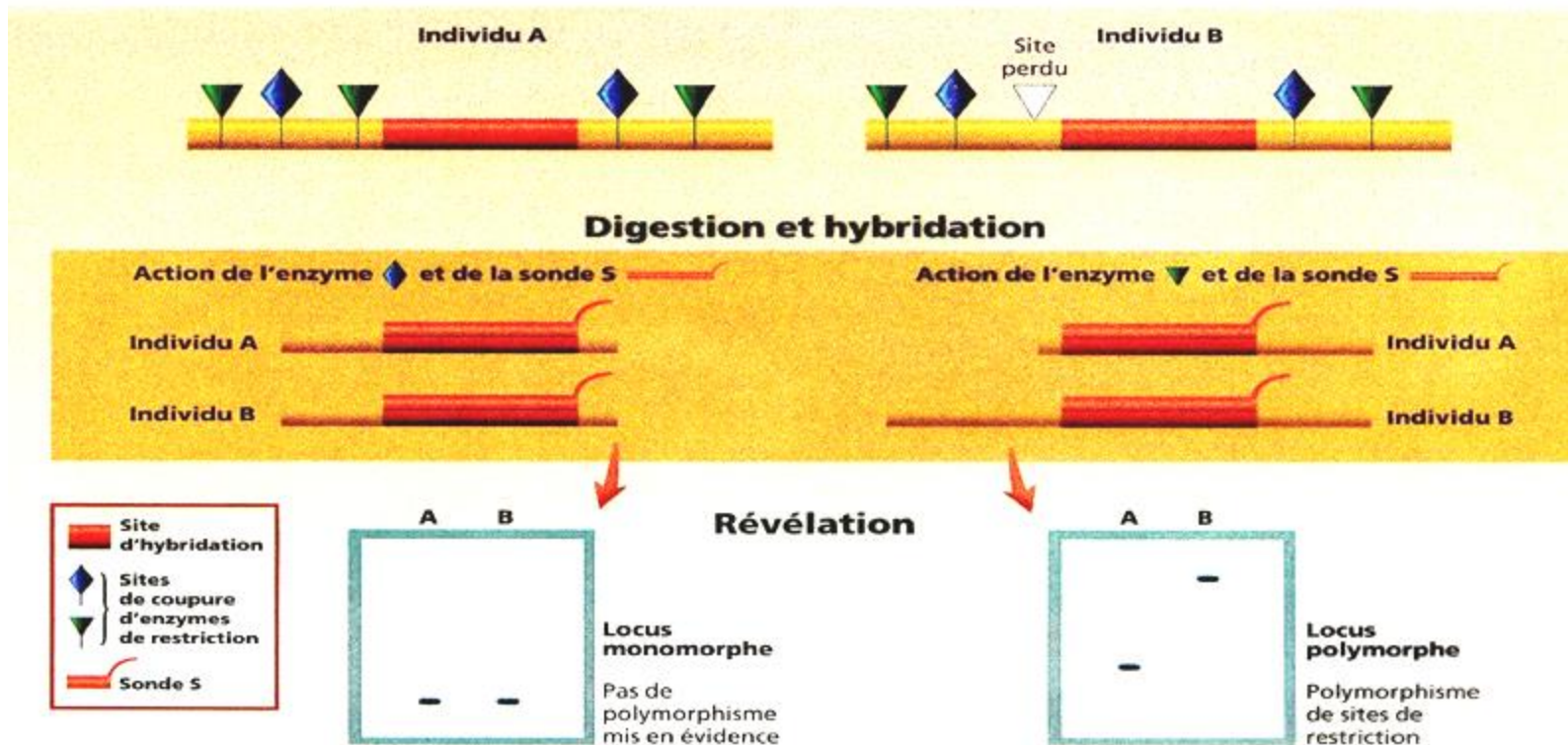
# 3.1.RFLP:

## Restriction Fragment Length polymorphism, Polymorphisme des fragments de restriction

- Toute modification des séquences d'ADN (mutation, addition, délétion ) **abolissant** ou **créant** un **site de restriction**.
- En 1980 Botstein et Coll ont publié une carte génétique du génome humain qui, pour la première fois, utilisait des marqueurs génétiques issus de la technique de polymorphisme de longueur des fragments de restriction **RFLP**. Cette technique a eu immédiatement un gros succès auprès des généticiens, car elle donnait accès à un nombre très élevé de marqueurs répartis le long du génome.
- En utilisant une **enzyme de restriction**, le génome d'un individu va être découpé en plusieurs parties dépendant du nombre de **sites de restriction** présents dans le génome de l'individu pour l'enzyme utilisée. Le nombre de ces sites et leurs positions diffèrent en fonction de l'individu. On a donc un **polymorphisme de longueur des fragments de restriction**.

- Lorsque deux individus homozygotes sont comparés après digestion de leur ADN par une enzyme donnée et hybridation avec une sonde donnée, on observera des profils d'électrophorèse différents chaque fois que des sites de restriction différeront.

### Les marqueurs RFLP



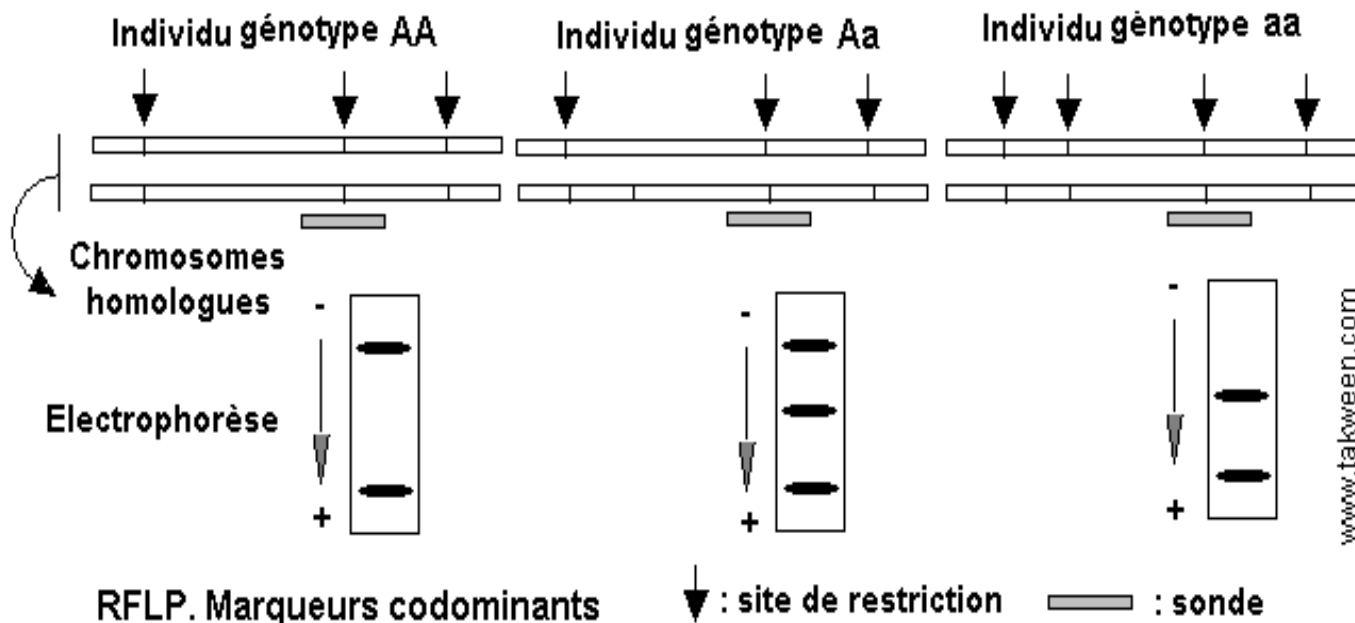
**Marqueur RFLP = couple enzyme / sonde**



- C'est le couple **enzyme/sonde** qui constitue le marqueur. La sonde révèle un locus polymorphe ou monomorphe.

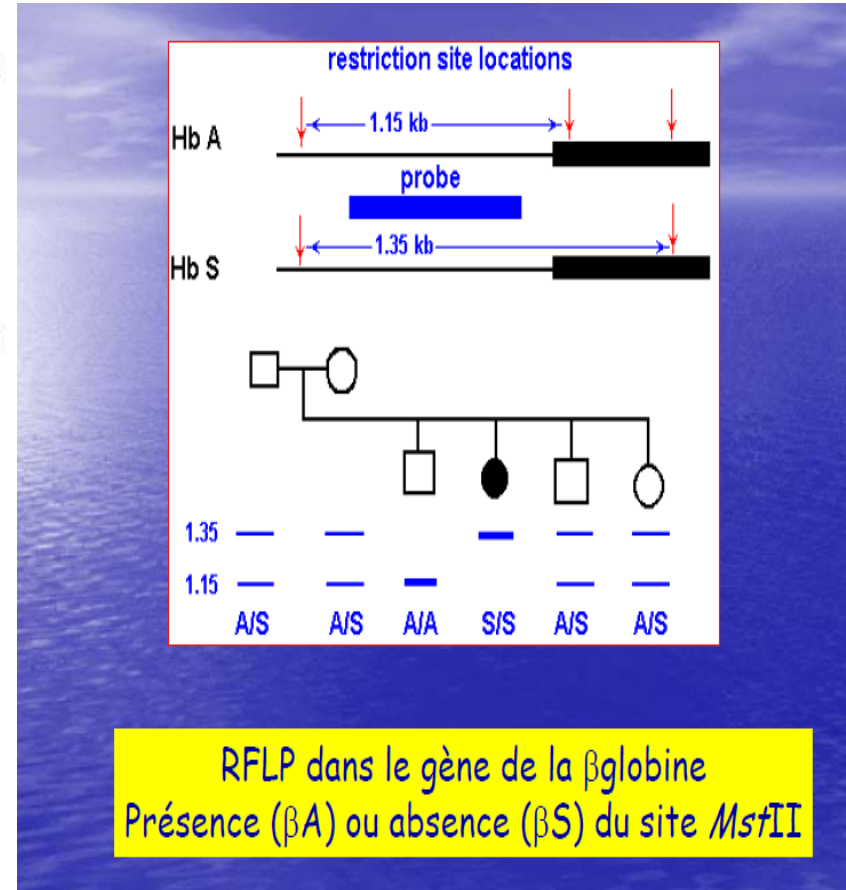
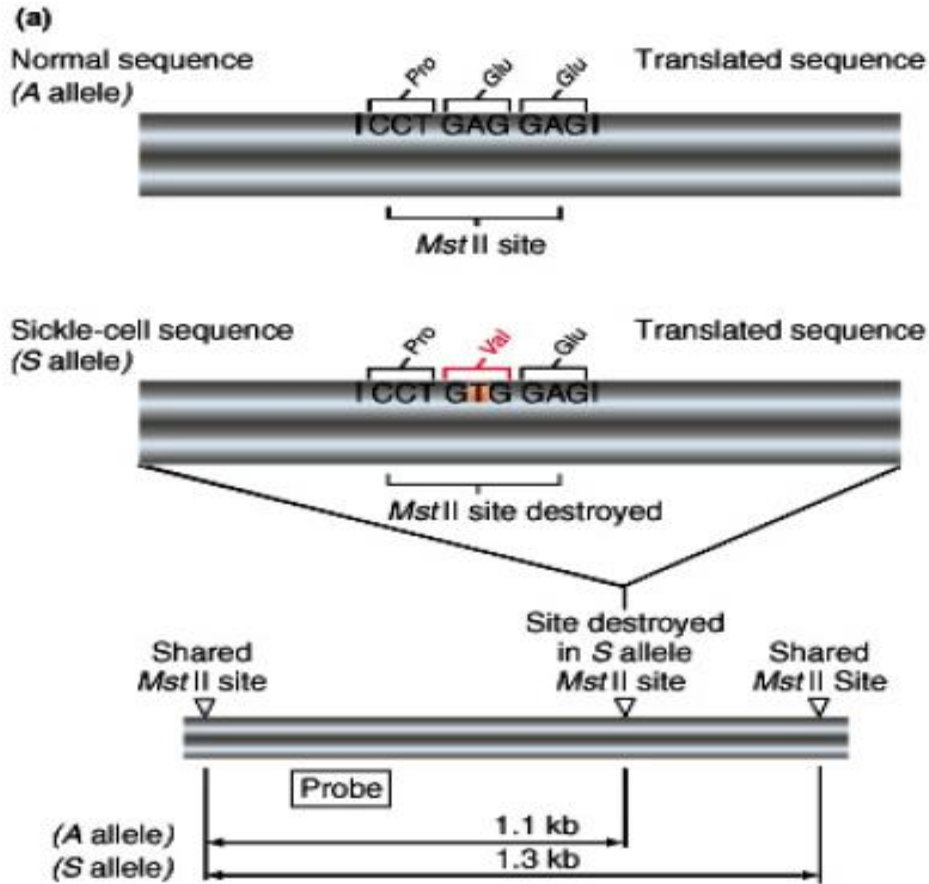
Les enzymes de restriction permettent de visualiser le nombre d'allèles détectables à ce locus dans une population.

- Les RFLP sont des marqueurs codominants





# Exemple de l'anémie falciforme : perte d'un site MstII



## Les étapes de la technique de RFLP

- Méthode du Southern Blot:

**1) Extraction de l'ADN** des différents géotypes à analyser.

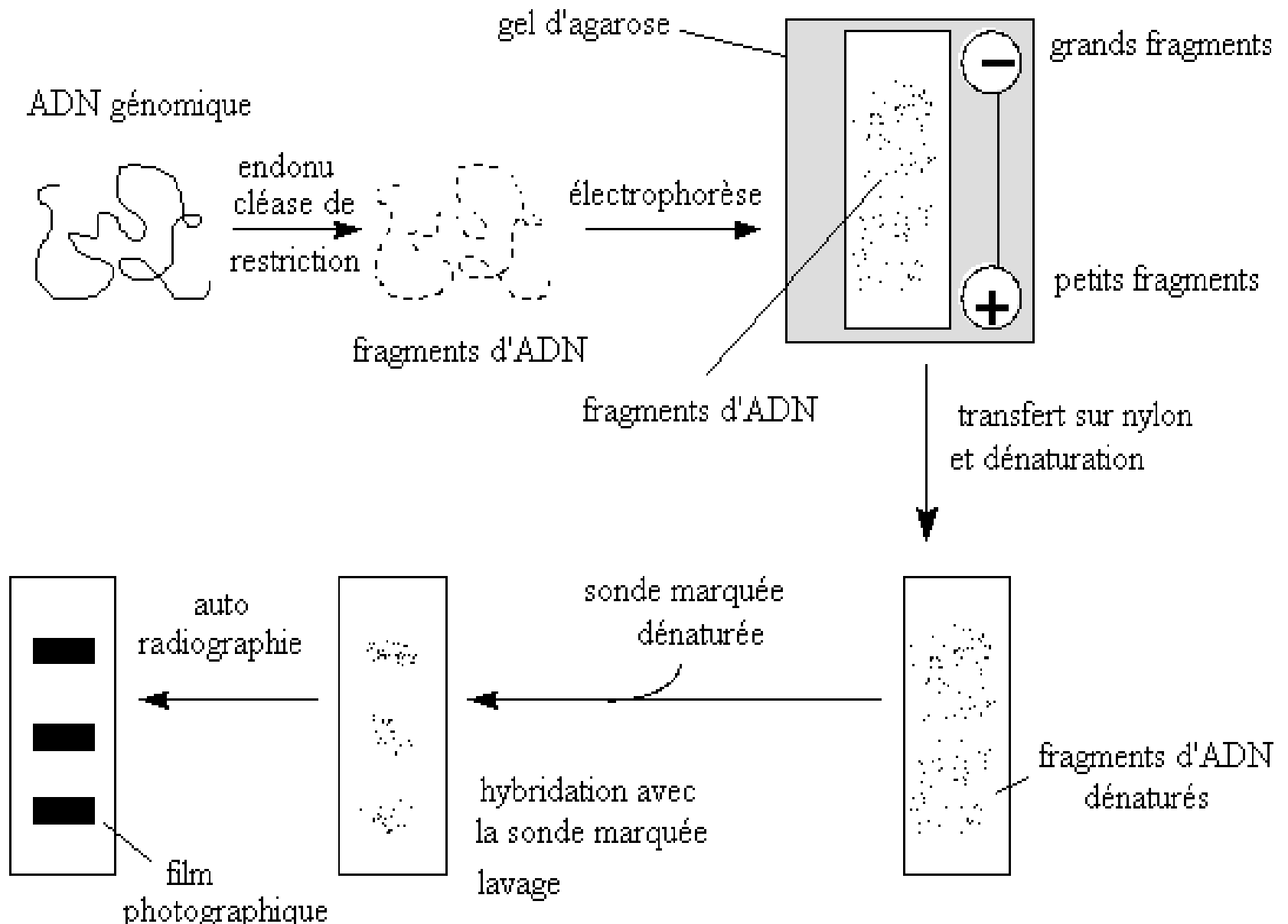
**2) Digestion de l'ADN** par une **enzyme de restriction**. On obtient alors des fragments d'ADN de longueurs différentes selon les individus.

Cette différence de taille est due à une différence du nombre de site de restriction qui varie selon les individus ; l'ADN à analyser est alors coupé en fragment plus ou moins long du au polymorphisme de longueur des fragments de restriction.

- Les étapes qui suivent servent à visualiser les différents fragments d'ADN obtenus :

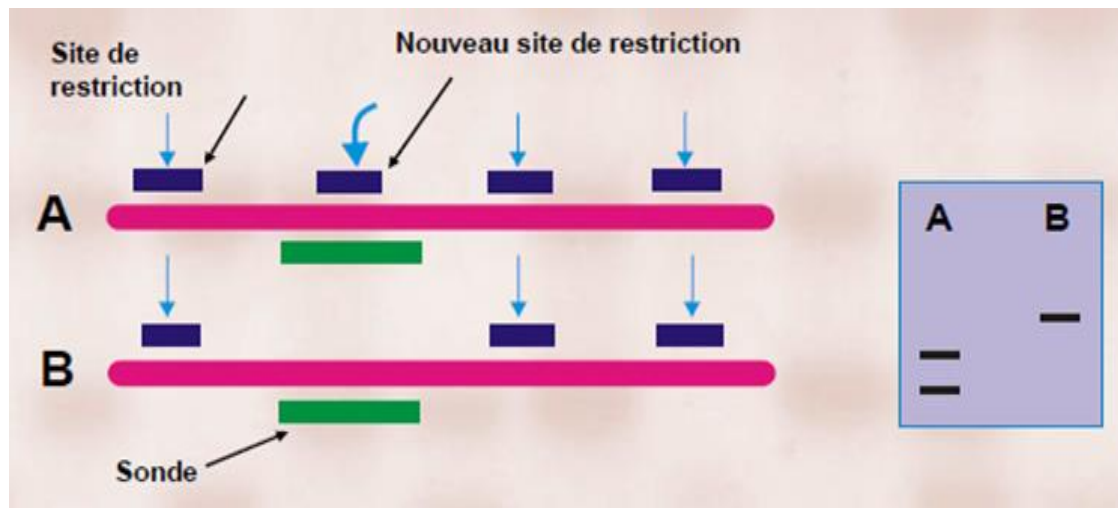
**3)** Les fragments de restriction sont séparés selon leur taille par une **électrophorèse en gel d'agarose**. L'ADN étant chargé négativement, il migre de la cathode vers l'anode. Les fragments les plus petits sont les plus rapides.

**4)** L'ADN est **transféré** sous forme dénaturée sur une membrane de nylon. La position relative des fragments d'ADN est préservée durant le transfert.

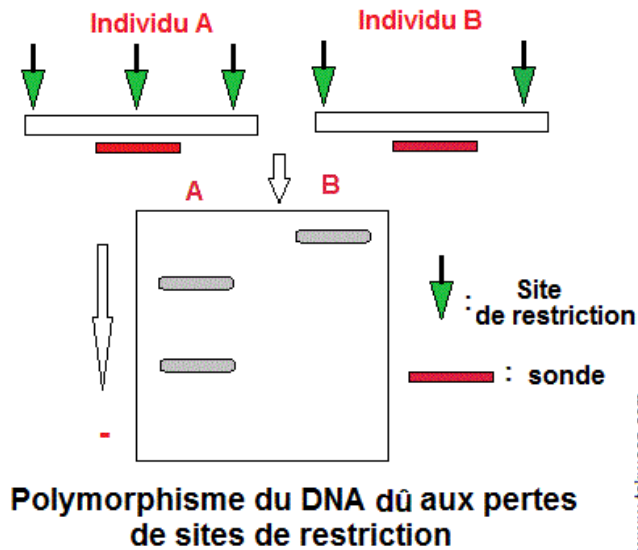


- La membrane est incubée dans une solution contenant une **sonde** marquée préalablement, soit par la **radioactivité**, soit **chimiquement**.
- La sonde s'hybride alors avec le ou les fragments d'ADN avec lesquels elle présente une homologie. On utilise couramment deux sources de sondes:
  - ✓ les sondes **génomiques (ADNg)**: Les sondes génomiques sont obtenues par digestion de l'ADN total du génome nucléaire de l'espèce étudiée à l'aide d'une enzyme de restriction.
  - ✓ les sondes **d'ADN complémentaire (ADNc)**: correspondent nécessairement à des gènes exprimés, puisqu'elles sont obtenues à partir des ARN messagers.
- L'endroit, ou les endroits, où la sonde s'est fixée sont révélés en plaçant la membrane au contact d'un film sensible à la radioactivité, ou en réalisant une réaction enzymatique colorée spécifique.

- **Origines du polymorphisme de restriction (RFLP) :**
  - - Perte ou gain d'un site de restriction (mutation ponctuelle)
  - - Mutation de type insertion-déletion ( indel).
  - Dans les deux cas les loci sont généralement bialléliques et l'expression des allèles est codominante.
- ❖ **Gain d'un site de restriction dans la région cible.**
- Deux bandes plus courtes sont alors détectées sur le film autoradiographique.

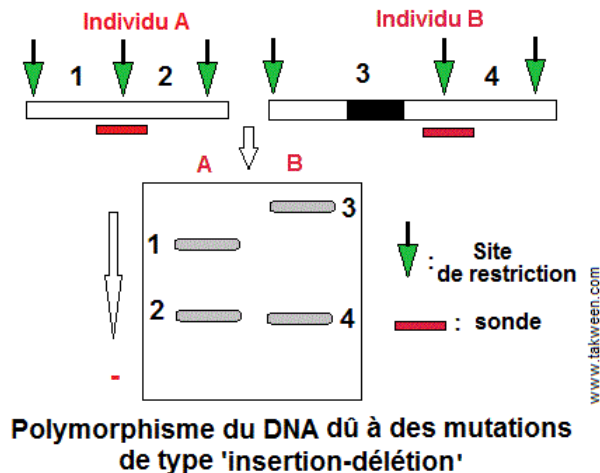


## ❖ Perte d'un site de restriction:



## ❖ Polymorphisme du aux mutations de type 'insertion - délétion'

- Une insertion d'une séquence d'ADN entre deux sites de restrictions flanquants produit un fragment de restriction plus grand.



- **Avantages des RFLP**

- Méthodologie très robuste, bien transférable entre laboratoires
- Hérité de façon codominante et, de ce fait, permet d'estimer l'hétérozygotie
- Bien adaptée pour la construction de cartes de liaison génétique

- **Inconvénients des RFLP**

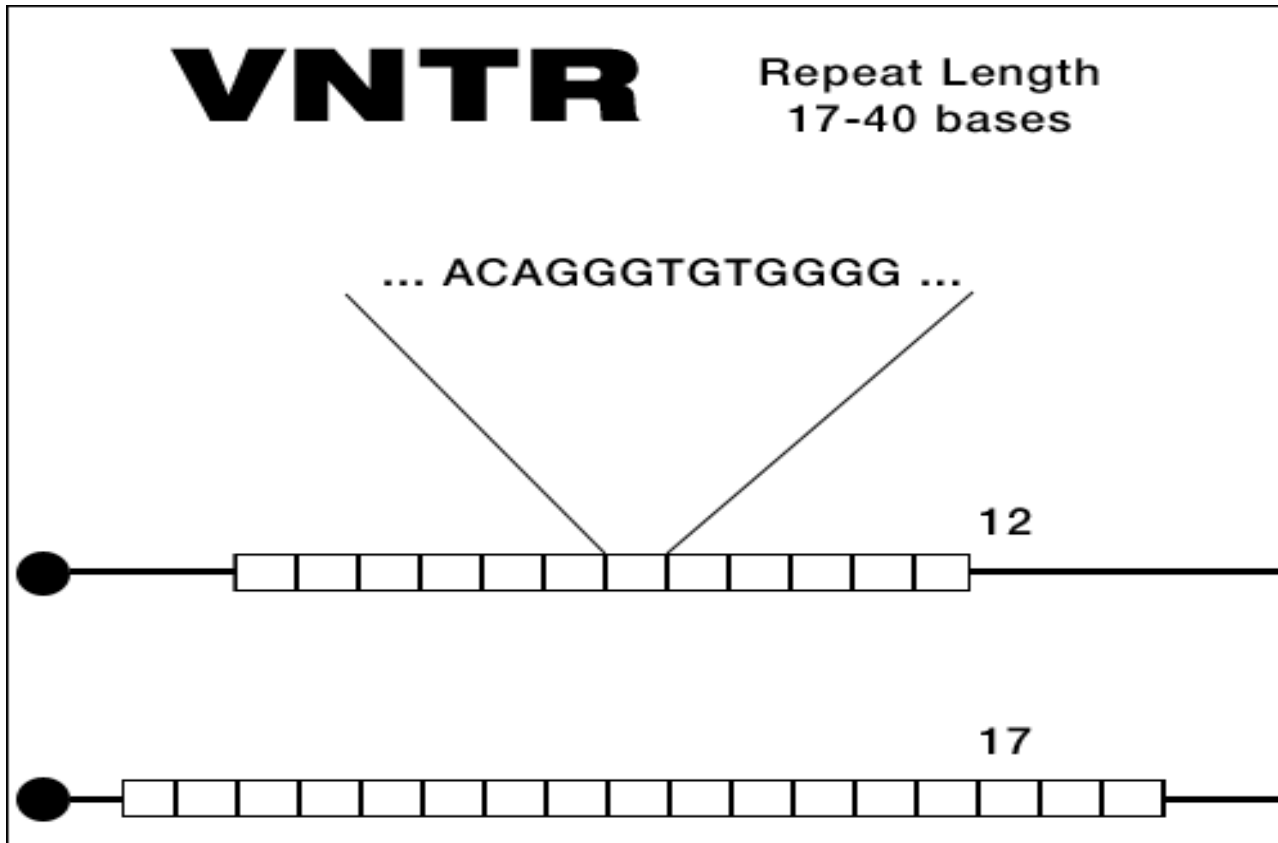
- Grandes quantités d'ADN nécessaires
- Automatisation impossible
- Faible niveau de polymorphisme chez certaines espèces
- Peu de locus sont détectés par expérimentation
- Nécessite une banque de sondes appropriée
- Coûteux

# les séquences répétées

- 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:
  - les minisatellites qui sont des répétitions de motif ayant 10 à 60 paires de bases (pb).
  - les microsatellites qui sont des répétitions de motif ayant 1 à 6 paires de bases (pb).



# les séquences répétées

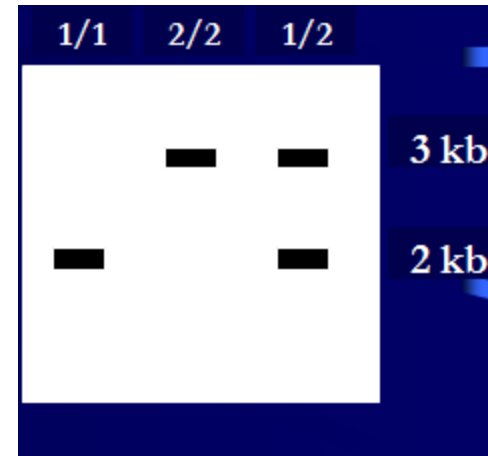
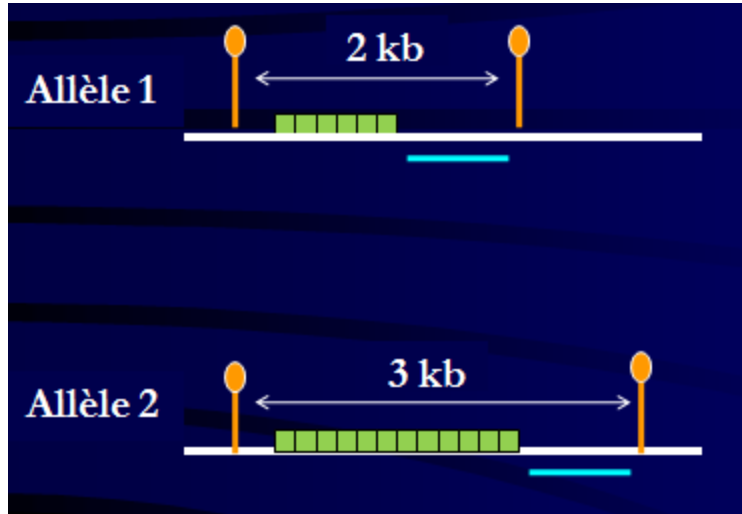


## 3.2. Les minisatellites

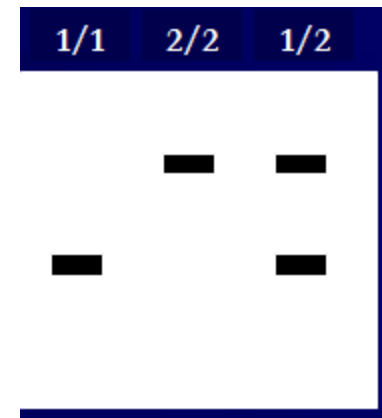
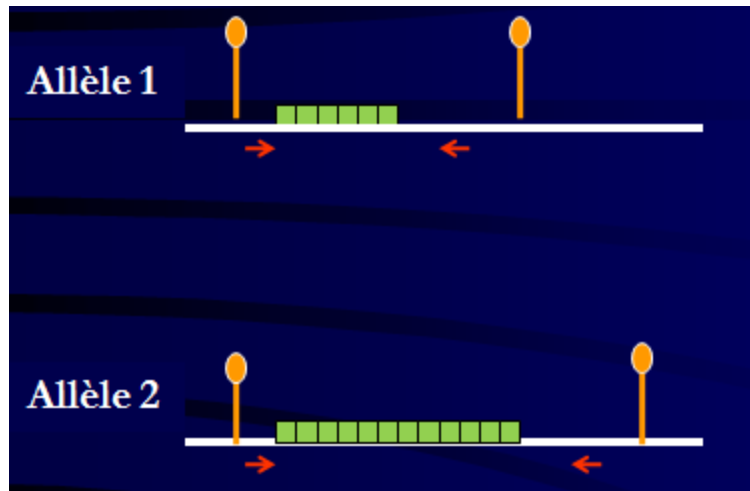
### Les séquences répétées en tandem ou minisatellites (VNTR)

- **Poly alléliques:** grand nombre d'allèles différents à un locus donné (en fonction du nombre de répétitions de la séquence) .
- **Multilocus:** une même séquence de minisatellite va se trouver à différents endroits dans le génome
- peuvent être détectés par RFLP en utilisant des enzymes de restriction qui coupent un grand nombre de fois le génome.
- 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.

- Southern blot

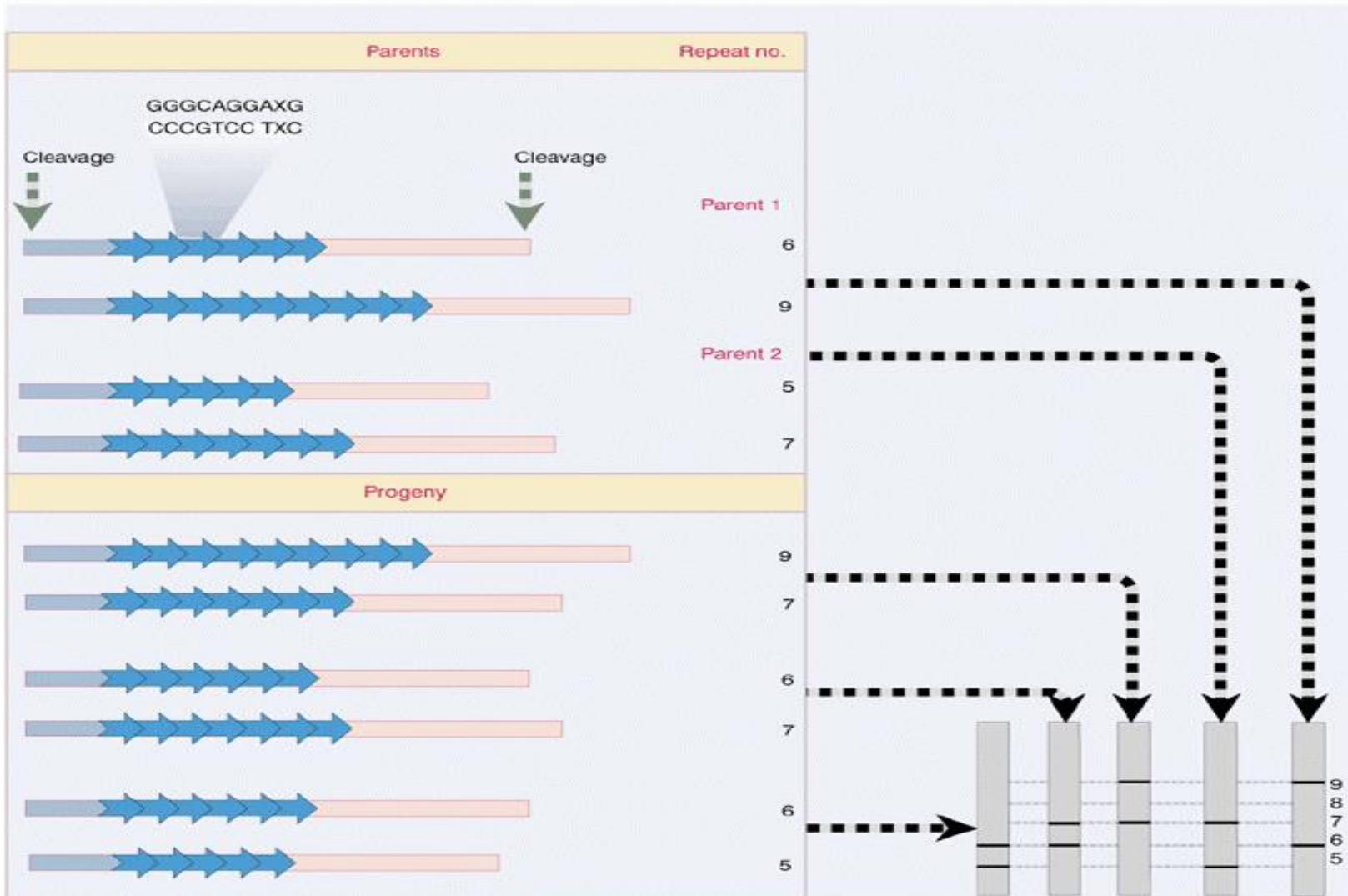


- PCR



- Southern blot

**Figure 4.22** Alleles may differ in the number of repeats at a minisatellite locus, so that cleavage on either side generates restriction fragments that differ in length. By using a minisatellite with alleles that differ between parents, the pattern of inheritance can be followed.



## 3.3. Les microsatellites:

STR : Short Tandem Repeats.  
répétitions de séquences simples

- Les microsatellites sont constitués de séquences de **di-**, **tri-** ou **tétra** nucléotides répétés en **tandem**. Ces éléments sont uniformément répartis en plusieurs exemplaires sur l'ensemble du génome d'une espèce et présentent un taux de polymorphisme élevé.
- Il s'agit donc ici aussi d'un polymorphisme de la séquence d'ADN où les différents **allèles** au **locus** considéré correspondent aux différents **nombre**s de **répétitions** possibles.
- Contrairement aux polymorphismes de restriction qui sont nécessairement bialléliques, les microsatellites sont le plus souvent **multialléliques**.
- Leur détection est obtenue par **PCR** du **locus** contenant **la séquence répétée** suivie d'une électrophorèse du produit de **PCR** qui permet de distinguer les allèles en fonction de leur taille.

# 3.3. Les microsatellites:

STR : Short Tandem Repeats.  
répétitions de séquences simples

nnnnnnCACACACACACACACACACACACAnnnnnnn



nb variable de répétitions

nnnnnnGTTGTTGTTGTTGTTGTTGTTGTTGTTnnnnnnnn



nb variable de répétitions

nnnnnnATGGATGGATGGATGGATGGATGGATGGnnnnnn



nb variable de répétitions

Allèle 1

9 répétitions (CA)

...nnnnnnnnnnCACACACACACACACACAnnnnnnnnnnn...  
...nnnnnnnnnnCACACACACACACACACAnnnnnnnnnnn...

Allèle 2

13 répétitions (CA)

...nnnnnnnnnnnnCACACACACACACACACACACACAnnnnnnnnn...  
...nnnnnnnnnnnnCACACACACACACACACACACACAnnnnnnnnn...

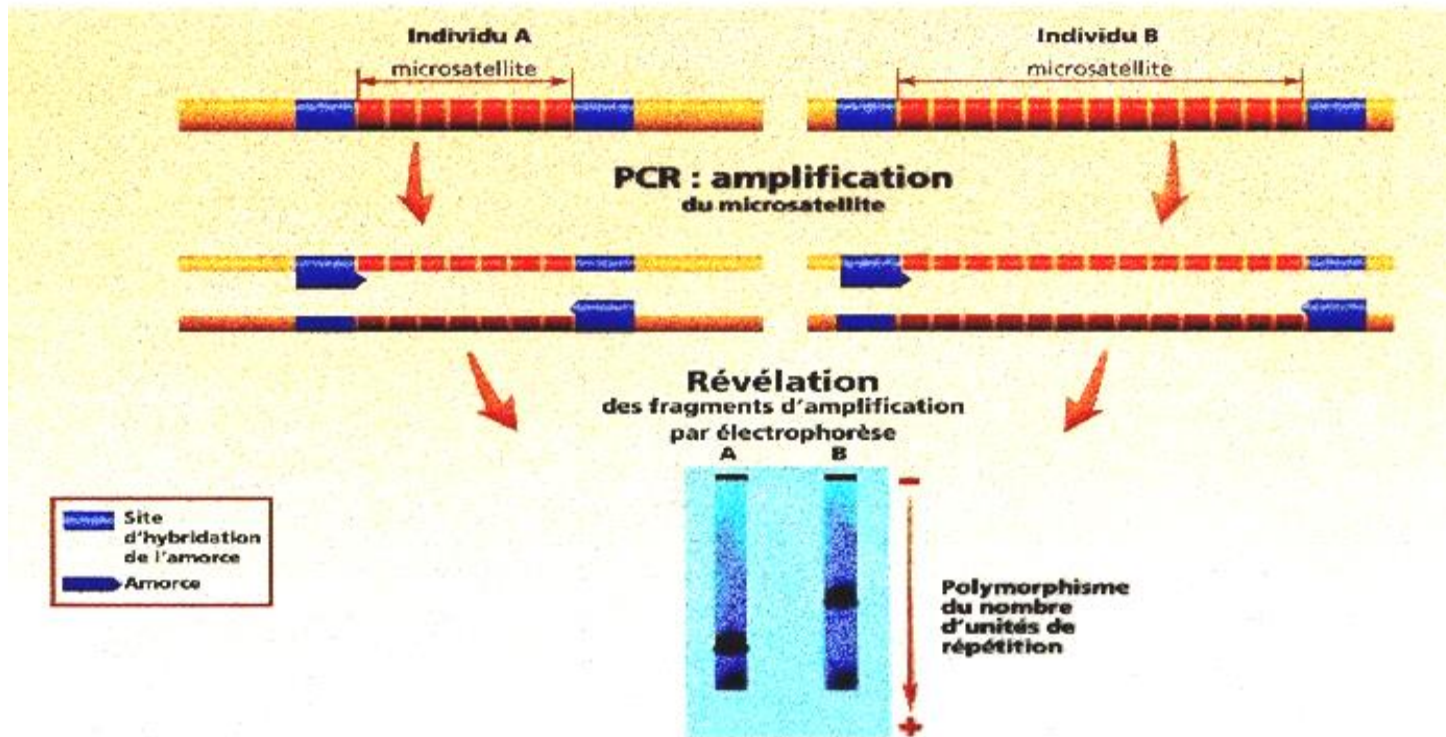
Allèle 3

11 répétitions (CA)

...nnnnnnnnnnnnCACACACACACACACACACACAnnnnnnnnnnn...  
...nnnnnnnnnnnnCACACACACACACACACACACAnnnnnnnnnnn...

- **Les étapes de la technique**
- C'est la technique de **PCR** qui est utilisée pour révéler le polymorphisme des microsatellites.
- **Les séquences flanquant** ces éléments répétés permettent de définir les **amorces** qui seront utilisées pour amplifier le même microsatellite chez différents individus.
- En effet, chaque microsatellite est bordé par des séquences uniques qui lui sont propres. Les fragments d'amplification sont ensuite révélés par électrophorèse.
- Un individu B, possédant plus d'unités de répétition que A, a un produit d'amplification qui migre plus lentement que A.
- C'est **la paire d'amorces** spécifiques des bordures droite et gauche du microsatellite qui constitue le marqueur.

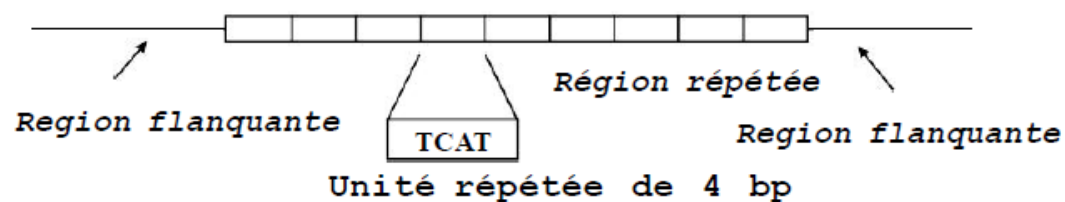
# Les marqueurs microsatellites



2

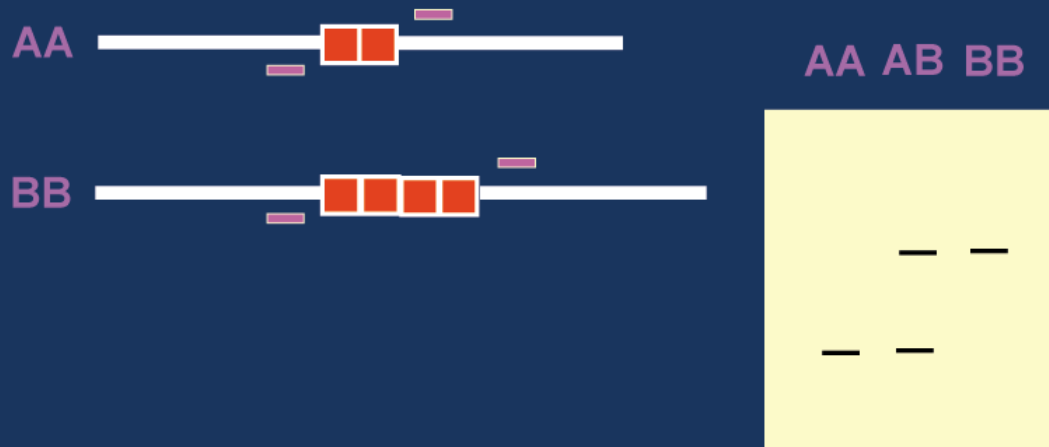
**Marqueur = paire d'amorces spécifiques bordant le microsatellite**

## Les microsatellites ou SSR ou STR

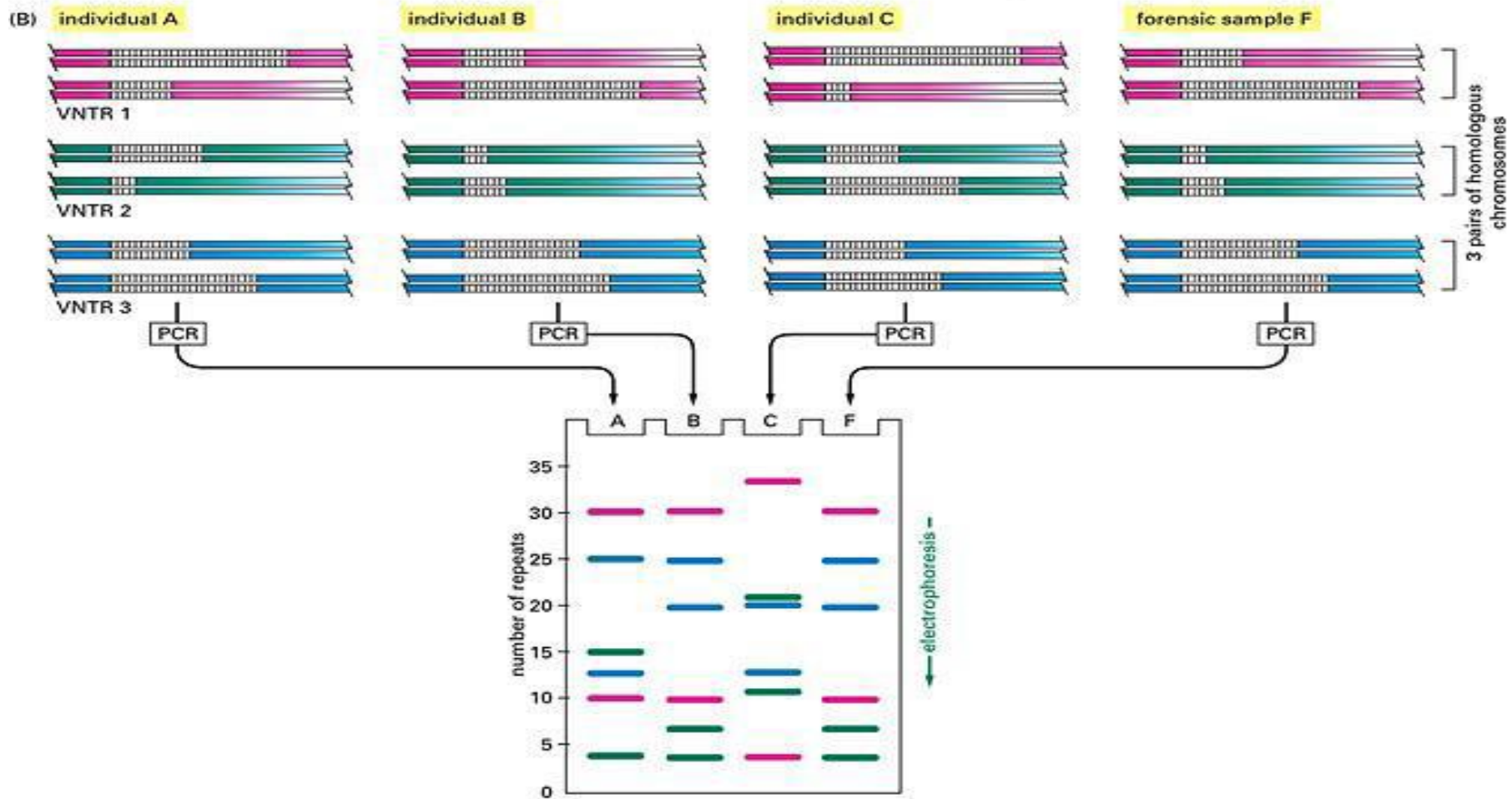
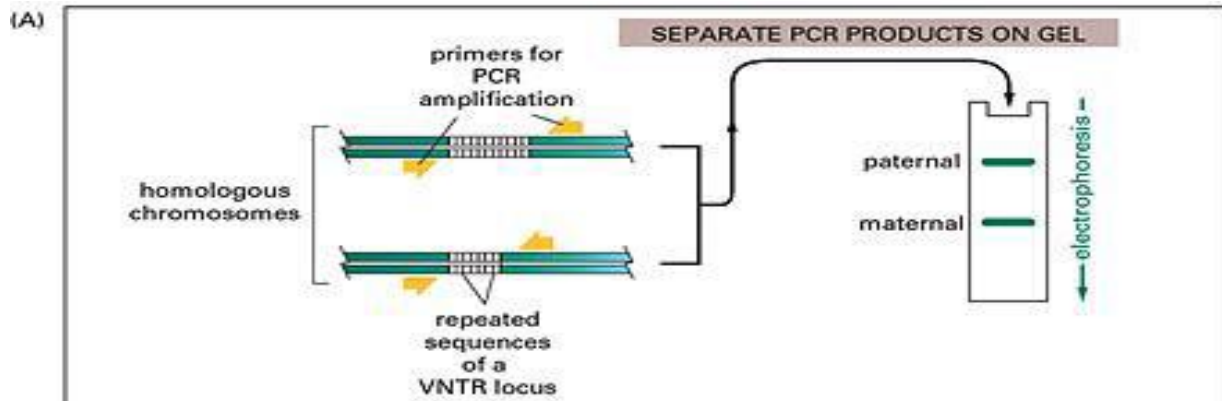




# VNTR/SSR(STR)



- La variation de taille des produits PCR est causée par des différences dans le **nombre d'unités de répétition du microsatellite**.
- Les polymorphismes STR peuvent être visualisés par **électrophorèse sur gel d'agarose ou de polyacrylamide**.
- **Les allèles microsatellites** peuvent être détectés en utilisant diverses méthodes: bromure d'éthidium, coloration à l'argent, radioisotopes ou fluorescence.

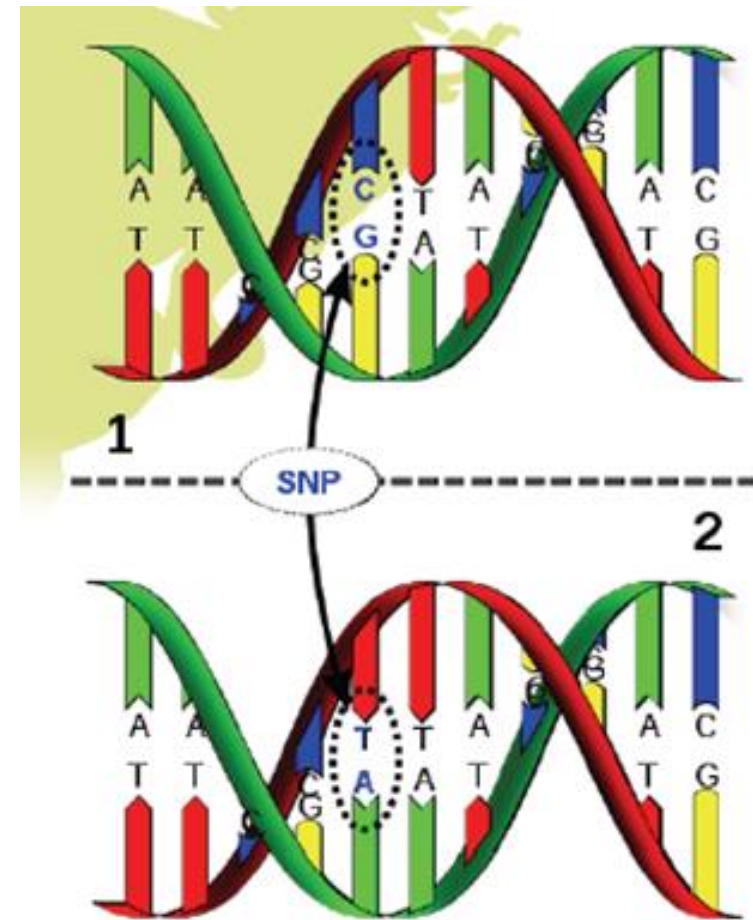




# Autre types de polymorphisme

## 3.4 SNP (single nucleotide polymorphism)

- Un polymorphisme SNP est défini comme étant la **différence d'une seule paire de bases**, au niveau d'un **locus** précis du génome, entre deux individus de la même espèce
- Ces SNP sont distribués uniformément dans tout le génome ( $5 \cdot 10^6$  à  $10^7$  dans le génome humain) et représentent 90% du polymorphisme génétique entre individus.
- Si un SNP modifie un **site de restriction**, il génère un **RFLP**.



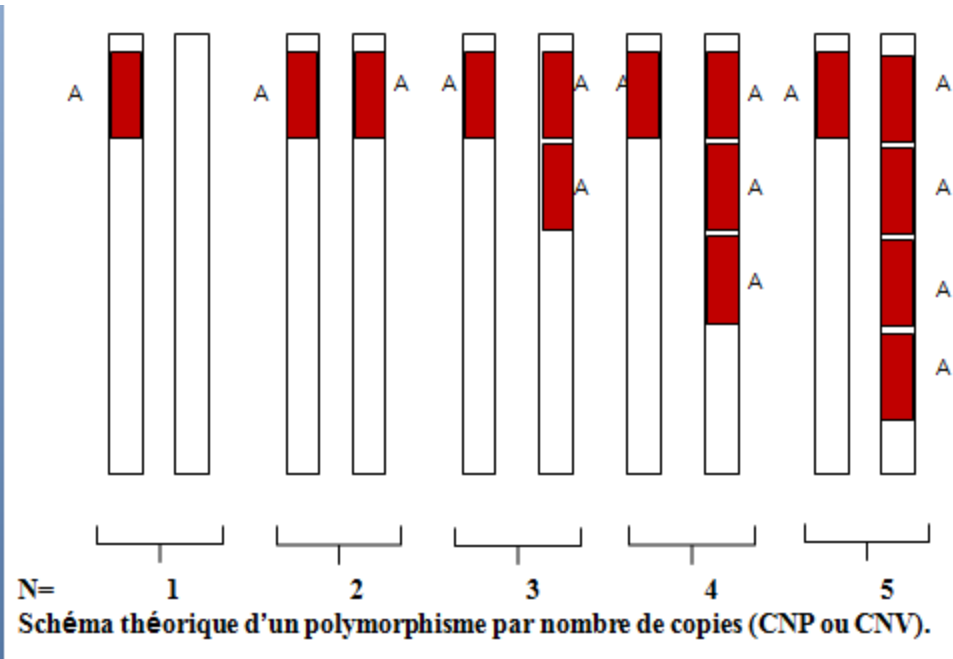
- C'est un polymorphisme bi-allélique qui survient toutes les 1000 pb
- Il est caractérisé par sa localisation précise
- On pense actuellement que les SNP pourraient être impliqués dans le développement de maladies complexes ex : HTA, diabète...

# 3.5. les CNV:

## Copy Number Variations

- Il s'agit de **la variation** du **nombre de copies** de **segments d'ADN de grande taille**, de plus de **1 kb** pouvant ou non contenir un ou plusieurs gènes. Ce type de polymorphisme, identifié récemment (2005), est détectable par des méthodes récentes utilisant les **puces**.
- Ils peuvent être sous forme de **délétions, insertions, duplication, translocation** ou de **variants complexes**. D'autres types de variations génomiques sont inclus parmi les variations structurales. Notamment les inversions qui contribuent aussi à l'instabilité génomique malgré leur nombre de copies balancé.
- Pour l'instant les **CNV** sont peu utilisés dans les applications diagnostiques, mais leur implication dans certaines pathologies, notamment multifactorielles, les rend de plus en plus intéressants et donc de plus en plus étudiés.
- Comme les séquences microsatellites, les CNV sont souvent **multialléliques**, chaque allèle étant représenté par un nombre de copies du segment d'ADN.

# 3.5. les CNV: Copy Number Variations





# IV. Applications des marqueurs généétiques:

- Les marqueurs génotypiques tels que: les RFLP, et les microsatellites sont des outils précieux pour l'analyse du génome humain, ils sont utilisés pour:
- Cartographie et clonage positionnel.
- La recherche de gènes nouveaux.
- Le diagnostic des maladies génétiques monofactotielles
- La mise en évidence d'empreinte parentale.
- L'analyse des anomalies des disjonctions chromosomiques.
- L'identification des individus

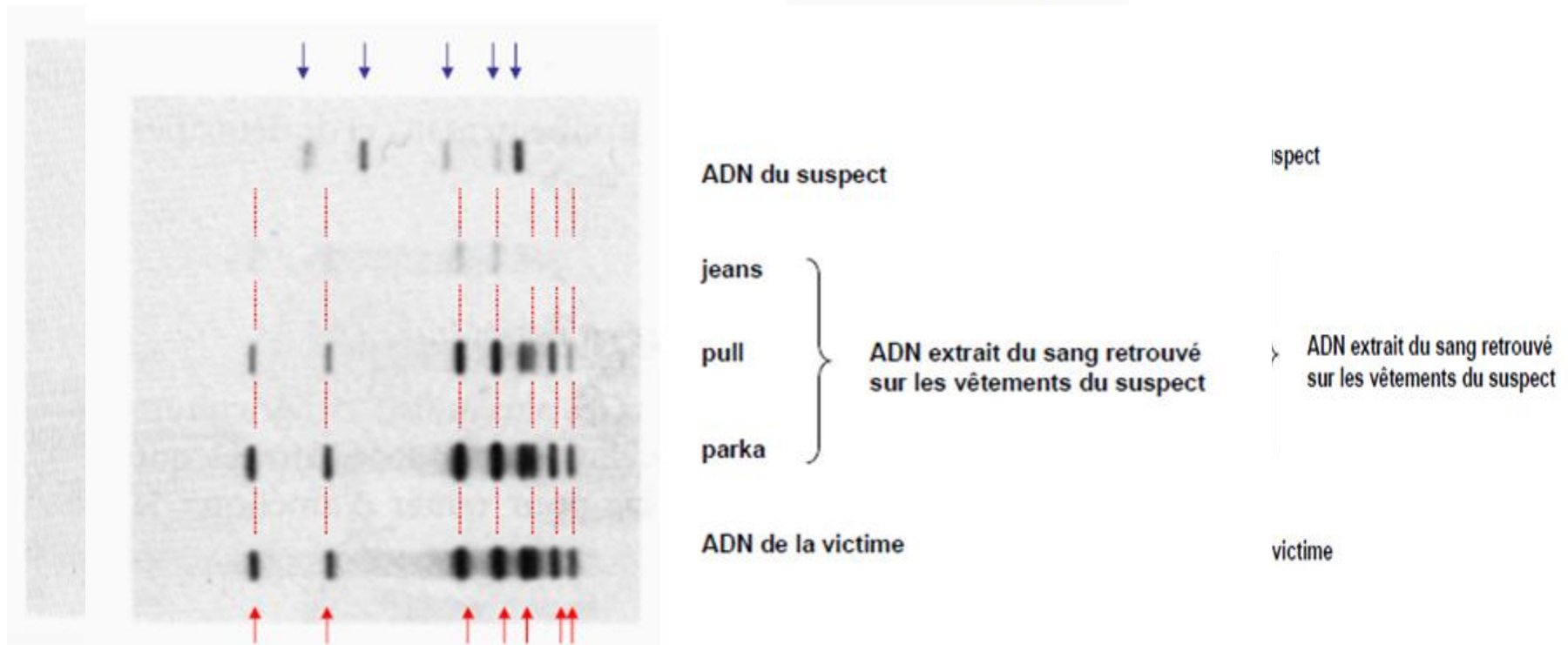
# Applications médico-légales:

## 1) Principe

- Les tests **d'identité** ou de **paternité** utilisent l'information apportée par **les microsatellites** qui présentent de nombreux allèles dans la population et sont donc variables d'un individu à l'autre.
- De plus, **les microsatellites** sont très nombreux et distribués de façon régulière dans le génome, si bien que la probabilité de trouver deux individus portant sur l'ensemble de ces loci les mêmes génotypes est quasi-nulle, ce qui revient à dire que l'ensemble des génotypes d'un individu est unique et constitue sa “ **signature** ” individuelle ou sa “ **carte d'identité** ” génétique.

## 2) Le test d'identité

Le test d'identité consiste à calculer la probabilité que deux prélèvements distincts (par exemple une tache de sang trouvée sur les lieux d'un crime et le prélèvement de sang d'un suspect) soient issus d'un même individu en comparant les allèles qu'ils portent au niveau de différents marqueurs microsatellites. Cette comparaison peut se faire facilement par une simple électrophorèse de l'ADN après amplification par PCR du locus microsatellite.



### 3) la recherche de paternité:

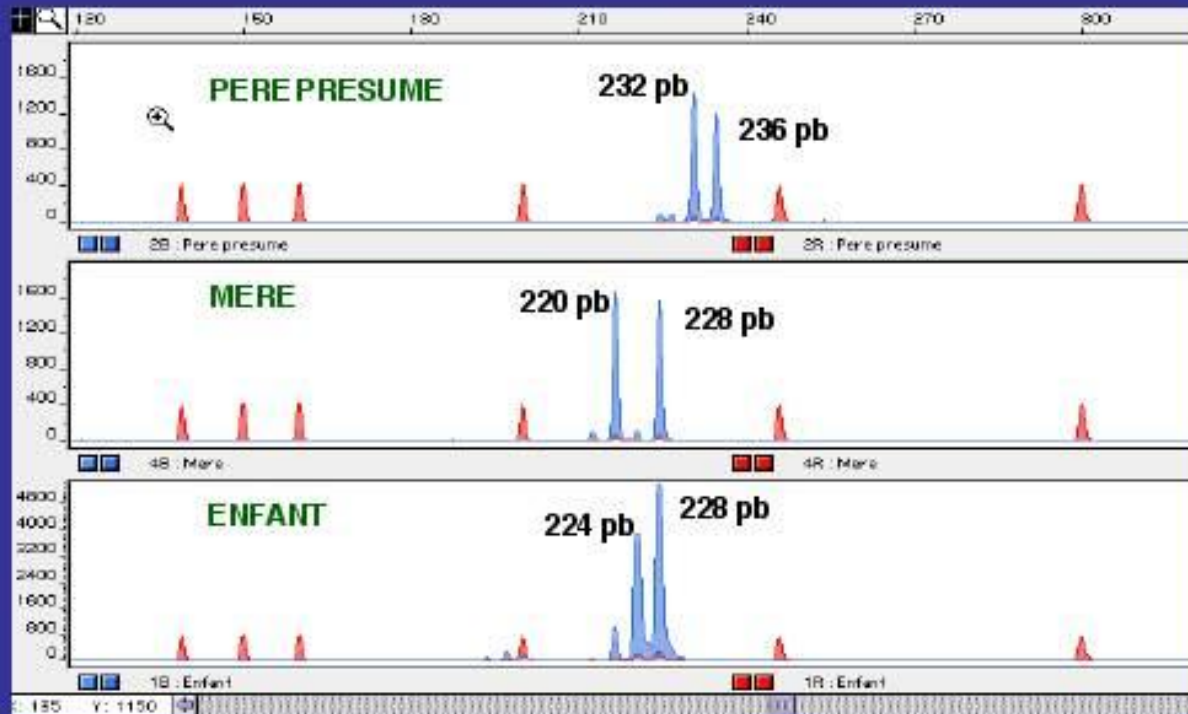
- Le test de paternité consiste à calculer la probabilité que l'enfant et le père présumé portent un allèle en commun par le fait du hasard et non par filiation.
- Dans les deux cas, le calcul repose sur les fréquences des différents allèles du marqueur microsatellite dans la population. L'utilisation combinée de plusieurs marqueurs permet d'obtenir des risques d'erreurs très faibles (de l'ordre de  $10^{-6}$  à  $10^{-9}$ ).

## Test de paternité par PCR-RFLP

L'enfant 2 est probablement illégitime !



## RECHERCHE DE PATERNITE



Exemple d'exclusion de paternité un locus microsatellite de type (CA)<sub>n</sub> est amplifié par PCR pour chacun des individus testés. Chaque produit de PCR est ensuite analysé sur un séquenceur automatique. L'un des oligonucléotides ayant servi à la PCR est marqué par fluorescence, ce qui rend le produit de PCR détectable sur un séquenceur automatique (pics de couleur bleue). Les pics de couleur rouge correspondent à un marqueur de poids moléculaire. Un fragment de taille 232 pb correspond à 12 répétitions du motif CA.

Génotypes : père présumé 12 / 14 ; mère 6 / 10 ; enfant 8 / 10