

# UE11 - Méthodes d'étude et d'analyse du génome



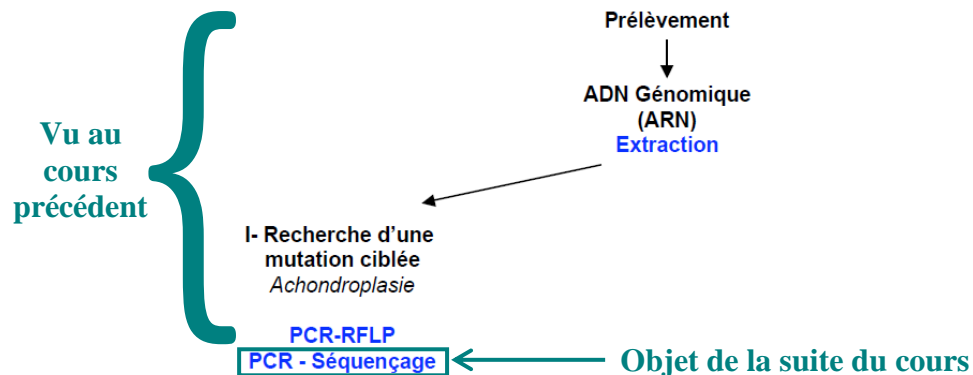
# Principales techniques de Biologie Moléculaire 2

## I. Introduction

Nous avons vu la dernière fois comment **diagnostiquer l'achondroplasie** chez un fœtus en identifiant une **mutation ciblée** grâce à une **digestion par des enzymes de restriction** suivie d'une **migration électrophorétique** (= **PCR-RFLP**) méthode **indirecte** car on ne séquençait pas ladite mutation.

On peut vérifier la présence de notre mutation par **séquençage**, méthode cette fois-ci **directe** car on lit directement la séquence nucléotidique du patient. En réalité en biologie moléculaire, on utilise toujours dans la mesure du possible **deux techniques** pour faire un « double contrôle » afin de **confirmer la présence de la mutation** que l'on recherche !

Ainsi, **pour l'achondroplasie**, nous venons de faire une **digestion par les enzymes de restriction**, suivie d'une **électrophorèse** (l'ensemble forme la **PCR-RFLP**) mais il va falloir que nous fassions une **vérification par séquençage**, autre technique de biologie moléculaire que nous allons détailler.



Nous avons déjà vu à partir de quels **types de prélèvements** nous pouvons travailler, comment en **extraire du matériel génétique** (ADN/ARN), et comment le **manipuler** pour identifier une **mutation ciblée** avec **PCR-RFLP**.

## II. Séquençage de l'ADN

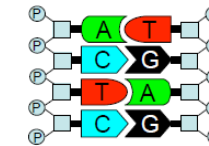
### A. Principe du séquençage

Le **séquençage** de l'ADN a pour but de déterminer la succession des nucléotides qui le compose. Il nous permet de **lire la séquence** d'ADN et de connaître la position de chaque nucléotide de ce variant.

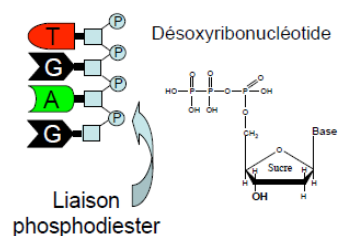
On peut **identifier** une mutation et **diagnostiquer** une éventuelle pathologie en séquençant directement la séquence du **gène concerné**.



La double hélice d'ADN



2 Brins



Liaison phosphodiester

**Rappel :** les deux brins d'ADN sont constitués de **dNTPs** reliés à leur nucléotide complémentaire sur **l'autre brin** par des **liaisons hydrogène**, et relié entre eux sur un **même brin** par des **liaisons de type phosphodiester**.

Nous allons voir le principe de fonctionnement de **LA méthode de référence** : la méthode de **Sanger** aussi appelée **méthode enzymatique des didésoxyribonucléotides (ddNTPs)** !

**L'ADN Polymérase synthétise, à partir d'une amorce, un brin complémentaire fidèle à la séquence d'ADN à étudier**

Dans le prochain cours vous verrez le **NGS** (Next Generation Sequencing) aussi appelé **Séquençage Haut Débit**, une nouvelle technique de séquençage beaucoup plus **puissante**.

Elle permet en effet de **séquencer le génome entier** et d'obtenir la séquence complète de tout le génome d'un individu en quelques jours.

La méthode de **Sanger** est en comparaison très **restreinte** : elle permet de séquencer des fragments d'environ **800 pb** mais pas davantage.

Elle reste cependant **LA technique de référence** car on n'a pas assez de recul pour n'utiliser que le NGS !

### Un NGS doit toujours être confirmé par un Sanger qui est LA méthode de référence

On a donc un fragment d'ADN dont on ne connaît pas la succession de nucléotides que l'on veut **séquencer**. Afin de déterminer cette séquence, on utilise une petite **amorce** de quelques nucléotides (entre 15 et 25), **simple brin** et **complémentaire** d'une partie de la **séquence connue** de notre fragment. A la différence de la PCR, le séquençage ne nécessite qu'**une seule amorce/primer** à partir duquel l'**ADN polymérase** synthétise un brin complémentaire **fidèle** à la séquence d'ADN à étudier, dans le **sens 5'-3'** !

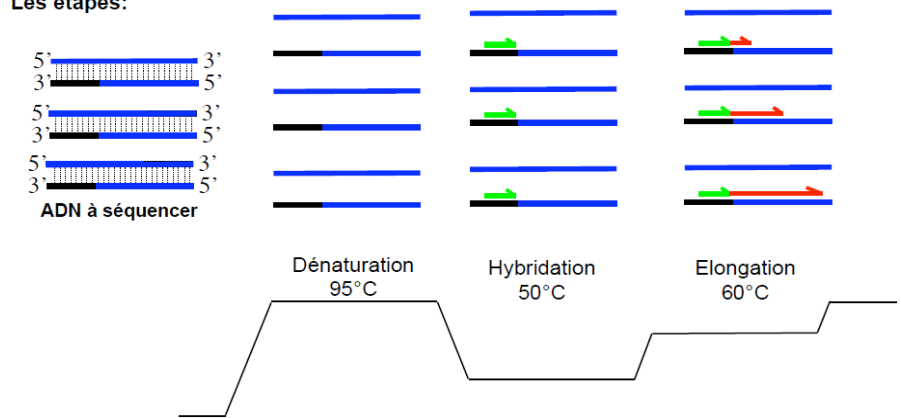


### Les étapes du séquençage sont les mêmes que celles de la PCR

On retrouve une succession de **30 à 35 cycles** successifs de **3 étapes** chacun :

- 1. Dénaturation (95°C)** : l'ADN double brin est **déstabilisé** et rendu **simple brin** par rupture des liaisons **hydrogènes** sous l'effet de la **chaleur**.
- 2. Hybridation (50°C)** : le **primer** s'hybride par **complémentarité** au brin.
- 3. Elongation (60°C)** : l'**ADN polymérase** synthétise le brin complémentaire.

Les étapes:



Lors du **séquençage**, on retrouve dans le **milieu réactionnel** l'ADN **polymérase**, les **désoxyribonucléotides (dNTPs)** ainsi que des nucléotides particuliers : les **didésoxyribonucléotides (ddNTPs)**.

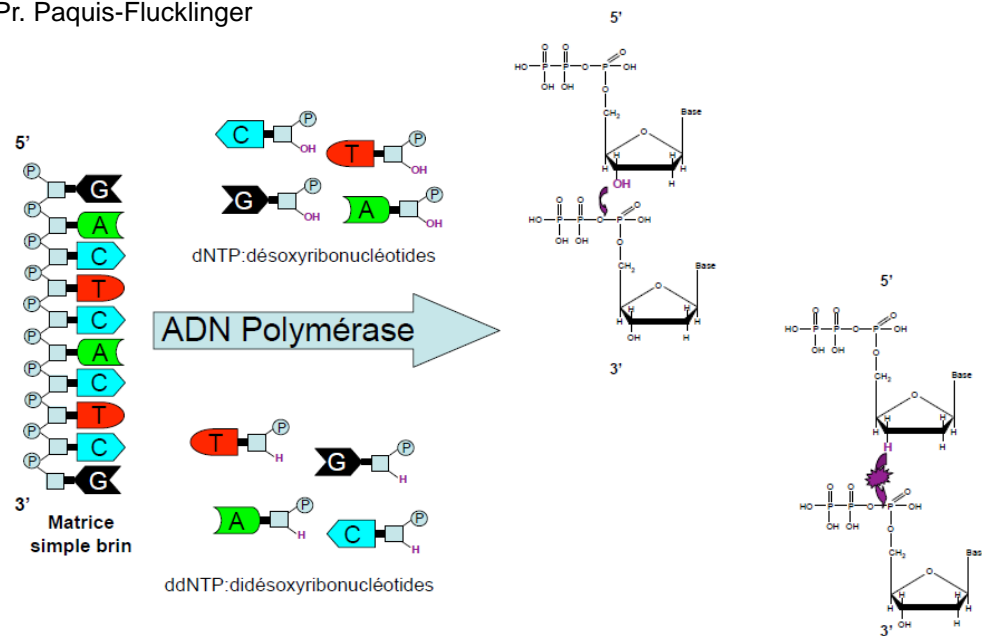
Un **2'-désoxyribose** est un ribose désoxygéné sur son carbone 2'. Maintenant, si on **enlève un deuxième oxygène** sur ce ribose, au niveau de son carbone 3', on obtient un didésoxyribonucléotide (un nucléotide contenant un ribose **désoxygéné en 2' et en 3'**).

Au lieu d'avoir un **groupement 3'-OH**, ces ddNTPs qui ont perdu leur oxygène n'ont plus qu'une **extrémité 3'-H** !

Or une liaison phosphodiester unissant deux nucléotides d'un même brin se fait entre une **extrémité 5'-P** et une **extrémité 3'-OH**. L'absence de l'oxygène **empêchera** donc la polymérase de former une **liaison 3'-5' phosphodiester** avec l'**extrémité 3'** d'un ddNTP !

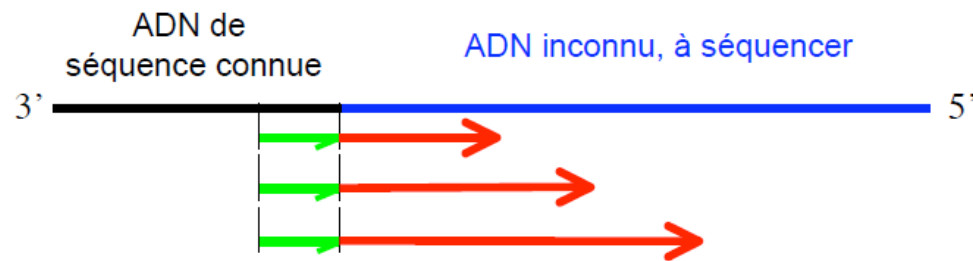
### L'incorporation d'un ddNTP stoppe la synthèse du brin !

La **polymérase** synthétise le brin complémentaire en incorporant **au hasard** des **dNTPs** ou des **ddNTPs**. Si c'est un **dNTP**, la liaison phosphodiester avec le nucléotide suivant **va pouvoir se faire** grâce à la présence du **groupement -OH**. Par contre, si c'est un **ddNTP** la liaison avec le prochain nucléotide est **impossible** : la polymérase arrête la synthèse de ce brin car elle ne peut plus ajouter d'autres nucléotides.



On répète ces étapes de « dénaturation/hybridation/élongation » **pendant x cycles**, de telle sorte qu'une fois la **réaction terminée**, on obtient un mélange de fragments d'ADN générés de **toutes les tailles possibles** !

On aura ainsi des brins ayant **incorporé un ddNTP en tout premier** et qui ne seront donc composés que d'**un seul nucléotide**, d'autre qui seront composés de **deux nucléotides**, d'autres de **trois**, d'autres encore de **quatre**... Et ainsi de suite jusqu'à ce que l'on obtienne selon toute probabilité des fragments ayant été **synthétisés entièrement** et n'ayant incorporé un ddNTP qu'à la **toute fin de la synthèse**, pour le tout **dernier nucléotide** de la séquence !



La synthèse s'est arrêtée à **tous les stades** si l'on considère l'ensemble des brins générés, de sorte que **tous les nucléotides** de la séquence **représentent la terminaison** (le ddNTP) **d'un brin** (vu qu'ils ont tous des **tailles différentes**).

## B. Méthode de Sanger (1977)

Historiquement, on pouvait réaliser **manuellement** ce type d'expérience, qui était relativement **long** et **peu précis** par rapport au NGS actuel.

On faisait à l'origine **quatre réactions indépendantes** dans quatre tubes différents contenant chacun **un seul ddNTP** (A, T, C ou G). On mettait donc **simultanément** dans **quatre tubes** le fragment d'ADN du patient à séquencer, une seule amorce, une ADN polymérase, les **4 types de dNTP** et **un type de ddNTP radiomarké**.

**Remarque** : il fallait faire attention à ne pas confondre les tubes !

Dans chacun des quatre tubes, on **répétait plusieurs fois** les étapes de « dénaturation/hybridation/élongation » et on obtenait à la fin un **mélange de produits** générés de **différentes tailles**.

**Ex** : si dans le 1<sup>er</sup> tube on met un **ddNTP** correspondant à une **Thymine**, **tous les fragments** de ce tube **finiront par une thymine radiomarkée**, puisque la seule façon de **bloquer la synthèse** est d'incorporer un **ddTTP**, les fragments se termineront donc obligatoirement par une thymine !

**Chaque tube contient un seul type de ddNTP (A, T, C OU G) ET les quatre types de dNTP (A, T, C ET G)**

On mettait ensuite ces quatre tubes à **migrer sur un gel d'acrylamide**, pour **séparer** les produits synthétisés en fonction de leur **taille** par **migration électrophorétique**. Grâce au champ électrique, les fragments les plus petits **migrent le plus loin du - vers le +**. Les fragments les plus **gros** et donc les plus **lourds** restent **en haut** du gel et **migrent donc le moins loin**.

Pour **déterminer la séquence** du fragment, il faut lire le gel **de bas en haut**, c'est-à-dire du plus **petit** au plus **grand** des fragments !

- **du plus petit** : les fragments ont incorporé un ddNTP rapidement, ils ont donc stoppé leur synthèse plus tôt, sont plus légers et migrent loin.

- **au plus grand** : les fragments ont incorporé un ddNTP plus tardivement, ils ont donc stoppé leur synthèse plus tard, sont plus lourds et migrent moins loin.

L'**ordre de lecture** est déterminé par la **taille des fragments** et la **nature des nucléotides terminateurs de chaîne** (ddNTP) de la piste de migration en question !

Pour faire simple, on **commence la lecture** de la séquence en **partant du bas** du gel et en **remontant**, on a une colonne, **un puits par type de ddNTP** (il y a donc **quatre puits** au total, un où tous les brins se terminent par une thymine, le suivant où ils se terminent tous par une adénine, celui d'après où c'est une guanine, et une cytosine pour le quatrième).

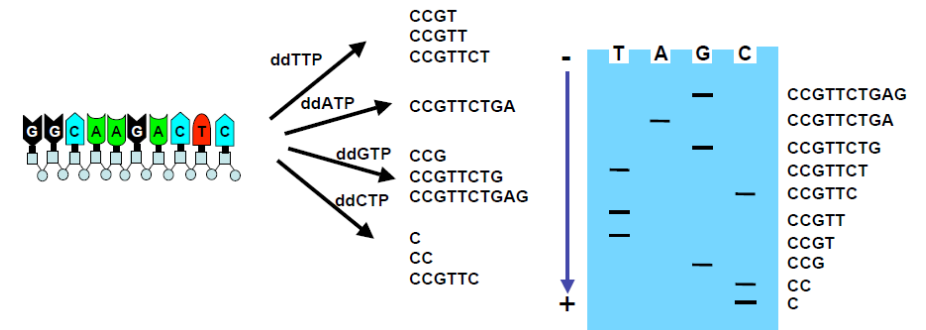
On regarde sur quelle piste est le fragment qui a **migré le plus loin**, cela signifie qu'il a **incorporé un ddNTP dès le départ** et que le ddNTP qu'il a incorporé correspond au **tout premier nucléotide de la séquence** ! Si ce fragment est sur la piste des **ddCTP**, cela signifie que le **premier nucléotide** de la **séquence lue sur le gel** est une **cytosine** !

On répète l'opération avec le fragment qui a incorporé un ddNTP en deuxième position et qui a donc migré un tout petit peu moins loin car il contient un nucléotide de plus (il est un peu plus lourd que le premier). On regarde sur quelle piste il se situe (par exemple s'il est encore sur la piste des **ddCTP**, cela signifie que le second nucléotide de la séquence est encore une **cytosine** ! Les deux premiers nucléotides de la séquence lue sont donc **5'-CC-3'** !

On répète ainsi l'opération jusqu'à remonter tout en haut du gel pour obtenir la séquence complète.

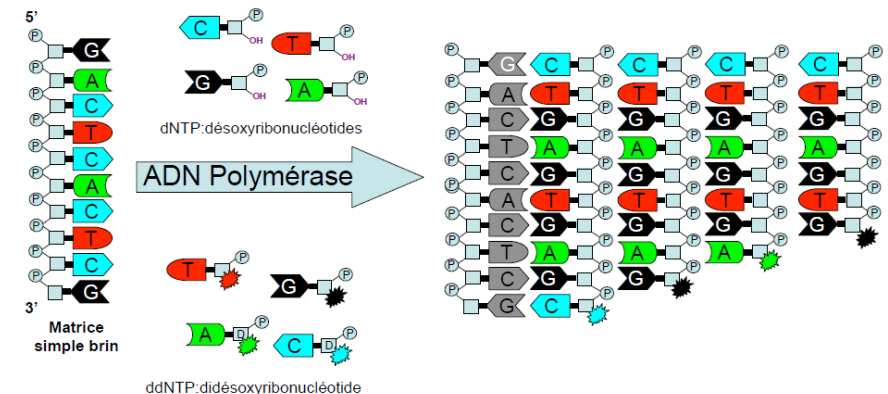
**Remarque :** il ne faut pas confondre la **séquence lue sur le gel** qu'on lit de bas en haut avec la **séquence du brin séquencé** (c'est-à-dire la séquence nucléotidique d'origine du patient chez qui on veut mettre en évidence une mutation). En effet tout se fait par **complémentarité** des bases, ce qui signifie que si **on lit un « C »** sur le gel, le brin à séquencer contient en réalité **un « G »**.

Cette méthode **longue** et **peu efficace** (lecture de 100 à 200 pb pour 2 /3 jours de travail) était tout de même très performante jusque dans les années 90'. Depuis, des sociétés se sont mises à fabriquer des **séquenceurs automatisés**.



### C. Méthode automatisée

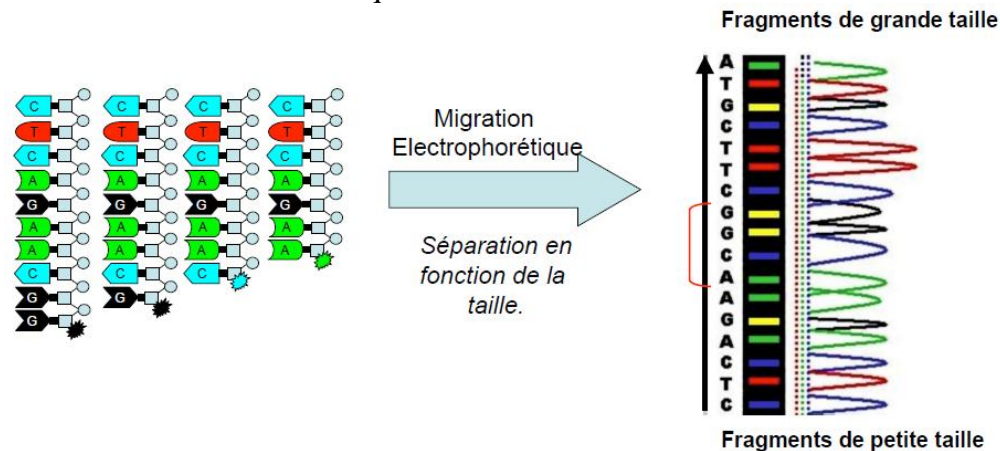
Avec les **séquenceurs automatiques**, on ne fait plus qu'**une seule réaction** au lieu d'avoir quatre tubes différents. Les **quatre ddNTPs** vont pouvoir être **mélangés** dans le même tube réactionnel car chacun est couplé à un **fluorochrome de couleur différente** pour les identifier. D'après la nomenclature, la **Thymine** est en **rouge**, la **Guanine** en **noir**, la **Cytosine** en **bleu** et l'**Adénine** en **vert** ! *(ne pas retenir les couleurs)*



Les fragments d'ADN de tailles différentes générés sont tous marqués d'un **fluorochrome** différent selon le ddNTP incorporé par la polymérase. Ces produits sont ensuite **séparés par migration électrophorétique dans un automate** ! Le principe reste le même : les fragments vont migrer dans un petit **capillaire** (= puits) qui contient un polymère (qui remplace l'agarose ou l'acrylamide). Ils seront séparés

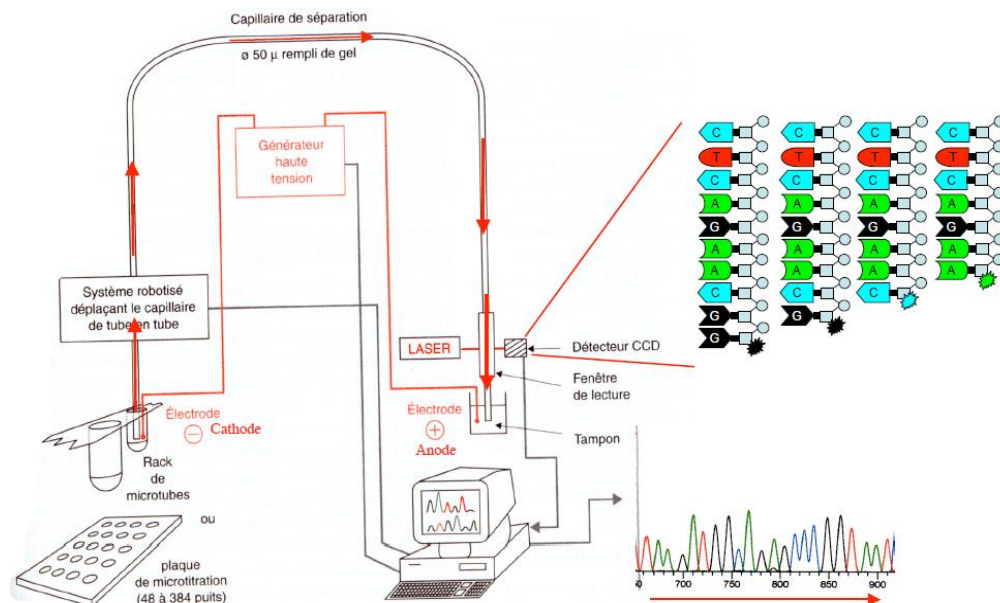


en fonction de leur **taille** ce qui donne un **ordre de lecture**.



Puis les fragments passent devant une **caméra laser** qui va **détecter** et **lire** la couleur de chaque fragment, permettant ainsi d'**identifier le nucléotide** incorporé dans la séquence (ce système est **utilisé en routine** dans les labos). Tout le système est relié à un **ordinateur** qui transmet à la fin de la réaction un **électrophorégramme**.

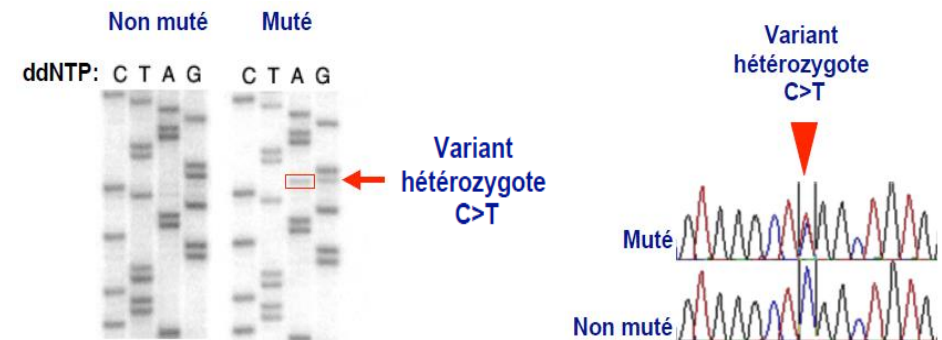
**Remarque :** si on a un séquenceur à **16 capillaires**, on pourra faire migrer simultanément **16 échantillons différents**.



Nom\_de\_Zeus!

Le Tutorat est gratuit. Toute vente ou reproduction est interdite.

## D. Exemples de résultats



### En séquençage Sanger (à gauche) :

On lit le gel de bas en haut : **5'-ACTTCTT...-3'** pour l'**électrophorèse de gauche** (non muté). On remarque ici qu'au niveau de la flèche, sur le gel du **patient muté**, il y a **deux produits de migrations** dans deux puits différents mais de **même taille** : il y a un fragment où un « **A** » a été incorporé, et un autre où c'est un « **G** ».

Il ne faut pas oublier que la séquence du brin lu sur le gel est **complémentaire** du brin séquencé du patient.

Si on lit un « **A** » sur le gel, on a un « **T** » dans la séquence d'origine. De la même façon, un « **G** » lu sur le gel correspond à un « **C** » dans le brin séquencé du patient.

Si on reprend notre exemple, la présence de deux fragments différents de même taille, l'un se terminant par un « **A** » et l'autre par un « **G** » témoigne d'un **variant hétérozygote C>T** ! (car C s'apparie avec G et T s'apparie avec A).

### En séquençage automatique (à droite) :

On peut également voir ce **variant hétérozygote** : au niveau du triangle on remarque une **superposition de deux pics** de couleur différente au niveau de la même **position nucléotidique** : on met en évidence le même **mélange de « C » et de « T »** !

### III. Recherche de mutations dans un gène

On part donc de notre prélèvement et on travaille à partir des **extractions de l'ADN génomique** du patient ou de l'ARN dans certains cas très particuliers (nous allons voir dans quelles mesures on s'en sert dans la suite du cours).

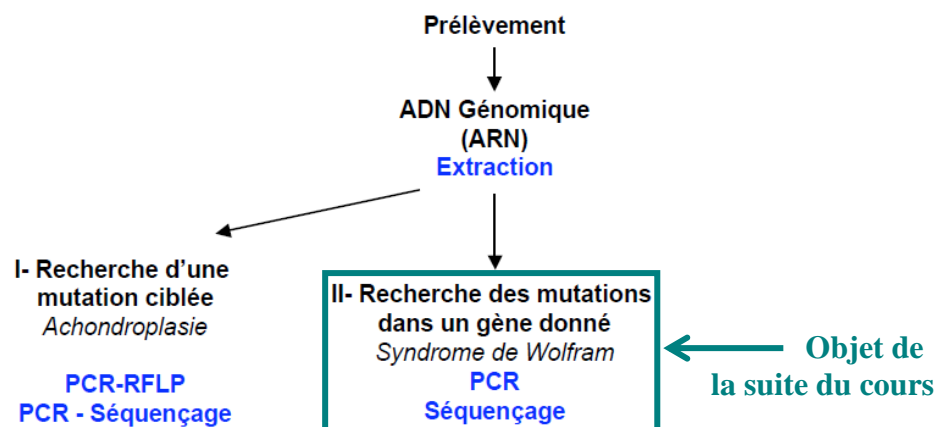
Dans la recherche d'une **mutation ciblée** (cf. cours précédent sur l'achondroplasie), on utilise la **PCR-RFLP** : acronyme anglais signifiant *Restriction Fragment Length Polymorphism*, littéralement « polymorphisme de longueur des fragments de restriction ». Le terme **RFLP** désigne ainsi une **migration électrophorétique** qui est faite une fois que nos fragments d'ADN ont été **digérés par des enzymes de restriction**. On **confirme** ensuite le diagnostic par **PCR - Séquençage**, ce que nous venons de voir.

Nous allons maintenant voir une situation un peu plus compliquée, la recherche de **mutations dans un gène donné**, en screenant le gène au complet.

On utilise toujours la **PCR** suivie d'un **séquençage** (= **PCR - Séquençage**) que l'on va utiliser de façon un peu différente.

Il existe en effet **certain cas** où le **résultat de ce séquençage** est dans un premier temps **ininterprétable** et on aura besoin d'utiliser des **techniques complémentaires** comme le **clonage moléculaire**, pour pouvoir tirer des conclusions et poser un diagnostic !

C'est le cas de **certain patients** atteints du **syndrome de Wolfram**, cas clinique de **pathologie monogénique** que nous allons à présent étudier :



### A. Syndrome de Wolfram

D'un point de vue clinique, ces patients présentent une **atrophie optique**, une **surdité**, des **troubles neurologiques** plus ou moins associés (qui démarrent chez les enfants de façon précoce en général) et du **diabète**.

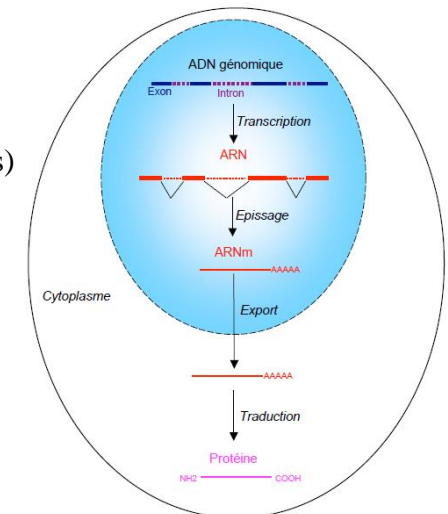
Le syndrome de Wolfram est dû à des mutations dans le **gène WFS1**. C'est un gène constitué de **8 exons**, avec l'**ATG** situé dans le **2<sup>ème</sup> exon**. Ce gène code pour la **wolframine**, une protéine impliquée dans le **flux de calcium de l'organisme** mais dont le rôle exact est inconnu...

**La transmission du syndrome de Wolfram est autosomique récessive**

**Remarque :** l'achondroplasie est une maladie autosomique dominante.

#### Rappel :

- Dans le **noyau**, les gènes sont **transcrits en ARN** (exons + introns)
- Les ARN sont ensuite **maturés en ARNm** après **épissage** (exons)
- Les ARNm sont ensuite **exportés dans le cytoplasme**
- Dans le **cytoplasme**, les ARNm sont **traduits en protéines**



Il existe **deux mutations** du **gène WFS1** :

- **mutation par substitution** (cas le plus simple)
- **présence d'un variant d'épissage** (cas le plus compliqué à l'origine de problèmes de diagnostic nécessitant l'utilisation de techniques complémentaires)

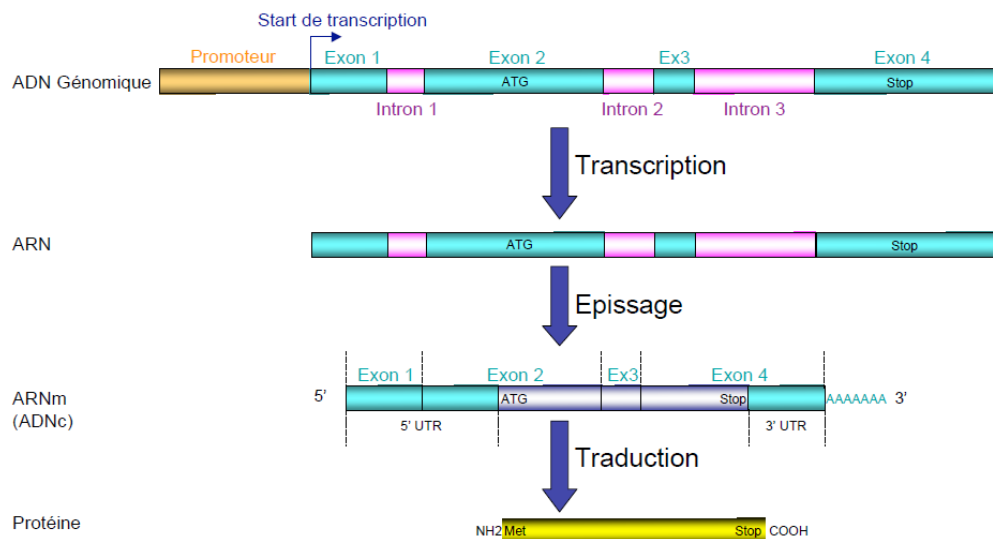
## B. Organisation d'un gène

- Il y a une région **promotrice** (en amont), un codon **START** qui détermine la position du premier exon du gène, puis une succession d'**exons** et d'**introns**.

- La **traduction** de l'ARNm se fait à partir de l'**ATG** mais ce dernier n'est **pas forcément** situé dans les premiers nucléotides de l'**exon 1** !

En **amont** de ce codon il y a la région **5' UTR** et en **aval** du codon **STOP** il y a la région **3' UTR**. Ce sont des régions non codantes impliquées dans la **traduction** ou dans la **stabilité** de l'ARNm.

**Tous les exons ne sont pas forcément codants, on a des régions 5' et 3' non codantes !**



Pour faire un diagnostic, on séquence les **exons**. Nous ne sommes en règle générale pas encore capables d'interpréter des **variants** situés dans les **introns** ou dans les **régions 5'UTR/3'UTR**.

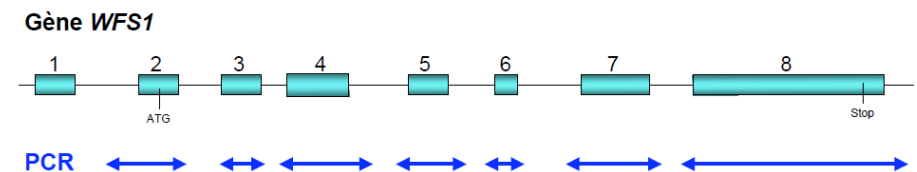
**On ne séquence que les EXONS !**

## C. Recherche de mutations dans un gène

Pour **diagnostiquer** le syndrome de Wolfram, il faut amplifier par PCR et séquencer les **exons** du gène WFS1 et éventuellement les **jonctions intron/exon** qui sont relativement bien caractérisées. On va cibler dans le gène uniquement les régions qui nous intéressent !

Notre stratégie d'exploration pour analyser la succession de nucléotides du gène **WFS1** à la recherche d'une éventuelle mutation consiste à faire **7 PCR** différentes suivies de **7 réactions de séquençage**.

On va donc faire au final **7 PCR – Séquençage** afin d'amplifier les **exons** et les **jonctions intron/exon** du gène.



**7 PCR différentes pour amplifier les exons**

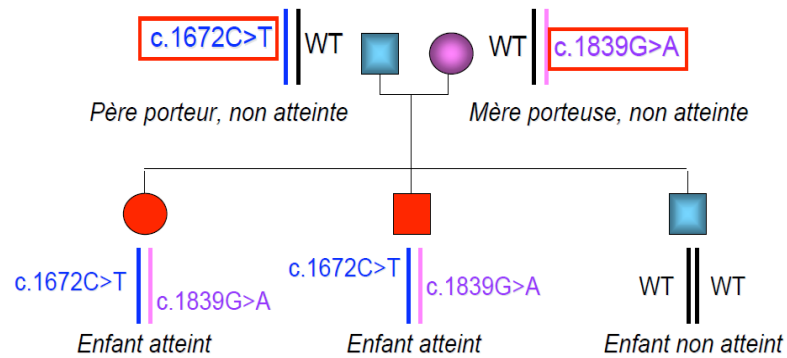
**Séquençage des 7 PCR**

**Remarque :** on ne va screener que les **exons 2 à 8 inclus** et non **pas l'exon 1** car le codon **ATG** étant situé dans l'**exon 2**, l'exon 1 fait partie de la **région 5' UTR non codante qui ne sera pas traduite** !

Le génome humain est **complètement séquencé** aujourd'hui, il est entièrement connu. On compare donc toujours nos résultats à une **séquence de référence**.

**Remarque :** les **mutations** dans les premiers nucléotides d'un **exon** ou d'un **intron** sont susceptibles de **modifier l'épissage** et donc la protéine !

## D. Exemples de résultats : Famille A



**Père** : porteur de la **mutation c.1672C>T** à l'état **hétérozygote**.

**Mère** : porteuse de la **mutation c.1839G>A** à l'état **hétérozygote**.

**Enfants atteints** : ce sont des **hétérozygotes composites**, c'est-à-dire qu'ils ont **deux allèles mutés différents**, deux variants hétérozygotes. Ils produiront donc au final **deux protéines anormales** codées par WFS1 et sont donc **atteints** du syndrome de Wolfram.

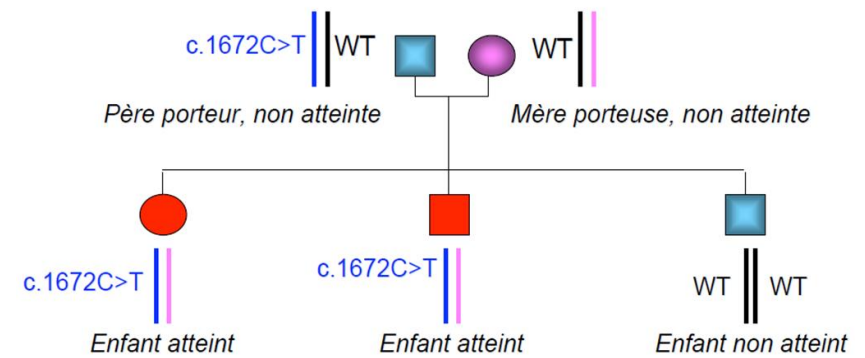
### Diagnostic possible : mutation par substitution dans un EXON !

Dans ce cas précis on a réussi à identifier les variants pathogènes en séquençant les **exons** du gène **WFS1** car les mutations ponctuelles du père et de la mère sont situées dans un exon. On **diagnostique** donc le syndrome de Wolfram !

**Remarque** : On peut aussi vérifier si les deux variants sont en « **cis** » ou en « **trans** ». S'ils sont portés par **chacun des allèles** on dit qu'ils sont en « **trans** », et s'ils sont portés par **le même allèle**, on dit qu'ils sont en « **cis** ». Dans les pathologies **autosomiques récessives**, il est important de connaître la **ségrégation** de ces variants : s'ils sont en « **cis** » on aura **un allèle doublement muté** avec deux variants et **un autre allèle Wild-Type** (sauvage, non muté). Du coup l'enfant ne sera **pas atteint** puisqu'il a **un allèle sauvage** codant pour une **wolframine fonctionnelle** tout à fait normale.

Le fait de vérifier que les deux parents sont tous les deux porteurs de la maladie est indispensable pour poser le diagnostic !

## E. Exemples de résultats : Famille B



**Père** : porteur de la **mutation c.1672C>T** à l'état **hétérozygote**.

**Mère** : a priori **aucune anomalie** n'a été trouvée après séquençage.

Le syndrome de Wolfram étant une pathologie **autosomique récessive**, il faut **deux allèles mutés** pour développer la maladie. Or, les enfants sont atteints, et on n'a identifié que la mutation du père, le séquençage des exons de la mère n'a révélé aucune mutation.

**Remarque** : le séquençage ne permet que de séquencer les **exons**, une mutation survenant dans un **intron** passera **inaperçue** !

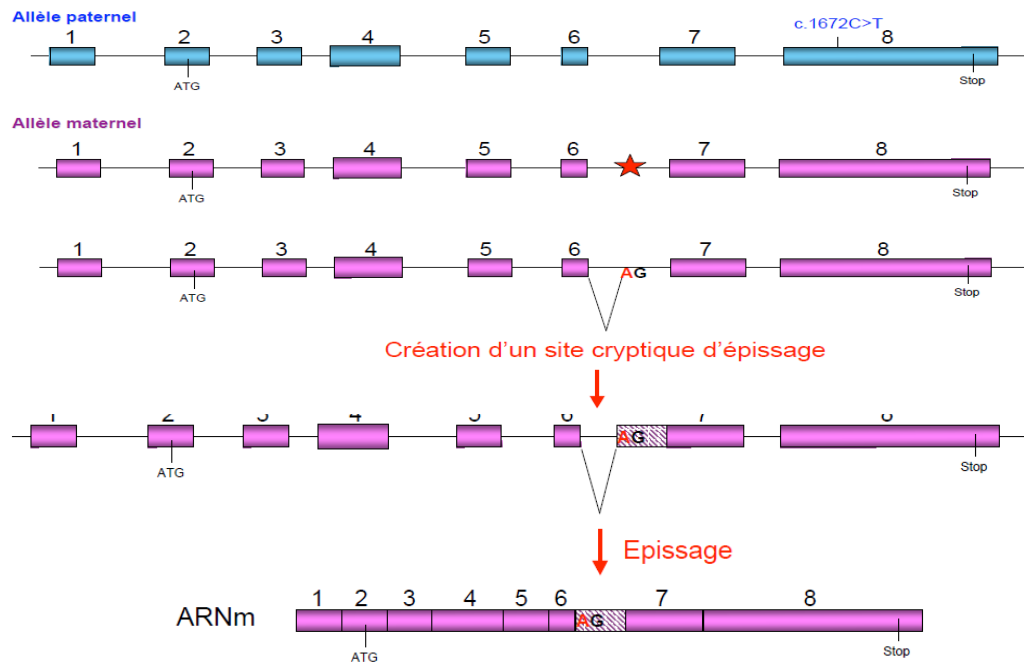
L'**hypothèse** serait donc que le variant est présent dans les **régions introniques** du gène **WFS1**, car en amplifiant et en séquençant uniquement les exons **rien n'a été détecté** !

*Mais alors que s'est-il passé ?* Au milieu de l'intron, il y a eu création d'un **site cryptique d'épissage AG** : une variation de la séquence nucléotidique a créé un **site accepteur**.

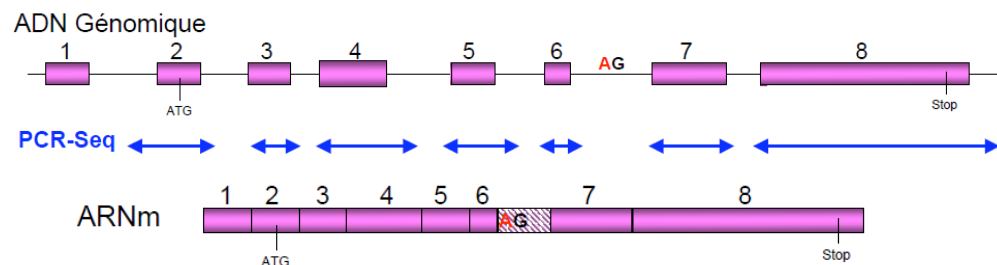
Cette région qui était alors considérée comme un intron va **devenir codante** et modifier la jonction exon 6 / exon 7 de l'ARNm.

Ce dernier va être **rallongé** par une petite séquence, et cela aura forcément des **conséquences** lors de la **traduction**, donnant ainsi une **protéine (wolframine)** différente et anormale, **non fonctionnelle**.





## F. Comment identifier les variants d'épissage ?



Pour identifier le variant d'épissage, on pourrait très bien **séquencer le génome complet** (et non pas que les exons comme on l'a fait jusqu'à présent). Mais le séquençage serait très long et de ce fait peu efficace.

L'autre possibilité est de travailler à partir des **ARNm** afin de séquencer directement ce qui est vraiment **exprimé chez le patient**, ce qui est plus rapide et plus logique. On va donc récupérer des ARNm du patient pour voir s'il a bien un variant d'épissage responsable du syndrome de Wolfram.

## G. Etude des ARN messagers

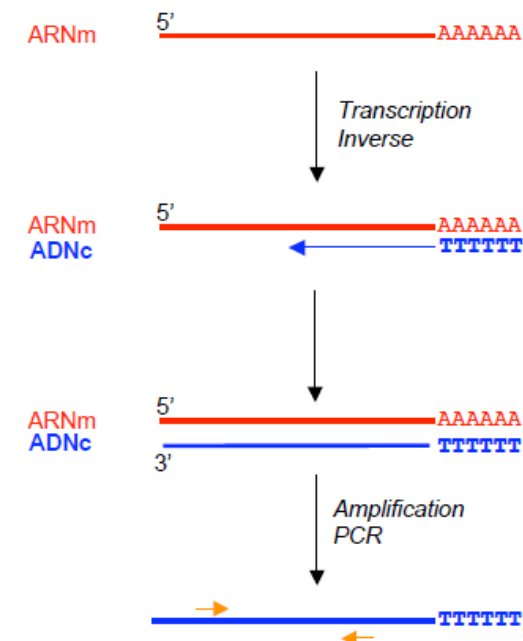
Il faut savoir qu'il est en général plus **difficile** de travailler sur les **ARN** car ils sont plus **instables** que l'ADN et nécessitent beaucoup plus de matériel. Mais dans certains cas comme celui du syndrome de Wolfram, quand on est sûrs qu'il nous **manque une mutation** par exemple, on peut être amené à travailler sur les ARNm et à les séquencer !

Pour isoler les ARNm, on fait une extraction au **phénol-chloroforme**. Mais attention : dans ce cas particulier le phénol doit être **ACIDE** pour permettre de **récupérer** et de **stabiliser** les ARN.

Puis, avant de pouvoir les séquencer il faut d'abord les **amplifier** par **PCR**, pour avoir suffisamment de matériel.

Mais il y a cependant un problème : l'enzyme utilisée dans la PCR est la **Taq Polymérase** (enzyme d'origine **bactérienne**). Or, il s'agit d'une **ADN polymérase**, et comme son nom l'indique, elle ne travaille qu'à partir de **fragments d'ADN** et ne pourra pas amplifier de l'ARN...

Ces ARNm ne peuvent donc pas directement être amplifiés par PCR, ils doivent être d'abord **copiés sous forme d'ADN** !



## 1. Etapes de l'identification des variants d'épissage :

- Mutation au niveau d'un **intron**
- Mutation détectée seulement à partir d'un **ARNm**
- Utilisation d'une **transcriptase inverse** car les ARNm ne peuvent pas être amplifiés directement par PCR
- Copie de l'ARNm sous forme d'ADN

## 2. Transcriptase inverse :

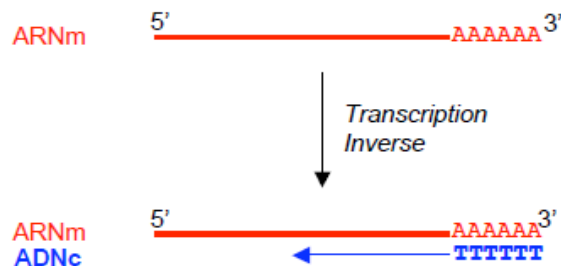
La **transcriptase inverse** est une enzyme isolée à partir de **rétrovirus**.

**Remarque :** les **enzymes de restriction** et la **Taq Polymérase** sont des enzymes isolées à partir de **bactéries** !

**La transcriptase inverse est une enzyme d'origine virale**

Cette **ADN polymérase** peut synthétiser un brin d'ADN complémentaire (**ADNc**) dans le sens 5'-3' en prenant un **ARN** comme **matrice** pour former un **hybride ADN/ARN** (mécanisme d'action des virus).

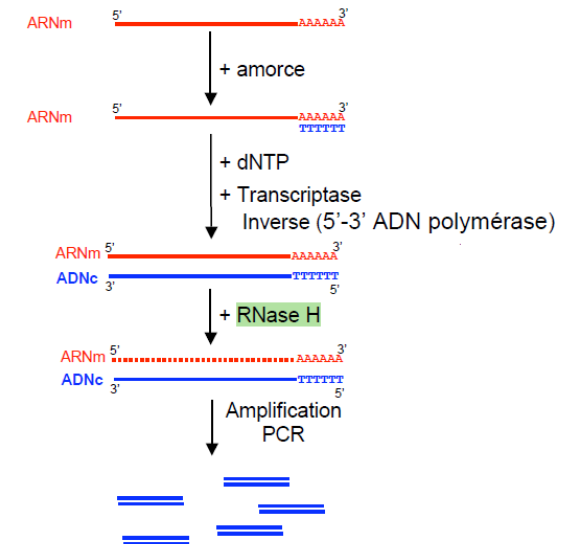
Cette enzyme génère un ADNc à partir d'un **primer** constitué de **plusieurs T successifs** (on parle de queue d'**oligo-dT**), pour qu'il puisse **s'hybrider** au niveau de la **queue poly-A des ARNm** !



**Remarque 2 :** ADNc = copie **complémentaire** de l'ARN **version ADN** !

**Remarque 3 :** à la fin de la transcription inverse, la **copie d'ADN simple brin** peut aller **directement en PCR** sans passer par l'étape de **dénaturation** !

Une fois l'**ADNc** obtenue, on se **débarrasse de l'ARNm** grâce à la **RNase H** : c'est une enzyme qui va **dégrader l'ARN uniquement** quand celui-ci est **hybridé à l'ADN** au sein d'un hybride ADN/ARN !



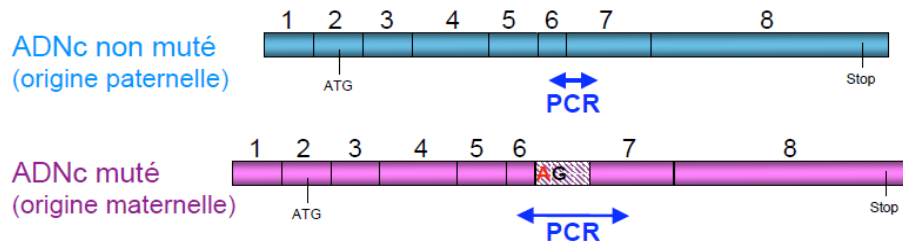
Cet ADNc pourra être **amplifié par PCR** après rajout des **2 amorces**, et les différents produits PCR pourront ensuite être **séquencés**.

On obtient donc en **version ADN uniquement** les portions du gène qui sont **restées dans l'ARNm mature après épissage** et qui seront donc traduites en protéines, à savoir les **exons** mais aussi la **portion intronique qui n'a pas été épissée** à cause du **site cryptique d'épissage** et que l'on veut identifier.

On a donc réussi à obtenir la séquence du gène WFS1 **sans les régions qui ne nous intéressent pas** (les introns non mutés) puisque la **seule portion intronique** contenue dans l'ADNc est celle qui était dans l'ARNm à cause du **site cryptique d'épissage** que l'on veut localiser ! Comme on connaît la séquence des **exons** de ce gène, il nous suffit d'isoler la séquence de l'ADNc qui ne correspond à rien pour en déduire qu'il s'agit de la séquence de l'**unique intron**, qui est donc **porteur de la mutation**. On pourra ainsi **localiser le site cryptique d'épissage** en regardant par quelles séquences exoniques cette séquence intronique est encadrée au sein de l'ADNc !

## H. Recherche de variants d'épissage : électrophorèse

On analyse nos produits PCR après **migration électrophorétique**.



Si on reprend l'exemple de la famille B, on a au niveau des allèles parentaux :

**Allèle paternel** : au niveau de l'épissage **rien n'a changé**.

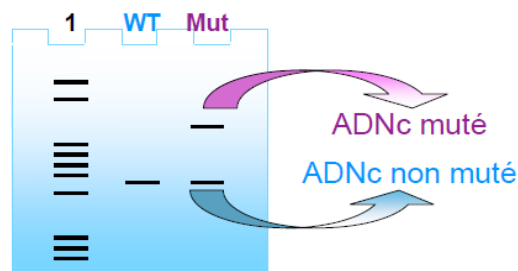
**Allèle maternel** : entre l'exon 6 et l'exon 7, il y a eu rajout d'une **partie supplémentaire** après l'épissage.

Une fois qu'on a obtenu notre **ADNc** à partir des ARNm issus de la transcription du gène WFS1 grâce à la transcriptase inverse, on va devoir **amplifier** notre matériel génétique par **PCR**.

Lors de la PCR, on met une amorce sur l'exon 6 et une seconde sur l'exon 7. On analyse ensuite les produits PCR obtenus sur gel en faisant une **électrophorèse**.

On s'attend à obtenir un **fragment plus important** pour l'ADNc d'origine **maternelle** à cause de la présence du **variant d'épissage** qui **rallonge** la séquence de l'**ARNm**.

Analyse des produits PCR  
après migration électrophorétique

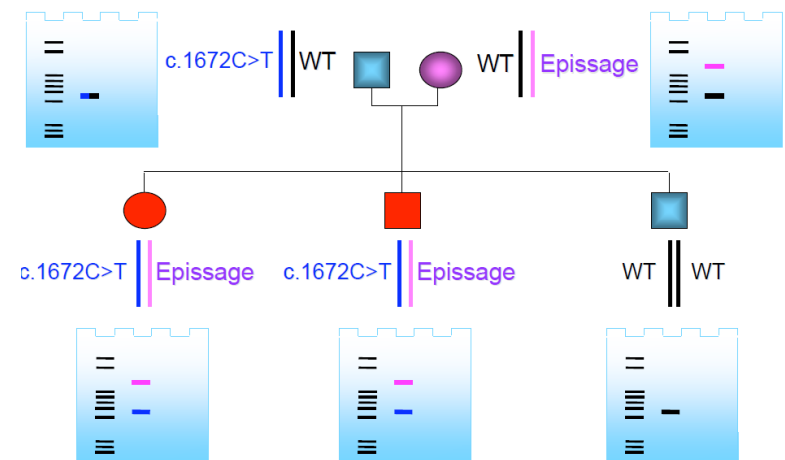


Concernant la migration électrophorétique précédente :

**Piste WT (Wild Type)** : correspondrait à la situation du **père**. La **mutation ponctuelle** 1672C>T au niveau de son exon 8 ne **modifie pas** la **taille** du fragment PCR car il s'agit juste d'une **substitution** de base.

**Allèle maternel** : correspondrait à la situation de la **mère** et des deux **enfants atteints** (le brin muté est plus lourd et migre moins loin).

Résultats de migration électrophorétique des membres de la famille B



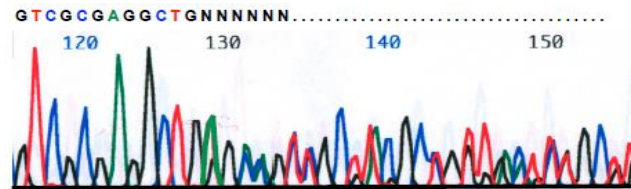
La migration électrophorétique nous permet de vérifier la présence d'un variant d'épissage qui rend l'ADNc plus lourd. Notre **hypothèse** semble donc être la bonne et on va pouvoir faire une réaction de **séquence** des produits PCR obtenus précédemment pour identifier ce variant d'épissage !

Tout le raisonnement précédent était **théorique**, dans notre tête, on a **imaginé** que la mutation était **entre les exons 6 et 7** à titre d'**exemple** pour mieux visualiser les mécanismes impliqués, mais en réalité elle peut se trouver dans **n'importe quel intron** d'un patient à l'autre, d'où la nécessité de faire un séquençage ! On est obligés de trouver précisément **les deux mutations** (via le séquençage) pour avoir le droit de poser un diagnostic car le syndrome est de **transmission récessive** et implique que **deux allèles soient mutés** !

## I. Identification de la mutation : séquençage

Si on veut identifier le variant d'épissage, on va faire un **séquençage** des produits PCR obtenus précédemment.

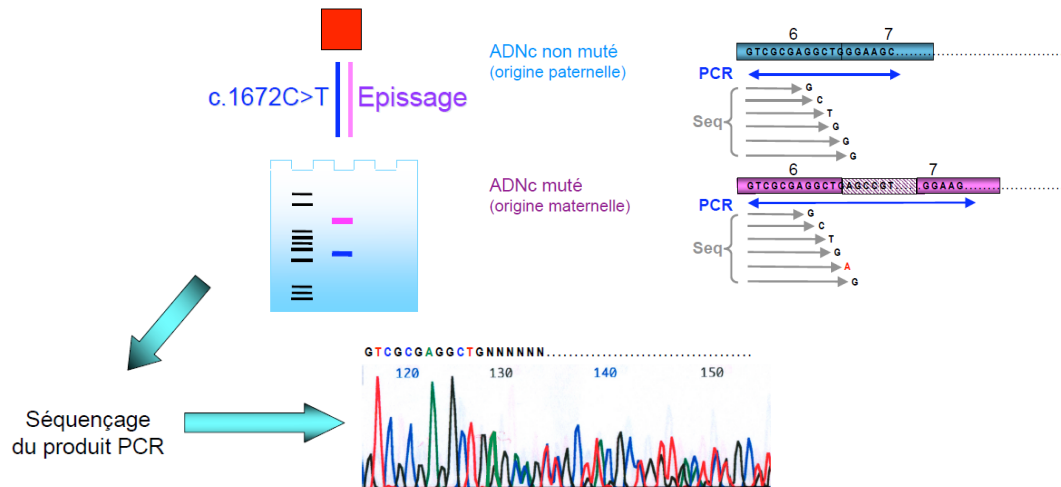
Mais on n'a pas d'autre choix que de séquencer nos **deux allèles simultanément** car les produits PCR sont mélangés !



A partir de la jonction exon6 / exon 7 (~ position 130) la séquence devient **illisible**, on a une **superposition de deux signaux**.

Ceci est dû au fait qu'on a un **mélange** de nos deux produits PCR de taille différente (à cause du variant d'épissage qui est plus long et qui décale la suite de la séquence). Ainsi, on a une **superposition** de la séquence de l'allèle paternel et de l'allèle maternel.

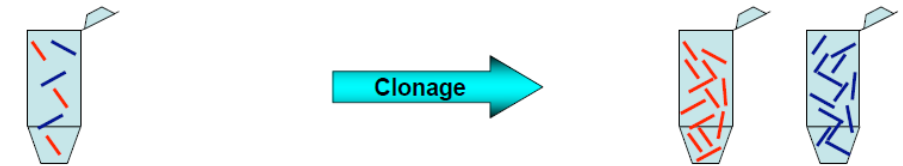
Il va donc falloir **séparer les deux populations** pour pouvoir les **séquencer individuellement** et déterminer le changement nucléotidique au niveau de l'intron. Pour cela, on va faire du **clonage moléculaire** !



## IV. Clonage moléculaire

### A. Principe du clonage

Le but du clonage est d'obtenir un **grand nombre de copies identiques** et **absolument pures** d'une séquence donnée d'ADN.



Pour ce faire, on va utiliser des **vecteurs** (ADN circulaire double brin) dans lesquels on va insérer à chaque fois un seul fragment d'ADN appelé **insert** (notre produit PCR d'origine paternelle ou maternelle).

Ces vecteurs sont ensuite introduits dans des **bactéries**.

**Le clonage moléculaire ne se fait qu'avec des cellules procaryotes (bactéries) !**

### Principe :

- 1- Intégrer un fragment d'ADN (=Insert) dans un **vecteur**.
- 2- Introduire le vecteur dans une **cellule hôte** (bactéries).
- 3- **Sélectionner, isoler et amplifier** les clones bactériens.
- 4- Obtenir un fragment d'ADN **pur en grande quantité**.





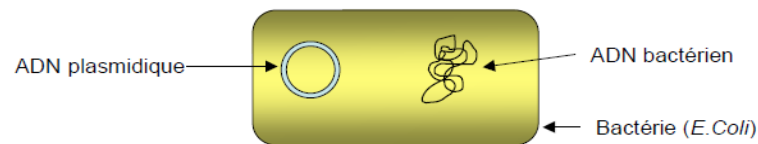
## B. Préparation de l'ADN recombinant

### 1. Vecteurs :

**ADN recombinant = vecteur + insert**

**Insert = ADNc = Produit PCR**

Un vecteur est un **ADN circulaire double brin** capable de **réplication autonome** indépendante de l'ADN de la cellule hôte. Il permet l'insertion d'un **fragment d'ADN étranger** (qu'on appelle **insert**) et il possède des **gènes de sélection** conférant un avantage à la bactérie et permettant ainsi de **sélectionner** les cellules hôtes ayant **intégré ce vecteur** !



**Le plasmide est un type de vecteur !**

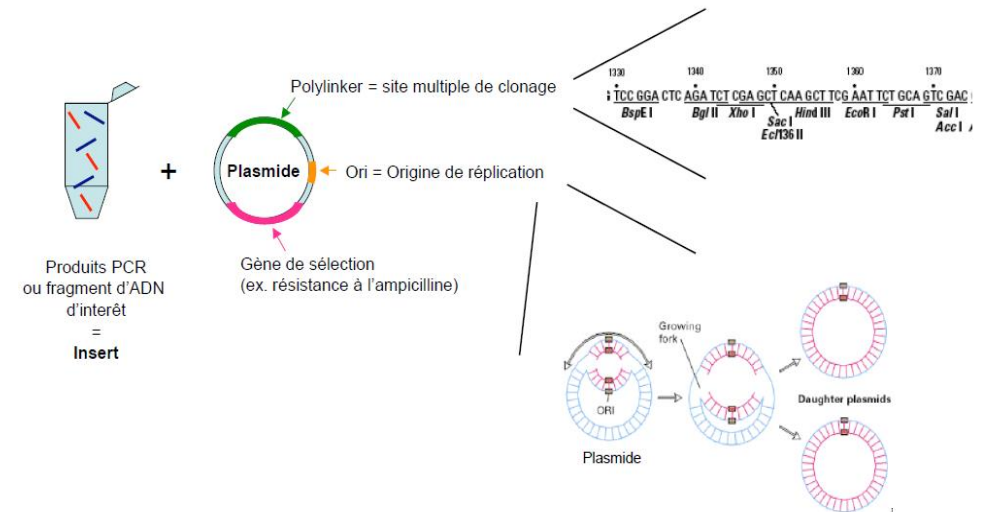
Il existe deux grandes catégories de vecteurs :

- **Vecteurs de CLONAGE** = **isoler** physiquement un fragment d'ADN d'intérêt et **amplifier** le nombre de copies de cet ADN
- **Vecteurs d'EXPRESSION** = **transférer** un gène dans une cellule hôte **eucaryote** et regarder son niveau d'expression et sa localisation dans la cellule !

En fonction de la **taille des inserts à étudier**, on distingue plusieurs **types de vecteurs**. Par exemple, les **vecteurs plasmidiques** peuvent contenir un insert de **maximum 20 kb** (kilo bases), mais si on veut insérer des fragments d'ADN **plus grands** on devra plutôt utiliser des vecteurs isolés à partir de levures.

**Le choix du vecteur utilisé dépend de la taille de l'insert**

### 2. Clonage dans un plasmide :



Pour être considéré comme un vecteur, le plasmide doit avoir :

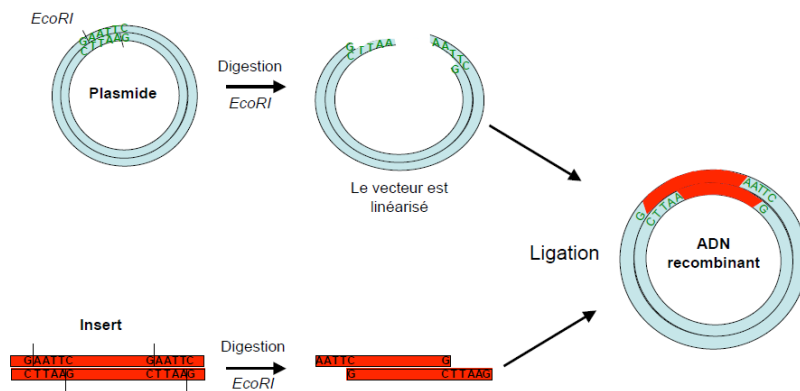
- **Un polylinker** ou **site multiple de clonage** : courte séquence **parfaitement** connue où peuvent agir différentes enzymes de restriction également connues. On les utilise pour couper l'ADN plasmidique et y insérer le produit PCR.
- **Une origine de réplication** : permet une réplication du vecteur **indépendante** de la réplication de la bactérie.
- **Un gène de sélection** : il faut pouvoir donner à la bactérie un avantage. Il s'agit généralement d'un gène de **résistance à un antibiotique** (ampicilline). De ce fait, dans un milieu avec un antibiotique, uniquement les bactéries ayant intégré un vecteur pourront se multiplier.

**Remarque :** les **ARNm** ne peuvent être introduits directement dans un vecteur, ils doivent être « **copiés** » sous forme d'**ADN**.

### 3. Préparation du vecteur et de l'insert :

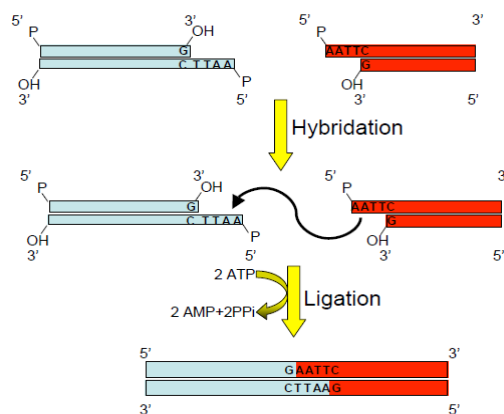
Les **mêmes enzymes de restriction** doivent digérer le vecteur et l'insert. L'idée est d'avoir une **correspondance** entre les extrémités de l'insert et du vecteur pour **faciliter l'insertion**.

Par exemple, l'enzyme **EcoRI** génère des bouts **cohésifs** au niveau du vecteur. On va alors préparer notre insert de la même manière : si on connaît la séquence reconnue par l'enzyme, on s'arrange pour avoir deux sites EcoRI aux extrémités de l'insert. Le but est d'avoir une **correspondance** entre les **extrémités** du **vecteur** et de l'**insert** pour **faciliter l'insertion** !



Ainsi, en mettant le vecteur et l'insert ensemble et en rajoutant une **ligase**, on va pouvoir insérer notre fragment PCR dans le vecteur !

La **T4 DNA ligase** catalyse la formation d'une liaison **phosphodiester** entre un 3'OH et un 5'Phosphate en présence d'ATP et d'ions divalents = liaison covalente entre deux fragments d'ADN.

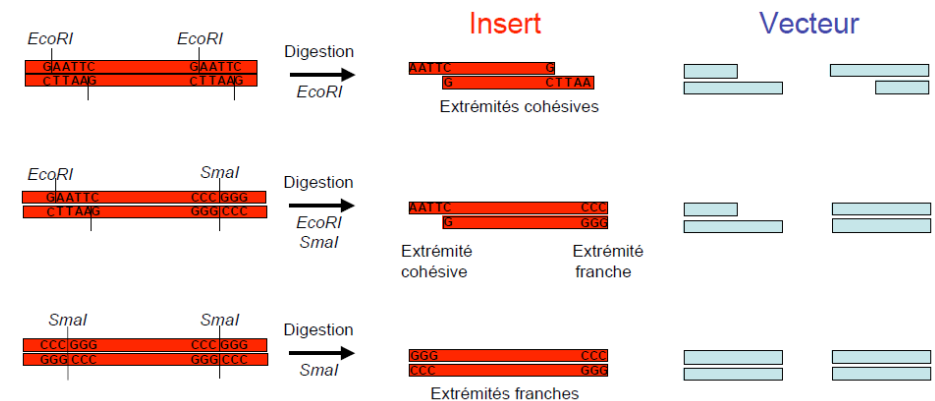


### 4. Stratégies de clonage :

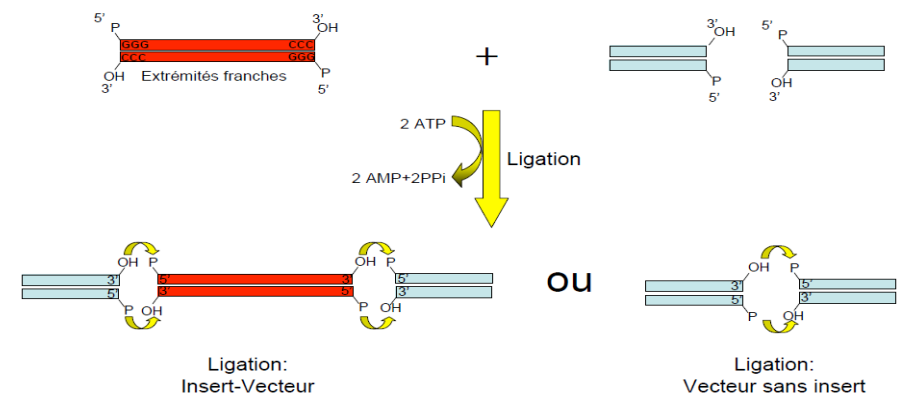
Plusieurs enzymes de restriction peuvent agir au niveau du **polylinker**. On peut également les utiliser en **combinaison** !

Suivant la stratégie utilisée, on peut être amené à travailler avec :

- des extrémités **cohésives** (EcoRI) : les deux brins du plasmide et de l'insert ne sont pas coupés exactement l'un en face de l'autre, un brin demeure plus long que l'autre, la coupure a un aspect « en biseau ».
- des extrémités **franches** (SmaI) : les deux brins sont coupés l'un en face de l'autre au nucléotide près, la coupure est franche et nette, bien droite.



**Lorsqu'on a des extrémités franches**, le vecteur (= plasmide) aura tendance à se refermer sur lui-même **avant d'avoir intégré l'insert** ! La **T4 DNA ligase** va créer une liaison phosphodiester **sans insert**.

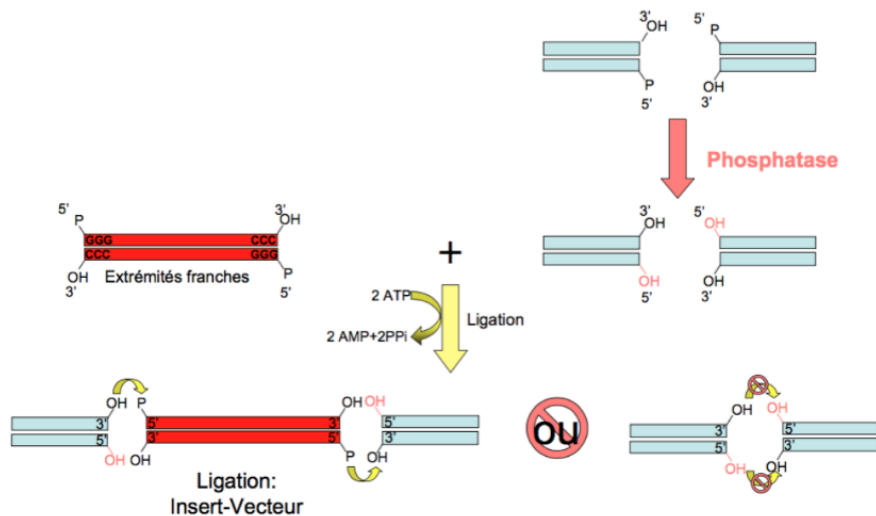


La **T4 DNA ligase** n'a qu'**une seule liaison phosphodiester** à former dans le cas où **l'insert n'est pas présent**, alors qu'elle doit en faire **deux** si l'insert est **intégré dans le vecteur**.

L'enzyme aura donc tendance à **privilégier** le cas le plus simple, c'est-à-dire celui où elle n'a qu'**une seule liaison à faire**, et donc où l'insert n'est **pas intégré** dans le vecteur !

Or ce cas de figure ne nous intéresse pas et on cherche à **l'éviter**, car on veut former notre **ADN recombinant**, on veut donc que le vecteur intègre l'insert !

Pour éviter ça, avant la ligation, une étape **facultative** de **déphosphorylation** est nécessaire **uniquement** dans le cas où **les extrémités sont franches**. Une **phosphatase** catalyse le retrait du phosphate à **l'extrémité 5' du vecteur**, empêchant ainsi la **liaison phosphodiester** de se former sans l'insert. Attention, **on ne touche pas à l'insert**, on ne **déphosphoryle** que le **vecteur** !



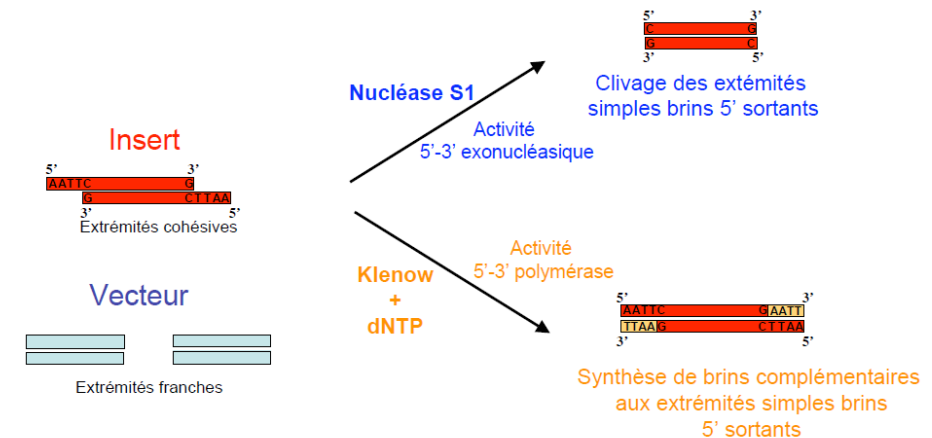
**Remarque :** si on a des extrémités **cohésives**, cette étape de déphosphorylation n'est **pas nécessaire**. En effet, les extrémités de l'insert et du vecteur ayant des **portions simple brin complémentaires**, il peut y avoir un appariement avec des liaisons **hydrogène** entre eux sans attendre l'action de la T4 DNA ligase ! L'enzyme viendra **consolider** l'appariement en formant des liaisons **covalentes** (phosphodiester) entre l'insert et le vecteur, mais le vecteur **ne pourra pas se refermer sur lui-même sans l'insert** à cause des extrémités simple brin (cohésives) **uniquement complémentaires de l'insert**.

On utilise de préférence le **même type d'enzymes de restriction** sur le vecteur et sur l'insert pour avoir les mêmes extrémités **franches** ou **cohésives**.

Cependant, il existe des situations où on ne peut pas utiliser les mêmes enzymes de restriction pour le vecteur (pour lequel on dispose d'un grand nombre d'enzymes utilisables) et pour l'insert.

On a donc recours à une **stratégie de clonage** car le vecteur et l'insert ne présentent pas le même type d'extrémités (on peut avoir un insert à bords francs et un vecteur à bords cohésifs ou l'inverse, dans les deux cas les deux ne sont pas complémentaires et ne pourront pas s'apparier/s'hybrider).

Ainsi, dans la situation où on a un **insert à bords cohésifs** et un **vecteur à bords francs** il est possible de transformer les **extrémités cohésives** de l'insert en **franches** :



- **Nucléase S1** : elle possède une activité **5'-3' exonucléasique** lui permettant de grignoter les extrémités **simple brin** sortantes : l'enzyme **s'arrête** de grignoter les nucléotides dès qu'elle rencontre de l'**ADN double brin** !
- **Klenow** : elle possède une activité **5'-3' polymérase** qui lui permet de combler l'ADN simple brin en synthétisant les brins complémentaires **en présence de dNTP** !

### C. Introduction du vecteur dans une cellule hôte

Nous avons vu en résumé qu'on peut travailler soit sur le **vecteur**, soit sur l'**insert**, le but étant d'**obtenir un ADN recombinant**, c'est-à-dire un vecteur ayant intégré un produit PCR.

**On considère qu'un seul produit PCR a été incorporé par vecteur**

On dispose pour ce faire de différents outils :

- **Couper** : grâce aux **enzymes de restriction** (= endonucléases qui coupent au milieu de la séquence) mais aussi grâce aux exonucléases (qui digèrent la séquence par ses extrémités).
- **Copier** : grâce aux **polymérases** qui synthétisent un brin complémentaire
- **Coller** : grâce aux **ligases**

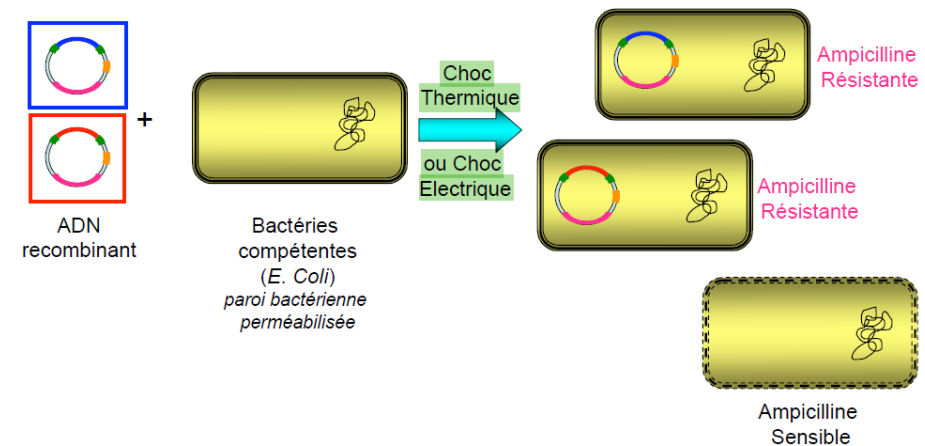
A ce stade, nous avons réussi à produire un **ADN recombinant**, c'est-à-dire un **plasmide** contenant un **insert (ADN complémentaire)** d'origine paternelle **ou** maternelle (puisqu'on s'arrange pour qu'un vecteur ne puisse **intégrer qu'un seul produit PCR/Insert/ADNC**) !

Nous n'avons à ce stade **aucun moyen de savoir** quels vecteurs contiennent un **insert d'origine maternelle** et quels vecteurs contiennent un **insert d'origine paternelle** mais ce n'est pas grave.

En effet pour le moment nous voulons juste **séparer nos inserts** pour pouvoir **séquencer séparément** l'allèle paternel et l'allèle maternel. Le but est toujours de **séquencer notre deuxième mutation** (celle de la mère) pour diagnostiquer le syndrome de Wolfram !

Pour rappel, le séquençage était **illisible** à cause de la **superposition** et du **décalage** de séquence entre l'allèle paternel porteur d'une **mutation ponctuelle**, et l'allèle maternel qui est notre **variant d'épissage**.

Le fait qu'un vecteur ne puisse contenir qu'un seul insert est une **première manière de séparer nos allèles** (nos produits PCR). On ne pourra **distinguer** l'allèle paternel de l'allèle maternel que lors de l'étape finale : le **séquençage**.



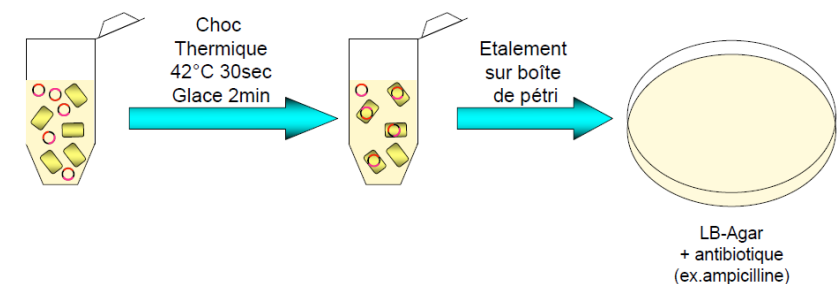
Pour faciliter l'entrée des ADN recombinants dans les bactéries, on **perméabilise leur paroi** grâce à un **choc thermique ou électrique**. On s'arrange pour qu'il y ait **un seul ADN recombinant par bactérie** : c'est une manière de **séparer encore un peu plus** nos produits PCR d'origine paternelle et maternelle !

**Introduction d'ADN dans une cellule bactérienne  
= TRANSFORMATION**

**Introduction d'ADN dans une cellule eucaryote  
= TRANSFECTION**

Dans le mélange, on retrouve :

- des bactéries ayant intégré l'ADN recombinant avec l'ADNc **paternel**
- des bactéries ayant intégré l'ADN recombinant avec l'ADNc **maternel**
- des bactéries **sans ADN recombinant** qui sont **ampicilline-sensibles**, vu que le gène de résistance est situé dans l'ADN recombinant...

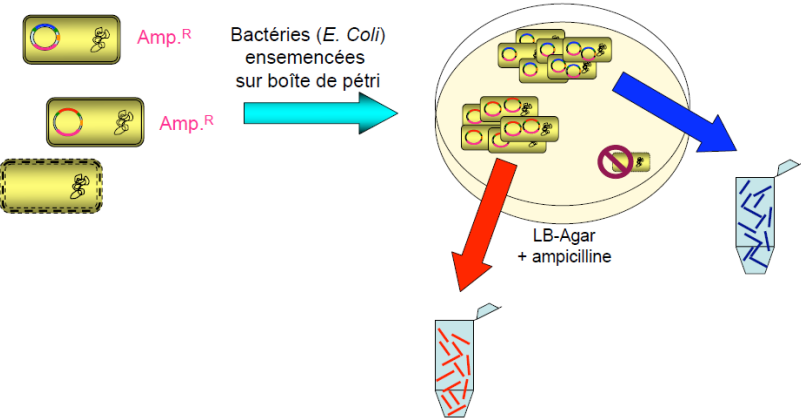




D. Sélection des clones bactériens

Sur les boîtes de pétri sur lesquelles on a étalé les bactéries, on a mis de l'**ampicilline** dans le milieu pour ne garder que les bactéries ayant incorporé un **ADN recombinant**, et donc le **gène de résistance**.

Seules les bactéries contenant le gène de résistance à l'antibiotique forment des colonies.



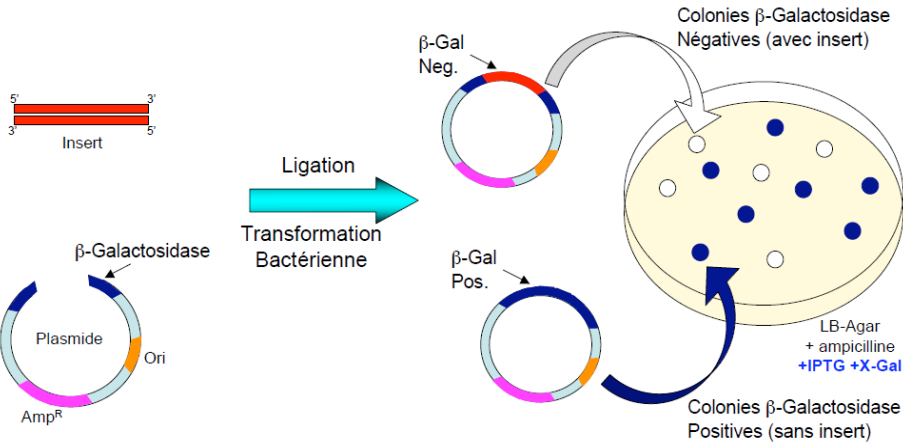
Si on repique dans un tube différent la colonie du haut ou celle du bas et qu'on extrait l'ADN, on aura réussi à **séparer** nos deux ADNc d'origine paternelle et maternelle !

Sélection Blanc / Bleu :

Même après l'étape de **déphosphorylation**, ou même dans des conditions optimales, on aura toujours des cas où le **vecteur se referme sur lui-même**. Il existe une solution pour sélectionner rapidement uniquement les bactéries qui ont **intégré un vecteur AVEC insert** !

L'ampicilline permet de sélectionner les bactéries ayant intégré un vecteur avec ou sans insert

La sélection Blanc / Bleu permet de sélectionner les bactéries ayant intégré un vecteur avec insert uniquement



Le **polylinker** est placé dans le gène qui code pour la **bêta-galactosidase**. A partir du moment où on va intégrer l'insert au niveau du polylinker, le gène de la bêta-galactosidase (*en bleu*) ne va **plus être fonctionnel** puisqu'il sera **coupé en deux** à cause de l'insert !

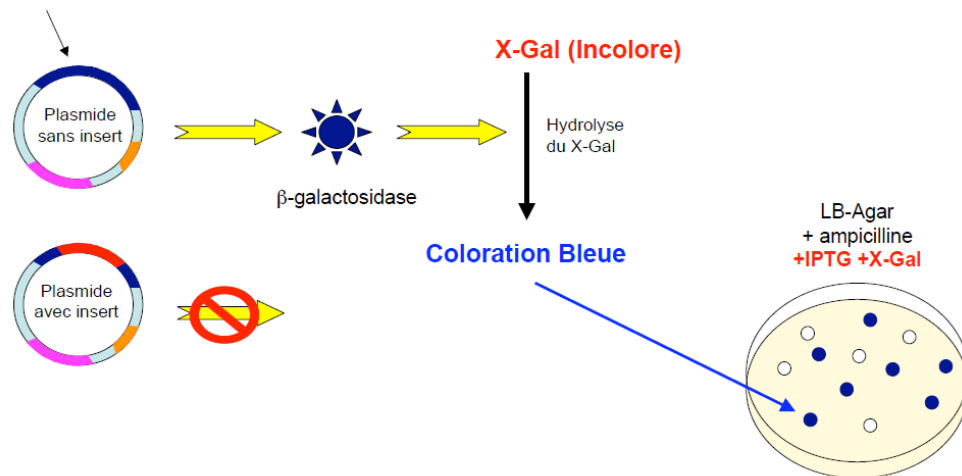
Par contre, les bactéries qui ont un vecteur qui s'est refermé sur lui-même (**sans insert**) auront une **bêta-galactosidase fonctionnelle**.

Si on ajoute dans le milieu de l'**IPTG** (agent qui induit l'expression de la bêta-galactosidase) et le substrat incolore **X-Gal**, l'hydrolyse de ce substrat va pouvoir donner une **coloration bleue**. On aura donc un **mélange** de colonies **blanches** et de colonies **bleues**.

Sur la boîte de pétri, on va donc **repiquer/garder uniquement** les colonies **blanches**, autrement dit celles qui ont une bêta-galactosidase non fonctionnelle à cause de la présence de l'**insert** !

AVEC INSERT	SANS INSERT
Gène bêta-galactosidase inactivé	Gène bêta galactosidase <b>activé</b>
X-Gal incolore <b>non hydrolysé</b>	X-Gal <b>hydrolysé colorant</b> le milieu en <b>bleu</b>
<b>COLONIES BLANCHES</b>	<b>COLONIES BLEUES</b>

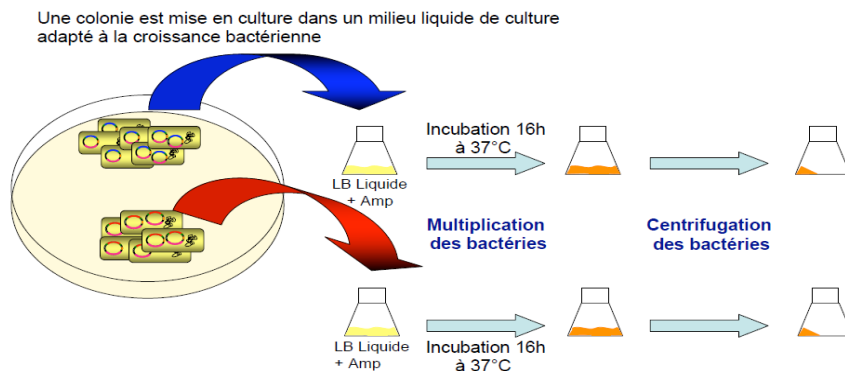
**IPTG** : Induit l'expression de la  $\beta$ -galactosidase



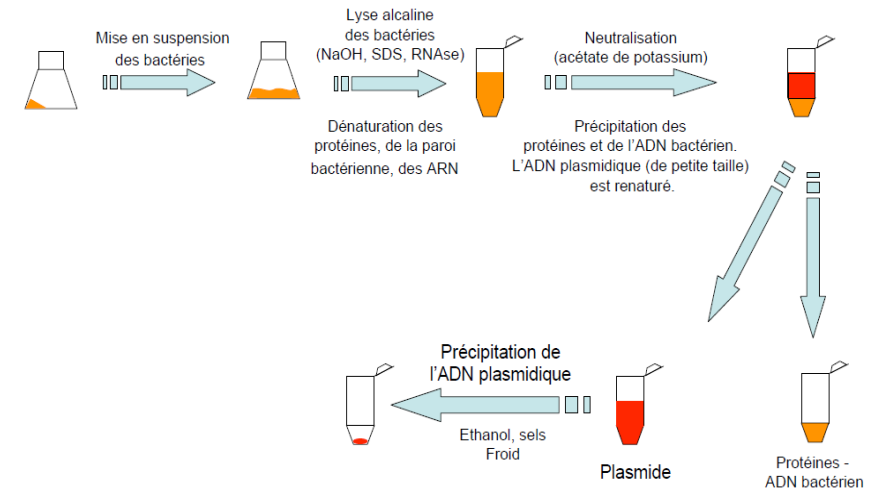
## E. Amplification des clones bactériens

On repique une colonie, que l'on met en culture dans un milieu **liquide** de culture adapté à la croissance bactérienne, toujours en présence d'**ampicilline**.

Chaque culture est mise en incubation pendant **16 h à 37°C**, les bactéries vont **se multiplier** et par la même occasion **amplifier** notre **ADN recombinant**.



S'en suit ensuite une **centrifugation**, une **lyse alcaline** des bactéries (**dénaturation** des **protéines** et de la **paroi bactérienne** par la soude ou un détergent pour récupérer les acides nucléiques) puis une **neutralisation** (permettant de **précipiter** les **protéines** et l'**ADN bactérien** dans le culot).



L'**ADN plasmidique** est **plus petit** que l'ADN bactérien, il est capable de **se renaturer** et va rester **en suspension** et **soluble** dans le tube !

Cet ADN plasmidique est **récupéré** : on le fait **précipiter** en rajoutant de l'**éthanol froid** et des **sels** pour former une **méduse d'ADN**.

On va ensuite le **purifier** et en récupérer en **grande quantité**, avant de pouvoir travailler dessus.

**Remarque** : on veut séparer les deux populations d'ADN (paternel et maternel) en isolant les bactéries sur une boîte de pétri et en les étalant de façon à obtenir plein de petites colonies.

## UNE SEULE bactérie est à l'origine de toute une colonie



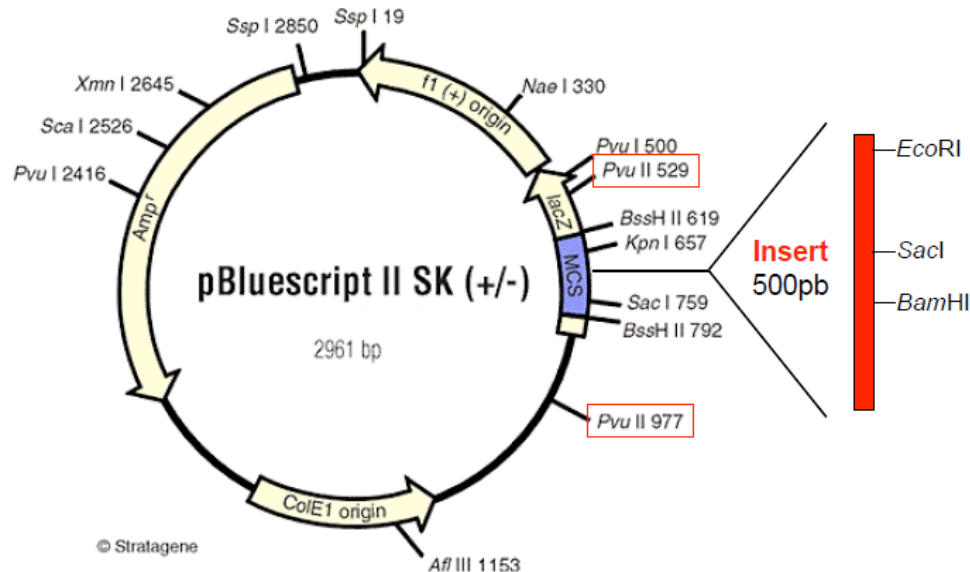
Un seul « point » correspond à une **colonie** et non **pas à une bactérie** ! Il s'agit en réalité d'une bactérie qui s'est multipliée pour donner une colonie : c'est la **même population**. Par contre d'une colonie à l'autre les bactéries ne contiendront **pas forcément le même ADN recombinant** (paternel ou maternel). On s'arrange aussi

pour que les bactéries soient suffisamment éloignées les unes des autres pour que les clones soient purs et qu'on en ait un seul type par colonie !

## F. Carte de restriction

On réalise des **cartes de restriction** pour vérifier si l'insert qu'on a récupéré est bon, ou du moins s'il est de la **bonne taille**.

Les cartes de restrictions sont des documents sur lesquels on retrouve la **séquence bien caractérisée d'un plasmide** : on connaît la position précise des **sites de coupure** de telle ou telle **enzyme de restriction**.



On veut savoir dans cet exemple si l'ADN recombinant récupéré à partir des bactéries contient notre insert de **500 pb**.

Les ADN recombinants purifiés vont donc être **analysés** par digestion enzymatique, et la **carte de restriction** réalisée.

L'**enzyme de restriction** utilisée est Pvu II : elle coupe en position 529 et 977.

En fonction de la **taille des produits de digestion** retrouvés sur le gel de l'**électrophorèse**, on saura si on a un plasmide **avec insert** ou un plasmide **sans insert** !

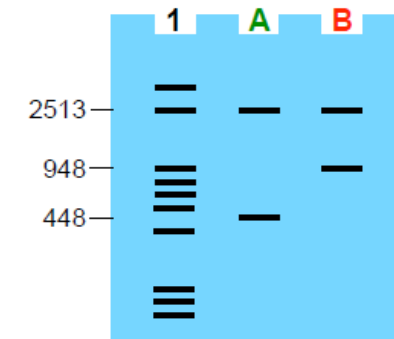
La digestion par Pvu II génère deux fragments, un petit et un grand. Selon la présence ou non de l'insert, la taille du fragment sera différente.

○ Dans le cas d'un plasmide **sans insert** (piste A) on retrouve :

- Un **petit fragment** de  $977 - 529 = 448$  pb
- Un **grand fragment** de  $2961 - 448 = 2513$  pb

○ Dans le cas d'un plasmide **avec insert** (piste B) on retrouve :

- Un **petit fragment** de  $977 - 529 + 500 = 948$  pb
- Un **grand fragment** de  $2961 - 448 = 2513$  pb



1- Marqueur de taille

La **taille** des produits de digestion nous permet donc de déterminer la **présence** ou non de l'**insert** dans le plasmide !

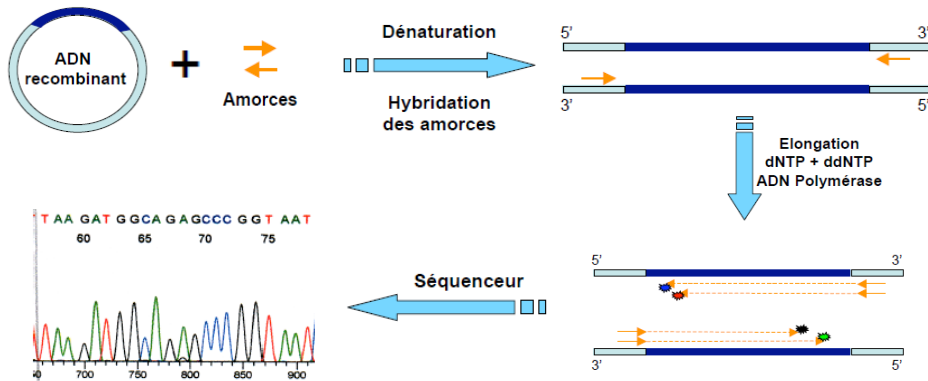
Pour la suite de l'étude, dans cet exemple précis, on ne garderait que les plasmides ayant généré un **fragment à 948 pb** (piste B), car ce sont eux qui ont **intégré un insert** !

On va donc récupérer ces plasmides, en **extraire l'insert** qu'ils contiennent et le **séquencer isolément** pour déterminer s'il s'agit de l'ADN recombinant d'**origine paternelle ou maternelle**, mais surtout pour **identifier la mutation intronique** maternelle à l'origine du site cryptique d'épissage ! Nous pourrions de cette façon **identifier la deuxième mutation** qui nous faisait défaut et **conclure à un diagnostic de syndrome de Wolfram** !

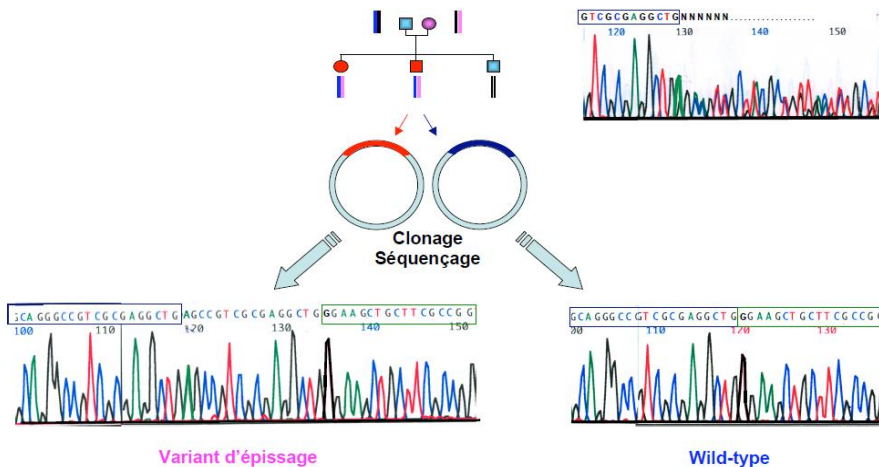
## G. Séquençage des fragments d'ADN pur

Tout à l'heure on avait un **mélange de variants** différents (d'origine paternelle et maternelle), d'où un **séquençage illisible**.

Maintenant que les ADN recombinants sont **purifiés** et **isolés**, on va pouvoir les **séquencer individuellement** !



- **Pour l'allèle maternel** : présence d'un **variant d'épissage** entre l'exon 6 et l'exon 7, dont la séquence peut être parfaitement déterminée.
- **Pour l'allèle paternel** : **jonction exon 6/exon 7 normale**.



Pour en revenir au **cas de la famille B**, on voit que les enfants sont bien **atteints du syndrome de Wolfram** car ils ont hérité de l'allèle paternel porteur d'une **mutation sur l'exon 8**, et de l'allèle maternel porteur d'un **variant d'épissage**.

## V. Clonage d'expression

Avec l'arrivée du **NGS** on peut maintenant séquencer facilement tout l'**exome** (l'ensemble des exons de tous les gènes connus) dont on arrive à déterminer les **variants** et les **modifications**.

On peut donc se demander quel est le **caractère pathogène** de ce nouveau variant ? Il y a encore d'autres techniques qui doivent se mettre en place pour nous permettre d'**interpréter** la **pathogénicité** d'un variant (clonage d'expression, étude d'expression, tests fonctionnels).

Dans l'étude d'une **protéine mutante** (comme la *wolframine*) on peut faire du **clonage d'expression** pour étudier artificiellement dans les cellules l'**expression** de cette protéine.

Les **vecteurs d'expression** vont pouvoir être exprimés dans les cellules **eucaryotes**, contrairement aux vecteurs de **clonage** qui sont réservés aux cellules **procaryotes**.

L'assemblage de l'ADN recombinant se fait de la même façon qu'avec les vecteurs de clonage : une **enzyme de restriction ouvre le plasmide** pour l'insertion d'un fragment d'ADNc.

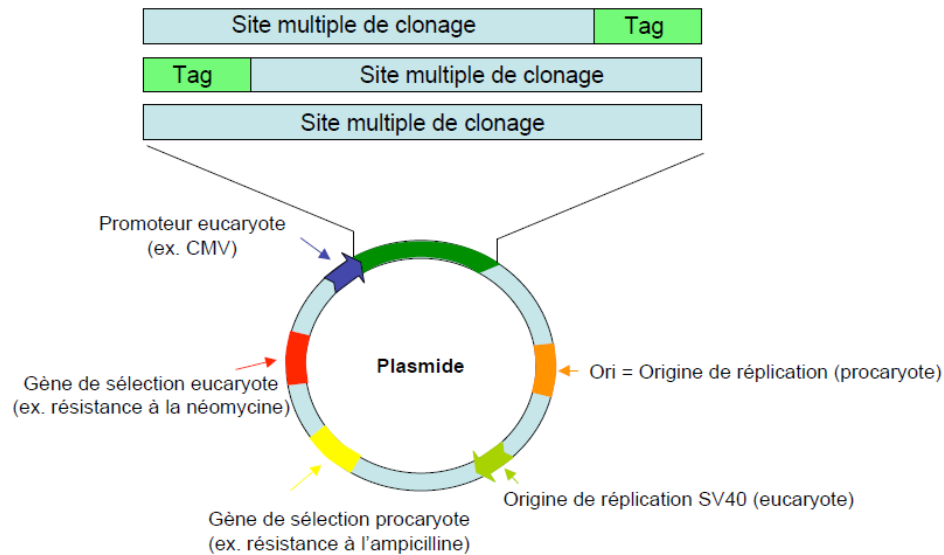
La multiplication se fait d'abord dans une **bactérie** (mais l'expression se fera dans une cellule eucaryote), d'où la nécessité d'avoir :

- **Un polylinker ou site multiple de clonage**
- **Une origine de répllication PROCARYOTE**
- **Un gène de sélection PROCARYOTE**

Puis les vecteurs d'expression seront ensuite isolés et intégrés dans des cellules **eucaryotes** pour se multiplier et **s'exprimer**, d'où la nécessité d'avoir **en plus** (contrairement à un vecteur de clonage) :

- **Une région promotrice EUCARYOTE** (origine virale)
- **Une origine de répllication EUCARYOTE** (SV40, virale)
- **Un gène de sélection EUCARYOTE**





## A. Protéines de fusion :

Le **vecteur d'expression** permet à une cellule eucaryote de **surexprimer** une protéine d'intérêt.

Notre protéine va s'exprimer a priori dans tel ou tel compartiment (mitochondrie, REG...).

Une fois la protéine **surexprimée**, il est nécessaire de pouvoir **la suivre** :

- **Soit il existe des anticorps (réaction Ag/Ac)**

Mais si c'est une protéine qu'on ne **connaît pas**, ou si on n'a **pas d'anticorps spécifiques**, on utilise d'autres outils tel que le marquage de la protéine :

- **Soit il faut utiliser des étiquettes (Tag)**

L'étiquette ou **Tag** est une répétition de petites séquences définies greffée en **N-term** ou **C-term** qui peut :

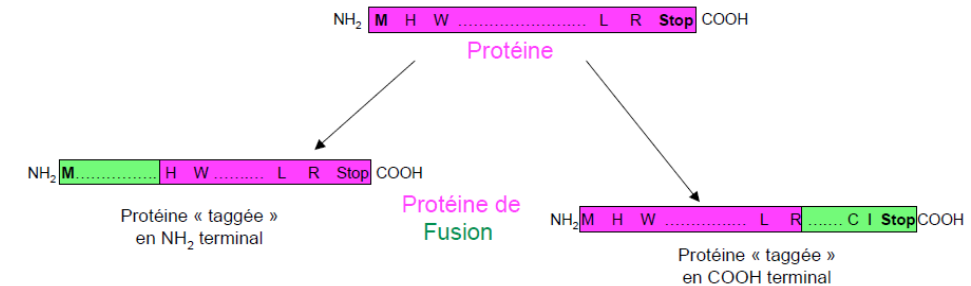
- soit être reconnu par un **anticorps** spécifique (petit peptide)
- soit être **fluorescent** (GFP, YFP)

On génère ce qu'on appelle des **protéines de fusion** : on synthétise la protéine d'intérêt directement avec à son extrémité **N-term** ou **C-term** un **petit peptide** ou une **protéine fluorescente**.

**Protéine de fusion = ADNc d'intérêt + étiquette**

Grâce à ces étiquettes on va pouvoir voir si le variant a un **impact** sur la protéine et surtout sa **localisation** intracellulaire !

Au moment de la traduction, il faut donc que le **Tag** soit **traduit avec la protéine**. Il y a deux situations :



- **Protéine taguée en N-terminal** : il faut **enlever le codon Start** au niveau de l'ARNm (codon pour la **méthionine**) pour que la traduction démarre au niveau du **codon Start de l'étiquette**.
- **Protéine taguée en C-terminal** : il faut **enlever le codon Stop** de l'ARNm pour que la traduction puisse se faire jusqu'au **codon Stop de l'étiquette**.

Il existe **différents Tags** :

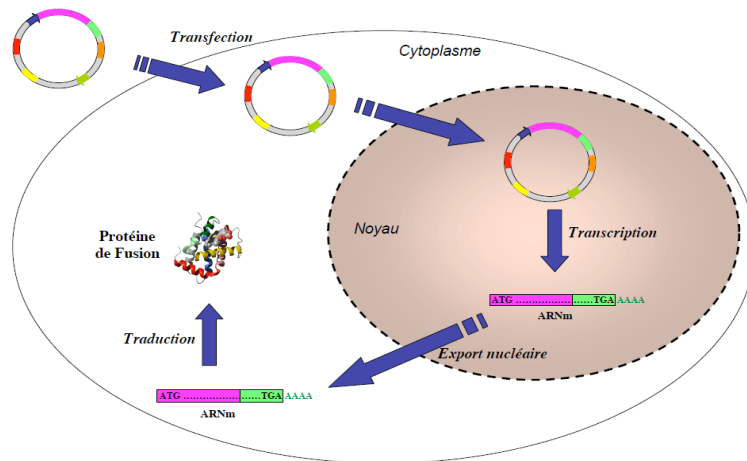
- **Peptide** : épitope reconnu par tel ou tel anticorps (V5, c-myc...)
- **Protéine fluorescente** : fluorescence sous un microscope particulier, il existe différentes couleurs (EGFP, EYFP, dsRED)

## B. Transfection des cellules eucaryotes :

L'ADN **exogène** d'intérêt préparé est transféré dans une cellule **eucaryote** pour être **surexprimé** par la machinerie cellulaire : on parle de **transfection**.

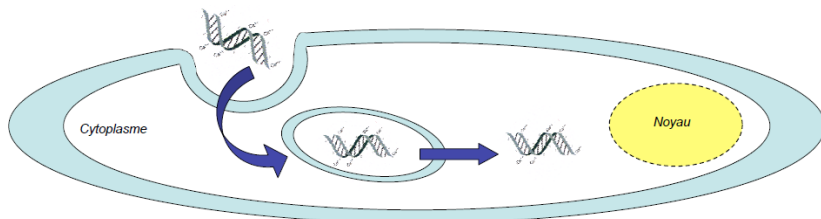
**Différentes techniques** sont utilisées :

- Réactifs **chimiques** (phosphate de calcium, DEAE Dextran, liposomes)
- Méthodes **physiques** (électroporation, micro-injection dans la cellule)
- Utilisation de **particules virales** (infection)



### Transfection par phosphate de calcium :

L'ADN est mélangé avec une **solution de chlorure de calcium**, puis on ajoute un **tampon phosphate** goutte à goutte. Il se forme ainsi un **précipité** de phosphate de calcium autour de l'ADN formant des « cristaux ». Ces cristaux de phosphate de calcium contenant l'ADN vont ensuite pénétrer dans les cellules en culture par **endocytose** ou **phagocytose**. L'ADN va être **libéré dans le cytoplasme** et en fonction des cas, il pourra **s'intégrer ou non dans le génome nucléaire**, pour ensuite être **transcrit en ARNm** et **exporté** dans le cytoplasme pour la **traduction**.



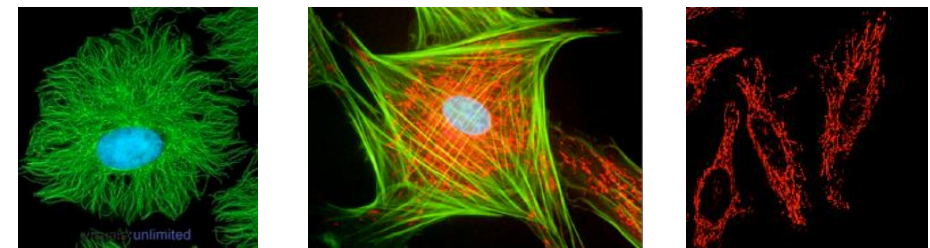
## C. Etude de la localisation des protéines :

Il est très utile de savoir où se répartissent nos **protéines mutantes** dans la cellule !

On va ainsi avoir recours à des techniques de **microscopie à fluorescence** nous permettant de visualiser ces protéines après **fixation** et **perméabilisation** des cellules :

- soit **directement** grâce à une étiquette, un **Tag fluorescent** (protéines taguées GFP par exemple)
- soit **indirectement** grâce à l'**immunofluorescence** : la protéine est révélée après fixation d'un Ac couplé à un marqueur fluorescent, la réaction Ag-Ac permet de révéler la protéine

**Les études de localisation utilisant des Tag fluorescents ou de l'immunofluorescence concernent uniquement les protéines**



**filaments de tubuline : vert / mitochondries : rouge / noyau : bleu**

Le réseau de filaments de tubuline a été marqué avec un tag fluorescent GFP (vert). Le noyau peut également être marqué grâce à la présence d'un composé qui **s'intercale dans l'ADN** et qui fluoresce en bleu. Le réseau des mitochondries a été marqué en rouge.

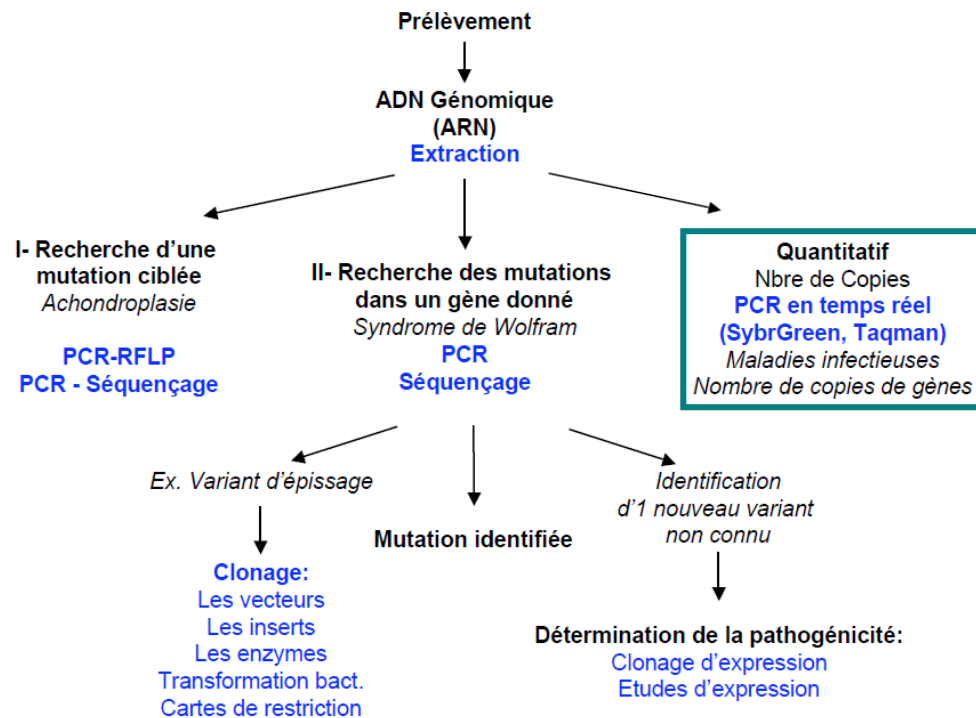
On peut ainsi **combinaison plusieurs tags** : le **cytosquelette** apparaît en vert et les **mitochondries** sont marquées en rouge.

En combinant ces différentes approches, on peut vérifier si **localisation de la protéine** est normale et même si l'on n'a **pas sa fonction exacte**, on va avoir une idée de la **pathogénicité** du variant !

## VI. PCR en temps réel

Nous avons vu jusqu'à présent dans les deux premiers cours comment trouver une **mutation**, que ce soit une **mutation ciblée** (achondroplasie) ou des **mutations dans un gène donné** (syndrome de Wolfram).

Dans ces deux cas nous avons manipulé notre matériel génétique dans l'unique but final de **séquencer** un gène du patient :



Nous avons donc cherché à **déterminer la nature** de quelque chose (en l'occurrence des nucléotides), ces techniques avaient toutes un but **qualitatif**.

Mais dans certains cas, on peut être amenés à **quantifier** quelque chose, on fera donc des études de **nombre** (on ne s'intéresse plus à la nature des nucléotides).

Notamment en **virologie**, si on veut déterminer la **charge virale**, il faut qu'on puisse savoir combien de **nombre de copies du virus** étaient dans notre échantillon (mais on se fiche de connaître leur séquence nucléotidique).

## A. Principe de la PCR en temps réel :

La PCR en temps réel utilise le **même principe que la PCR classique**, avec toujours les **trois mêmes étapes** que l'on répète **n cycles** :

**1. Dénaturation** (95°C)

**2. Hybridation** (55°C)

**3. Elongation** (60/72°C en fonction de l'enzyme utilisée)

L'**avantage** de cette technique est de pouvoir **mesurer la quantité de produits PCR générée en temps réel tout au long de la réaction**, et non pas seulement après 35 voire 40 cycles !

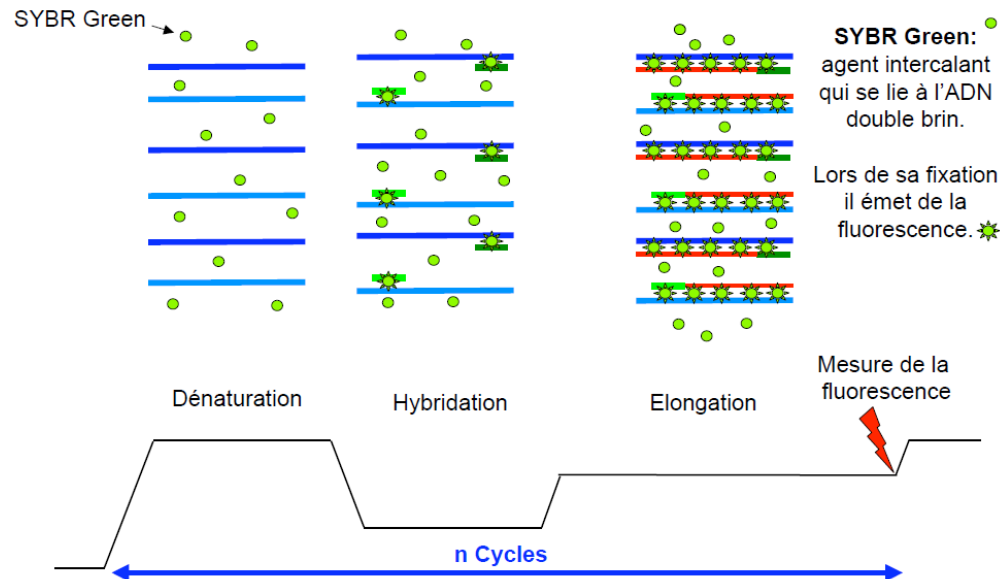
PCR classique	PCR en temps réel
Mesure de la quantité d'ADN <b>après 35-40 cycles</b>	Mesure de la quantité de fluorescence directement <b>après chaque cycle</b> grâce à un <b>thermocycleur</b> Utilisation de l'agent intercalant <b>SYBR Green</b>

La **mesure** se fait grâce à un **agent fluorescent** dont l'émission de fluorescence est **proportionnelle** à la **quantité d'amplicons produits**.

Le **SYBR Green** est une petite molécule qui va **s'insérer** dans l'ADN **double brin** et qui a la propriété d'émettre une fluorescence uniquement à partir du moment où elle est intercalée !

**La mesure de la fluorescence se fait à la fin de chaque cycle après l'étape d'élongation**

En toute logique, **plus on a de produits générés, plus le SYBR Green est incorporé dans de l'ADN double brin**, donc plus y a d'amplicons (qui sont justement des ADN double brin) et **plus il y aura de fluorescence mesurée** par l'appareil !



- **PCR classique** : Après 40 cycles, une certaine quantité de produits PCR est générée. Un agent intercalant est ajouté dans le milieu réactionnel, ce qui rend le système **fluorescent sous lumière U.V.**

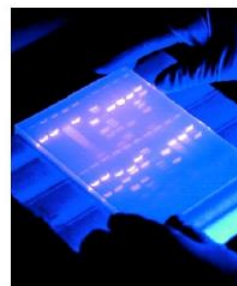
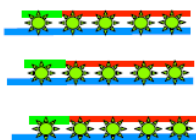
⇒ Technique **non quantitative**

- **PCR en temps réel** : On mesure la fluorescence après chaque cycle car l'agent intercalant est incorporé au fur et à mesure de l'avancement de la réaction, on ne l'ajoute pas à la fin !

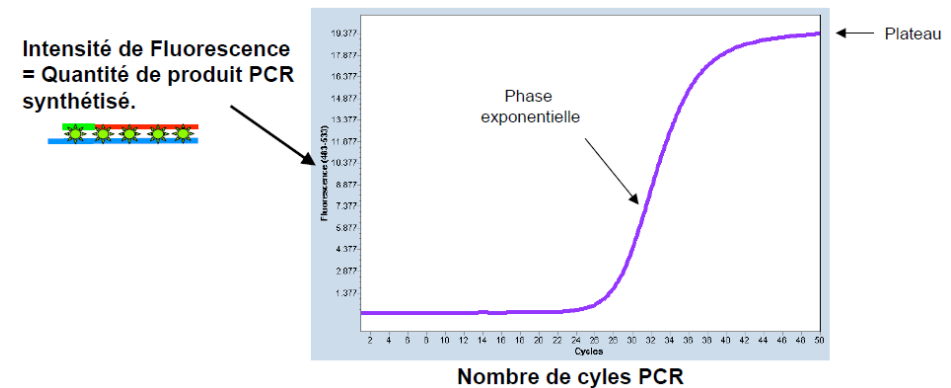
⇒ Technique **quantitative**

PCR « classique » :  
35-40 Cycles

PCR en temps réel:  
Mesure de Fluorescence  
à chaque cycle



## B. PCR quantitative :



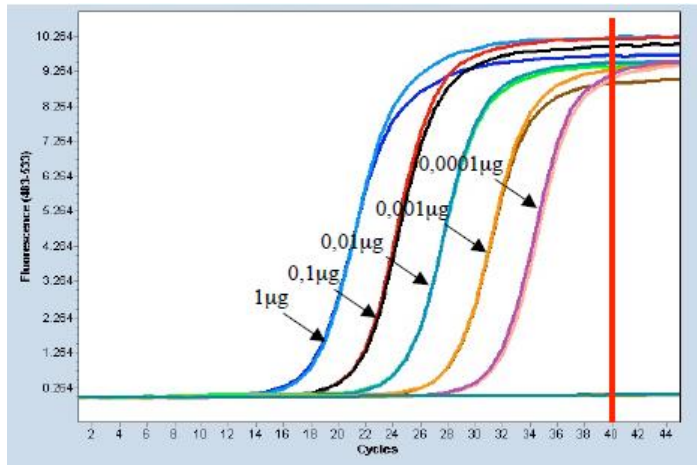
La courbe représente l'**intensité de fluorescence** des produits PCR générés en fonction du **nombre de cycles**, et présente trois phases :

- **Phase de plateau** : la quantité de fluorescence générée au niveau des premiers cycles n'est **pas suffisante** pour que le système la **détecte**.
- **Phase exponentielle** : au bout d'une **vingtaine de cycles**, les produits PCR sont générés **exponentiellement** jusqu'à **saturation du système** (au bout de 40 cycles environ).
- **Phase de plateau** : le système est **saturé** car le **milieu réactionnel** est **épuisé** (plus de dNTP par exemple) et la **Taq Polymérase** fatigue.

**Plus on a d'ADN au départ, moins il faudra de cycles pour observer la fluorescence (point d'inflexion plus rapide)**

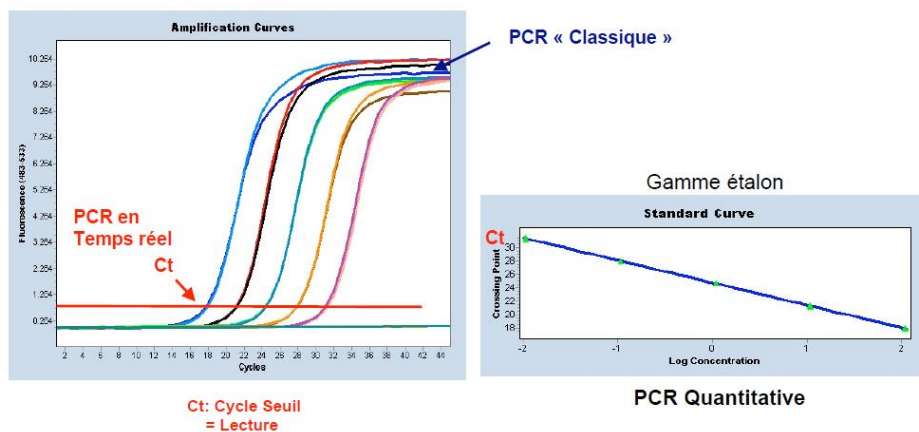
En mettant **1 microgramme d'ADN** au départ, on obtient la courbe bleue (tout à gauche). Si on part avec une concentration de **0,1 microgramme** (courbe rouge), soit **10 fois moins d'ADN**, on voit qu'il faut **plus de cycles** pour que la fluorescence soit **détectable** : la courbe se décale vers la droite !





Au **moins** on met d'ADN au départ, au **plus** il faut de cycles pour que la fluorescence soit **détectable par le système**. Mais, **après 40 cycles**, l'**intensité des produits PCR** générés est **identique** quelle que soit la quantité d'ADN génomique utilisée au départ !

Si on détermine la quantité de fluorescence au **point d'inflexion**, c'est-à-dire au niveau du **cycle seuil Ct**, l'écart entre chaque courbe est **proportionnel** à la **quantité de départ** :



Avec la **gamme étalon**, on obtient une relation **linéaire**. On peut **quantifier le milieu réactionnel** et connaître exactement la **quantité d'ADN de départ** !

### C. Utilisation des sondes TaqMan :

En dehors du **SYBR Green**, une autre technologie plus **spécifique** peut aussi être utilisée : les **sondes TaqMan**. C'est une seconde technique de **PCR en temps réel** où l'on rajoute un **troisième fragment** (simple brin) dans la réaction en plus des deux primers de la PCR.

Ce fragment va être **complémentaire** de la région d'ADN à amplifier, où il ira **s'hybrider**. Cette sonde possède à ses extrémités un **quencher** et un **fluorochrome**.

**Sonde TaqMan = quencher + fluorochrome**

Le **quencher** a pour rôle d'**éteindre la fluorescence du fluorochrome** de la sonde TaqMan quand ils sont tous les deux **physiquement proches** sur la sonde.

Lors de l'étape d'**élongation**, la **polymérase synthétise** le brin complémentaire de 5' en 3'. Elle possède également une **activité 5'-3' exonucléasique** qui lui permet de **grignoter la sonde TaqMan** quand elle rentre en contact avec celle-ci. Elle va donc « libérer » le fluorochrome du quencher, et en **séparant ainsi les deux**, il pourra **fluorescer** !

Là encore la **mesure de la fluorescence** se fait à la **fin de la phase d'élongation**, à la **fin de chaque cycle**.

Et la **quantité de fluorescence générée** sera également **proportionnelle à la quantité d'ADN au départ**.

