

Biologie Moléculaire



Outils de base et comparaison de génomes

I. Généralités

La biologie moléculaire utilise des outils/techniques variées

Les outils de Biologie moléculaire permettent d'**identifier** et d'**analyser** les gènes, voire de les **remplacer** !

A. Ils permettent de manipuler l'ADN

Nous allons faire la liste des manipulations que l'on sait effectuer sur l'ADN :

- **Couper** : Les **endonucléases de restriction** coupent l'ADN au niveau de **séquences spécifiques**.
- **Copier** : Les **polymérases** constituent une famille d'enzymes chargée de copier, comme l'**ADN Polymérase** pour la **réplication**, ou la **Polymérase delta/epsilon** pour la **transcription**.
Il existe aussi des **vecteurs de clonage**.
- **Coller** : Les **ligases** relient entre eux des fragments d'ADN.
- **Rechercher** : Les **sondes d'hybridation** (de courtes séquences d'ADN ou d'ARN couplées à un **marqueur**) repèrent une **cible** dans le **génome** et déclenchent un **signal** en s'y **hybridant**.

B. Ils permettent d'analyser et de lire des séquences génomiques

Quand on veut travailler sur de l'ADN, il est systématiquement nécessaire d'**amplifier** les **séquences génomiques**, car elles sont trop **petites** pour identifier une éventuelle **mutation** : le risque de **contamination** de l'échantillon est **trop élevé**, la modification de la séquence ne serait **pas visible**.

Amplifier un gène consiste à faire un grand nombre de **copies** d'une même séquence d'ADN pour ensuite la manipuler.

Nous avons à notre disposition **deux méthodes** pour **amplifier** une portion d'ADN :

- Technique de **clonage** d'ADN dans un **vecteur** (ex. plasmide).
- Technique de **PCR** (*Polymerase Chain Reaction*) qui peut être suivie d'un **séquençage**.

Une autre méthode permet uniquement de **déterminer la séquence** nucléotidique d'une portion d'ADN :

- **Séquençage haut débit**, puis le séquençage de nouvelle génération **NGS** (*Next Generation Sequencing*), non traités ici.

II. Techniques d'amplification de l'ADN

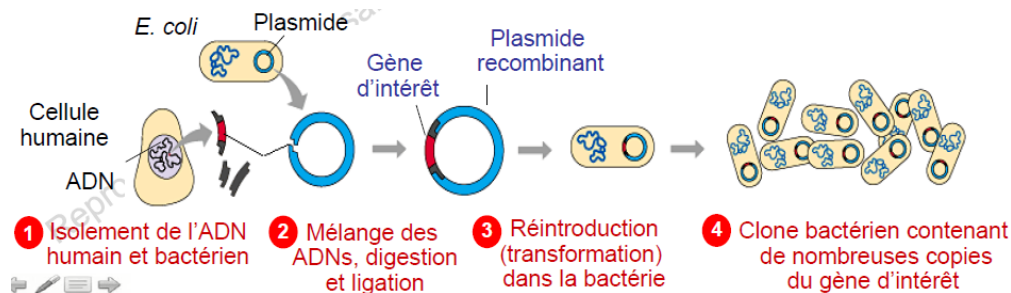
A. Clonage par vecteur

La technique de clonage utilise un vecteur (~ 1970)

Un **vecteur** est une molécule d'ADN qui permet l'introduction d'ADN **étranger** dans une cellule, sa **réplication** voire son **expression** !

Ex : plasmide bactérien.

- Le **gène d'intérêt**, ou **insert** (en **rouge**) est introduit dans un plasmide après **digestion** (= coupure) des ADNs par des **enzymes de restriction** et **ligation**.
- Le **plasmide recombinant** est introduit dans des bactéries par **transformation**.
- Les bactéries prolifèrent et « amplifient » le **plasmide** et le **gène d'intérêt**.



B. PCR (Polymerase Chain Reaction)

L'autre technique majeure est la technique de PCR (~1980)

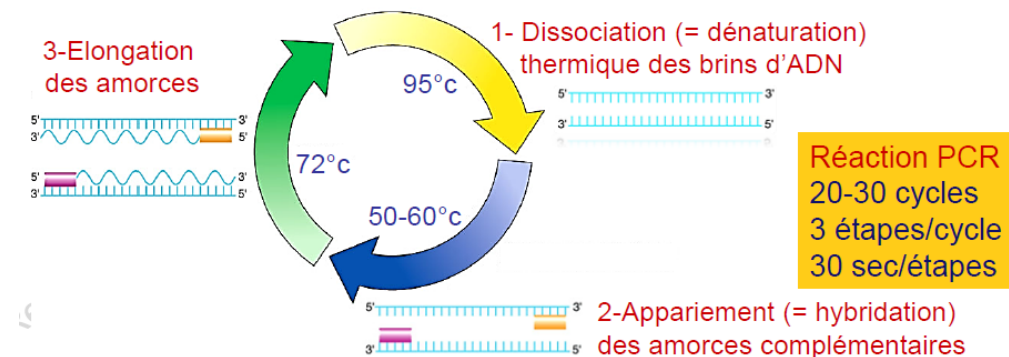
La PCR permet l'**amplification exponentielle** d'une **séquence spécifique** d'ADN.

La réplication *in vitro* d'une copie sur **n cycles** donne **2 puissance n** copies de la cible !

- **Thermocycleur** : Appareil faisant varier la **température** de façon **cyclique**.
- **Réactifs** : ADN contenant la cible, **ADN Polymérase thermostable** (Taq Polymérase), **dNTPs** et **primers** (= amorces de synthèse) encadrant la cible.

Etapes de la PCR :

1. **Dénaturation** (dissociation des brins d'ADN) à **95°C**.
2. **Hybridation** (appariement des amorces) à **50-60°C**.
3. **Elongation** des amorces à **72°C**.



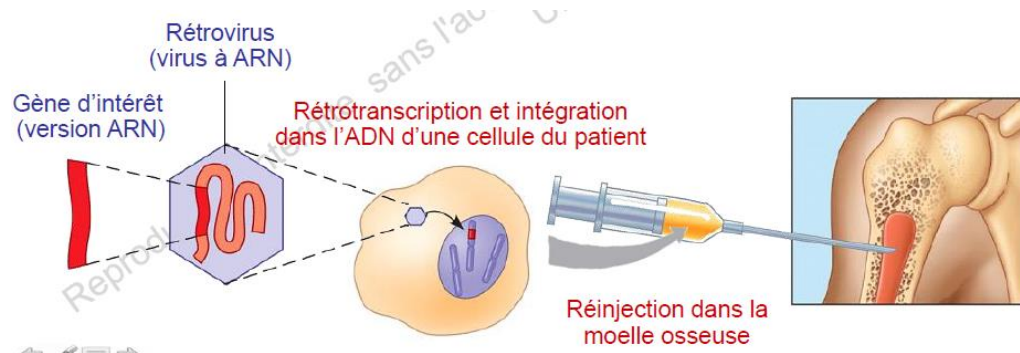
C. Applications

1. Applications du clonage :

Les applications du clonage sont nombreuses

En médecine, il est utilisé à des **fins thérapeutiques** :

- Production de **médicaments recombinants**, de **vaccins**...
- **Insuline**, **hormone de croissance**, **facteurs de coagulation**...
- **Thérapie génique** (leucémies...)
- Un **virus** utilisé comme **vecteur** peut introduire un gène dans une cellule et **remplacer un gène défectueux** ou la rendre **sensible** à un **médicament** !



2. Applications de la PCR :

L'analyse d'ADN après PCR à de multiples applications

○ Diagnostic génétique des maladies héréditaires :

- Diagnostic **post-natal**, **prénatal** (interruption de grossesse), **préimplantatoire** (fécondation *in vitro*) avec analyse sur **quelques cellules** seulement.

○ Diagnostic en virologie, parasitologie, bactériologie, criminologie, etc...

- **Détection** et **quantification** de virus, parasites, ou bactéries.
- **Identification** de coupables par l'ADN retrouvé sur les scènes de crime...

3. Séquencer le génome :

Ces techniques ont permis de lire la séquence du génome

Ex : Séquençage Sanger par PCR et électrophorèse capillaire (1990)

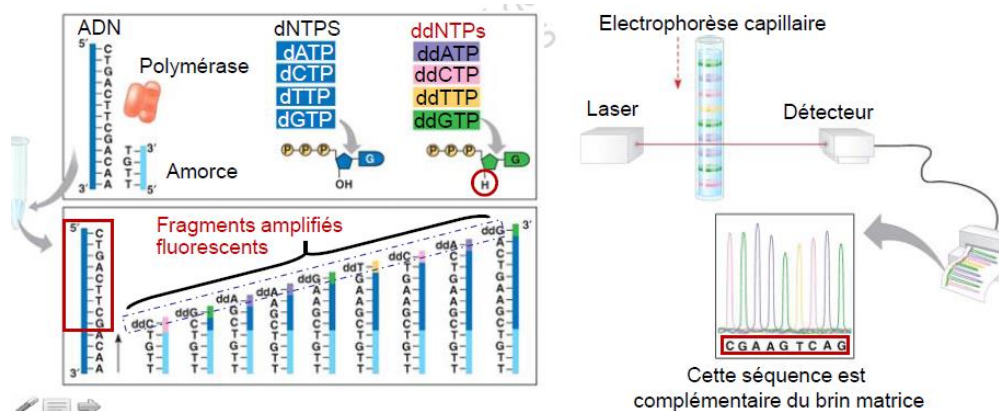
Cette technique utilise des **terminateurs de chaîne** (**dd**idésoxyribonucléotide, **dd**NTP) fluorescents !

La réaction de PCR **s'arrête** à chaque incorporation possible d'un **ddNTP** (la polymérase ne pourra plus ajouter de nouveaux nucléotides après).

On obtient un **mélange de fragments** dont la taille diffère d'un nucléotide et dont la **fluorescence** indique chaque position d'un dNTP.

Chacun des **quatre types** de ddNTPs (dd**A**TP, dd**G**TP, dd**T**TP, dd**C**TP) possède une couleur de fluorescence qui lui est **propre**.

La séparation en taille des fragments par **électrophorèse** puis la **lecture de la fluorescence** nous permet de déterminer la **séquence** nucléotidique.



**Le génome d'autres organismes est séquencé :
procaryotes, eucaryotes monocellulaires et pluricellulaires**

Le génome de l'**homme** et du **chimpanzé** sont identiques à **98,6%** !

Mais la comparaison des génomes (génomique comparative) peut-elle expliquer les différences entre espèces ?

B. Les questions fondamentales de la génomique comparative

**Quelle(s) différence(s) fondamentale(s) entre
procaryotes et eucaryotes ?**

Le **nombre de gènes** est à peu près similaire, il est donc **sans rapport** avec la **complexité** d'un organisme !

Et entre eucaryotes monocellulaires et pluricellulaires ?

Le rapport **nombre de gènes / taille du génome** décroît : plus un organisme est **complexe**, moins son génome est **riche** en gènes !

	Taille du génome (10 ⁶ bases)	Nombre Gènes	Nombre gènes / Taille du génome
Procaryotes (<i>E. coli</i>)	4.6	~ 4000	~ 1/1000 pb
Eucaryote monocellulaire (<i>S. cerevisiae</i>)	12	~ 6,000	~ 1/2000 pb
Eucaryote pluricellulaire (<i>C. elegans</i>)	100	~ 14,000	~ 1/7000 pb
Homme	3 000	~ 30,000	~ 1/100 000 pb

III. Génomique comparative

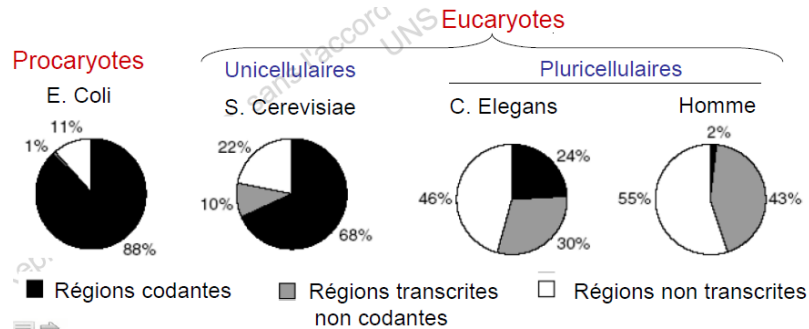
A. Généralités

Le génome humain contient 3 milliards de paires de bases

- Le nombre de gènes est estimé entre **20-30 000** :
- La fonction de **plus de 50%** de ces gènes est **inconnue**.
- Les **séquences codantes** représentent **< 5%** du génome.
- La fonction de **95%** du génome est **non / mal connue**.
- Le génome de deux individus est identique à **99,9%** :
- Ces différences neutres appelées **polymorphismes** expliqueraient certaines différences entre individus.

Quelle(s) différence(s) fondamentale(s) entre procaryotes / eucaryotes ?

Elle réside dans la proportion de **séquences non codantes** dans le génome : paradoxalement, plus un organisme est **complexe**, **moins** son génome contient de **séquences codantes**, **plus** il contient de **séquences non codantes**, transcrites ou non !



C. Les réponses de la génomique comparative

La différence réside dans la proportion des séquences non codantes

1. Séquences non codantes transcrites :

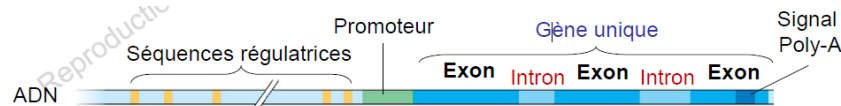
Les séquences **non codantes transcrites** correspondent aux **introns**.

Les gènes **procaryotes** sont regroupés en **opérons** et **dénués d'introns** !



Les gènes **eucaryotes** sont régulés **individuellement** et **morcelés** (introns).

Les introns sont **transcrits** avant d'être **éliminés** ce qui explique l'abondance de **séquences non codantes transcrites** des génomes eucaryotes.

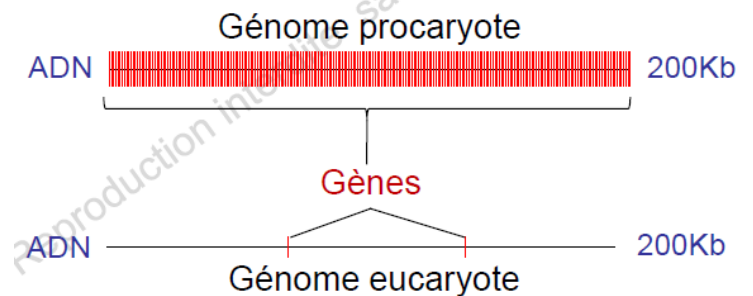


2. Séquences non codantes non transcrites :

Les séquences **non codantes non transcrites** sont les **régions intergéniques**.

Le génome **procaryote** est **très compact** (peu de régions intergéniques) : la **densité** de gènes est **élevée** (un gène toutes les 1000 bases) !

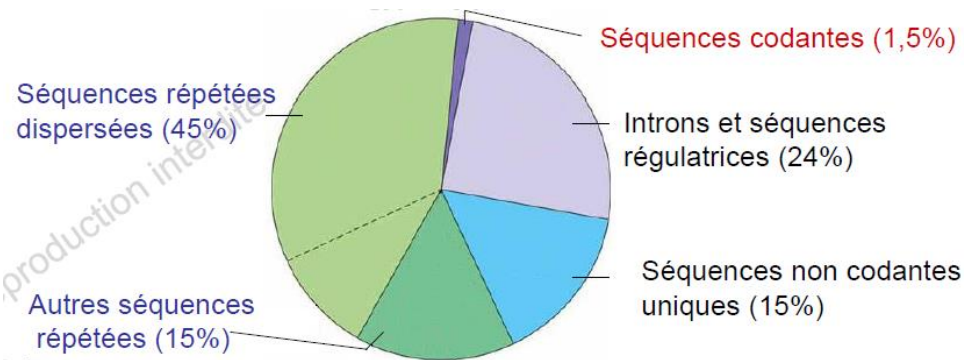
Le génome **eucaryote** contient de **vastes régions intergéniques** («désert») : la **densité** de gènes est **faible** (un gène toutes les 100 000 bases) !



3. Séquences répétées du génome :

Les régions **intergéniques** sont en majorité des **séquences répétées** dont :

- Les **séquences répétées dispersées** (45% du génome humain)
- **Transposons** et **rétrotransposons** (séquences *L1* ou *Alu*)
- D'autres séquences répétées (15%) sont des séquences **répétées en tandem** (5%) ou des **duplications** de larges séquences génomiques (5%)



4. Cas des introns :

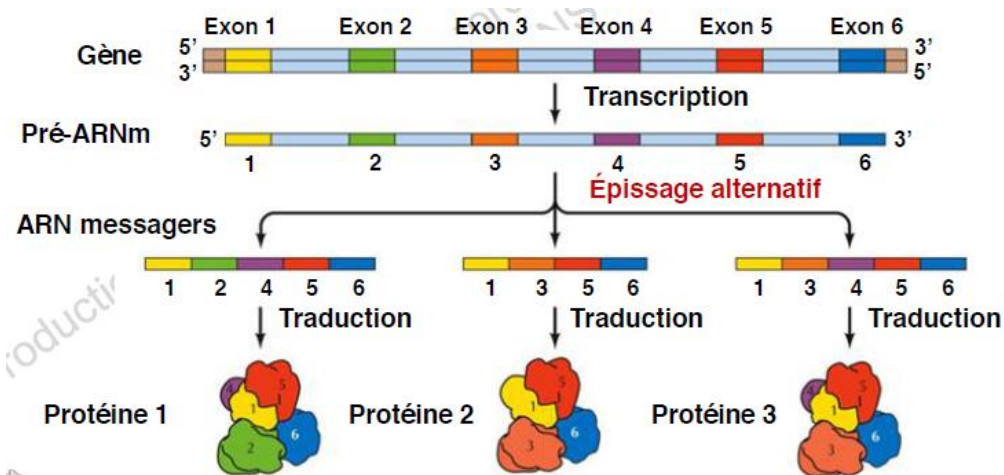
Les introns ont favorisé la complexification des espèces

Leur existence est à la base du phénomène d'épissage alternatif.

Ils permettent donc de produire **plusieurs protéines** à partir d'**un gène**.

Chez l'homme, **20-30 000 gènes** codent pour **200 000 protéines** différentes.

C'est le **nombre de protéines** d'un organisme qui reflète sa **complexité** !



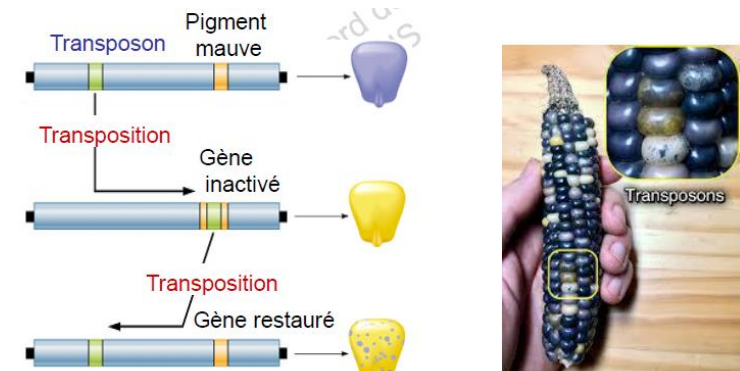
5. Cas des transposons :

Les transposons peuvent inactiver des gènes

Ils se **déplacent** et se **multiplient** dans le génome (gènes « sauteurs »).

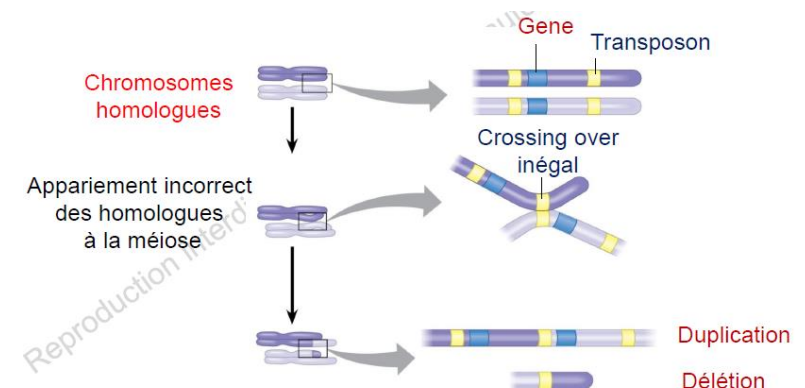
Ils peuvent **déplacer** des **exons** ou des **gènes entiers**, ou **inactiver** des gènes !

Exemple : Les variations de **couleur** des **grains de maïs** sont à l'origine de la **découverte des transposons** (Barbara McClintock, 1947).



Ils favorisent les crossing-over inégaux entre chromatides

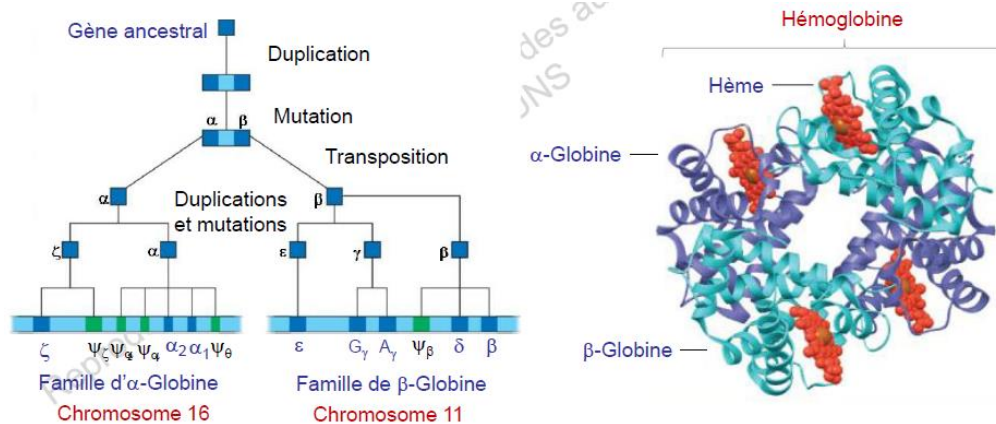
Le résultat est d'un côté un chromosome avec **délétion** d'une région et des gènes qu'elle contient, et de l'autre côté un chromosome avec **duplication** de cette même région.



Certains gènes ont été dupliqués et forment des familles multigéniques

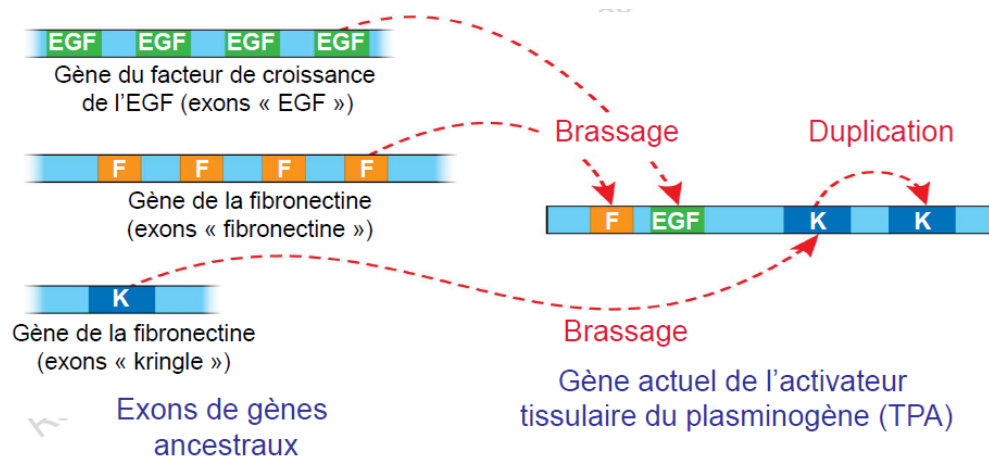
Elles sont issues de **duplications** et **mutations** à partir d'un gène ancestral.

Ex: Famille des gènes codant les chaînes de **globine** de l'hémoglobine (trois gènes de la famille **alpha**, cinq de la famille **bêta** et cinq **pseudo-gènes**).



Les transposons ont permis la création de nouveaux gènes

Un **exon** peut par exemple être **déplacé** d'un gène à un autre par l'intermédiaire des **transposons** qui l'**encadrent**.



D. Conclusion

Introns et séquences répétées ont favorisé l'évolution

A partir des **premières formes de vie** (nombre minimal de gènes), l'**évolution des génomes** a contribué à l'apparition de **nouvelles espèces**.

Séquences répétées et non codantes seraient à la base de l'évolution des espèces

Grâce aux **duplications**, **réarrangements**, **mutations**, **épissage alternatif**...

En contrepartie, elles peuvent favoriser l'apparition de **maladies génétiques**.

IV. Points clés

- Parmi les outils de la Biologie moléculaire
 - La réaction de **PCR** est à la base de nombreux procédés de **diagnostic**.
 - Elle permet de détecter des **mutations connues ou non** et est notamment une étape préalable au séquençage du génome.
- Grâce à ces outils, le génome de différents organismes est séquencé
 - Les génomes **eucaryotes** et le génome **humain** contiennent **peu** de séquences **codantes** et de **nombreuses** séquences **non codantes**.
 - Les séquences **non codantes introniques** favorisent la **diversité** et la **complexité** des organismes en permettant un **épissage alternatif**.
 - Les **séquences répétées** (transposons, etc.) ont **favorisé l'évolution** du génome mais peuvent être la source de **mutations pathogènes**.