L’avènement du séquençage haut débit

Présenté par Dr. Philippe GLASER (Directeur Genopole Pasteur)

# Historique

Depuis les années 1977, le séquençage des acides nucléiques est réalisé par la méthode Sanger, grâce aux terminateurs de chaînes.

C'est par cette méthode que le consortium international à réaliser le séquençage du génome humain entier, qui a débuté en 1989. Ce projet a été achevé en 2003 pour un coût global de près de 3 milliards de dollars.

Cependant, cette méthode a un coût élevé à la base séquencée et ne permet pas de générer un nombre important de séquences.

C'est dans l'optique de répondre à une demande grandissante que sont alors apparus les séquenceurs de nouvelle génération.

# Les connaissances du génome humain

De par l'arrivé des séquenceurs de nouvelles générations, les connaissances sur les génomes et particulièrement le génome humain, ont considérablement évoluées.

En effet, au cours des dernières années, le nombre de gènes composant le génome humain à énormément évolué. On estime aujourd'hui que le génome humain est composé d'environ 23 000 gènes et que seulement 1 à 2% du génome coderait des protéines.

Cet essor technologique a notamment permis de considérablement augmenter nos connaissances sur le génome humain. Cela a notamment permis de découvrir de nouveaux gènes, de faire de la prédiction de fonctions, de mieux comprendre les voies de signalisation et métaboliques ainsi que de mieux comprendre le protéome et le transcriptome.

# Le séquençage à haut débit

### Le début d'une nouvelle ère

C'est à partir de 2007 que les séquenceurs de nouvelles générations ont envahis le marché mondial.

L'apparition de "modèles de paillasses" à coût plus faible a notamment permis d'équiper de nombreux centres et laboratoires.

La possibilité de séquencer massivement, rapidement et à faible coût a considérablement bouleversé le monde de la recherche. Dans un futur très proche, on espère pouvoir séquencer un génome entier pour moins de 1 000$...

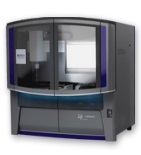
Les applications du séquençage haut débit sont diverses :

- le séquençage *de novo*

-le reséquençage de génomes (mutations, SNP, Indel, CNV, …)

- la quantification de molécules (RNA-seq, ChIP-seq, ..)

Il existe plusieurs technologies sur le marché, dont les principales sont :



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Séquenceur | 454 GS FLX | HiSeq 2000 | SOLiD 5500XL | Ion Proton System |
| Temps run | 10 heures | 8 jours | 10 jours | 2 heures |
| Taille des lectures (pb) | 1000 | 2x100 | 2x75 | 300 |
| Nombre de lectures | 1.106 | 3.109 | 1.4.109 | 3.106 |
| Données générées | 1 Gb | 600 Gb | 300 Gb | 1 Gb |
| Amplification | PCR en émulsion | PCR "Bridge" | PCR en émulsion | PCR en émulsion |
| Réaction séquençage | Pyroséquençage | Terminateur de chaîne réversible | Ligature | "proton-séquençage" |

Toutes nouvelles technologies de séquençages s'appuient sur une amplification clonale d'un fragment d'ADN, permettant ainsi de s'affranchir du biais dû à l'étape de clonage présente dans la méthode Sanger.

### Principe de la technologie 454

* **Préparation de la banque**

L’ADN que l'on souhaite séquencé est dans un premier temps fragmenté par nébulisation ou sonication afin d'obtenir des fragments d'environ 300pb. Les extrémités cohésives créent lors de la coupure vont être réparées afin d’obtenir des extrémités franches permettant l’ajout des adaptateurs.

Par le biais d'une ADN ligase, on ajoute ensuite des adaptateurs qui contiennent notamment une séquence MID (Multipled IDentifier) qui permettra d'identifier chaque échantillon lors du séquençage. L'adaptateur B a la particularité d'être biotinylé à son extrémité 5' afin de les purifier par la suite.

**Adaptateur A** : **Amorce A** – **Clé** – MID

****

**Adaptateur B** : **Amorce B** – **Clé** – MID

****

On purifie ensuite nos ADN pour ne conserver que les fragments bordés par au moins un adaptateur B à leurs extrémités. Cette purification est réalisée sur des billes magnétiques recouvertes de streptavidine qui vont fixer spécifiquement les adaptateurs B biotinylés.

Les fragments d'ADN bordés par deux adaptateurs A seront donc éliminés, tandis que les fragments bordés par deux adaptateurs B seront conservés mais éliminés lors de la PCR en émulsion.

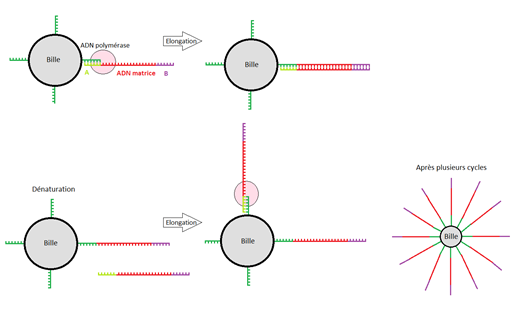
* **Amplification clonale par PCR en émulsion**

On réalise ensuite une PCR en émulsion afin d'amplifier une molécule d'ADN en plusieurs millions de copies.

Pour cela, des sondes complémentaires de l'adaptateur A sont fixées sur des billes magnétiques afin de réaliser la PCR en émulsion. Seuls les fragments ayant un adaptateur A à une extrémité seront donc fixés sur les billes et amplifiés. Les fragments ayant des adaptateurs B et B à chaque extrémité seront donc éliminés.

Cette PCR est réalisé dans un mélange eau/huile afin de former des microgouttelettes (ou microréacteurs) contenant les billes et l'ensemble des "acteurs" de la PCR. Afin de réaliser une amplification clonale, on travaille en conditions limitant, de telle manière à obtenir 0 ou 1 fragment d'ADN par microréacteur.

Un enrichissement est ensuite réalisé sur des billes recouvertes de streptavidine qui vont permettre de purifié uniquement les billes où il y a eu une molécule d'ADN et donc amplification.

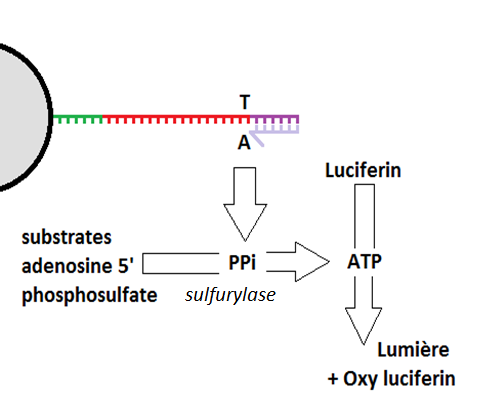


* **Réaction de séquençage : le pyroséquençage**

La réaction de séquençage dans la technologie 454 se fait par pyroséquençage.

Les 4 bases de l'ADN sont ajoutées de façon cyclique : une base après l'autre avec des étapes de lavage entre chaque base. Lorsque la base ajoutée est complémentaire du brin matrice, il y aura ajout de la base et libération d'un PPi qui sera transformé en ATP par une sulfurylase. Cet ATP est ensuite capté par la luciférase pour émettre un signal luminescent qui sera ensuite détecté par une caméra CCD.

Un des inconvénients de cette technologie (qu'on retrouve dans la plupart des autres technologies) est que dans des régions homopolymériques, le signal capté par la caméra CCD n'est pas totalement proportionnel au nombre de bases ajoutées.



### Un pas vers le futur : les séquenceurs de 3ème génération

L'objectif futur est désormais de s'affranchir du seuil de détection élevé et donc d'une amplification par PCR.

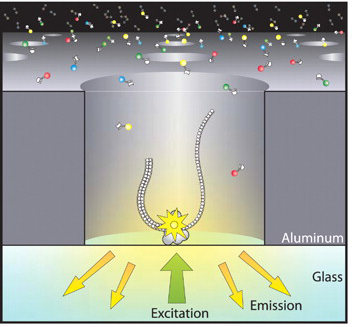
C'est chose faite avec l'apparition des séquenceurs de 3ème génération permettant de séquencer en temps réel une molécule unique d'ADN, d'ARN voire protéique!

Cela permet de séquencer des séquences beaucoup plus longues (de l'ordre du kb) et en un temps beaucoup plus court (quelques minutes à une heure).

La société Pacific Biosciences à développer un modèle commercial basé sur la technologie SMRT (Single Molecule Real Time).

Cette technologie s'appuie sur fixation de l'ADN polymérase sur un support de plaque à fond non opaque. Chaque base est marquée par un flurochrome différent, et l'incorporation est suivie en temps réel par une caméra CCD placée sous le support.

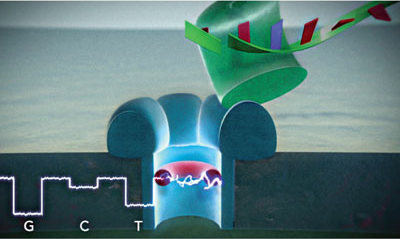
Le principal défaut de cette technologie est son taux d'erreur (15%!) mais qui est "compensable" par une couverture plus élevée.





La seconde technologie des séquenceurs de 3ème génération est celle des nanopore réalisé par la société Oxford Nanopore.

L'ADN passe le long d'un pore qui est formé par une première protéine qui permet de séparer les deux brins d'ADN. Puis le passage de l'ADN simple brin au sein de la seconde protéine provoque un courant électrique caractéristique de la base qui à traversée le pore.

ONTIl n'y a pas encore eu de commercialisation pour cette technologie qui est encore à l'état de prototype.