



Opinions Libres

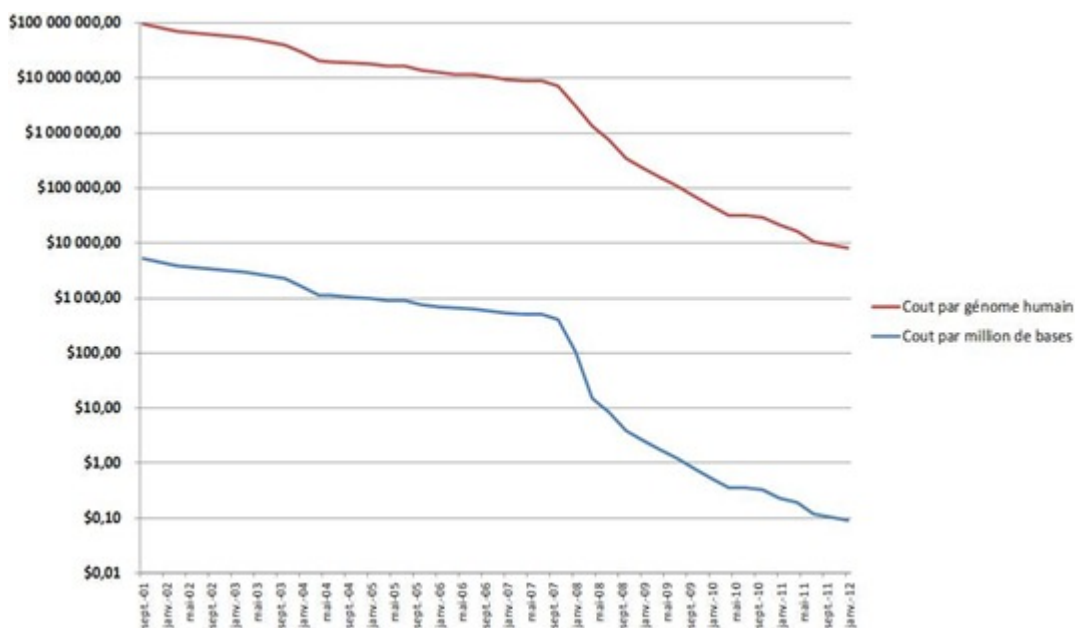
le blog d'Olivier Ezratty

Les technologies de séquençage du génome humain – 3

Dans l'épisode précédent de notre série estivale, nous avons décortiqué le processus qui précède le séquençage de l'ADN et toutes les étapes de préparation. Avec la sélection des cellules, leur préparation pour en extraire les longues molécules d'ADN, leur pré-découpage pour simplifier le processus du séquençage et leur multiplication avec la PCR pour augmenter la taille de l'échantillon disponible pour le séquençage. La plupart de ces étapes précèdent le séquençage mais nous verrons que certaines méthodes permettent de simplifier le processus en évitant notamment la PCR.

Nous passons maintenant au cœur du sujet, le séquençage, sachant que nous allons à la fois expliquer les basiques du séquençage avec les principales méthodes utilisées et aussi nous attarder sur les dernières générations de séquenceurs massivement parallèles.

Les technologies de séquençage ont été créées au milieu des années 1970 et ont été continuellement améliorées depuis. Au départ, le processus de séquençage, popularisé par la méthode de Sanger que nous allons décrire était lent et générateur d'erreurs. Les techniques et machines qui sont apparues depuis ont à la fois automatisé le processus, apporté des variantes, parallélisé le séquençage et réduit les erreurs. Résultat, alors qu'il a fallu plusieurs années et des centaines de millions de dollars pour procéder au premier séquençage générique du génome humain en 2000, on peut obtenir le même résultat en moins d'une journée et pour quelques milliers de dollars, l'objectif étant de descendre rapidement en dessous de \$100 par génome.



Un concours a ainsi été lancé qui rappelle les belles heures de l'aviation, le **Archon Genomics Xprize**. \$10m récompenseront l'équipe qui créera un appareil permettant de séquencer 100 génomes humains de centaines

en moins de dix jours et pour moins de \$1000 par personne. L'idée est de découvrir les gènes qui protègeraient de certaines maladies et assureraient la longévité humaine. Ce prix est lancé par la **XPrize Foundation**, une association non lucrative américaine qui a dans son board des personnalités aussi diverses que Larry Page (Google), Ray Kurzweil (entre autres, cofondateur de la **Singularity University**), Ariana Huffington, le réalisateur James Cameron, le biologiste **Craig Venter**, l'astronaute d'origine iranienne Anousheh Ansari, Ratan Tata (du groupe indien Tata) et Will Wright, le chief designer officer d'Electronics Arts. Le prix est sponsorisé par Express Scripts, une société de services dans la santé aux USA. L'une des sociétés à participer à ce concours est **Ion Torrent** dont nous décrirons plus loin la machine, celle-là même qui avait attiré mon attention lors du CES 2012 !

Mais avant d'en venir là, il faut remonter aux sources et notamment à l'une des premières méthodes de séquençage...

Méthode de Sanger

La méthode de séquençage de l'ADN développée par Frederick Sanger au milieu des années 1970 a été la plus répandue pendant près d'une vingtaine d'année. Elle concurrençait au départ celle de Maxam et Gilbert créée en 1976-1977. Cette dernière s'est rapidement éclipsée car elle s'appuyait sur des produits toxiques et sur la meure de radioactivité. Mais les principes étaient voisins d'un point de vue conceptuel et on les retrouve jusqu'à aujourd'hui dans certaines technologies de séquençage massivement parallèles.

Sanger est l'un des quatre scientifiques à avoir gagné deux prix Nobel dans l'histoire. Ici, il s'agissait du prix Nobel de chimie : l'un en 1958 pour la découverte en 1951 de la séquence d'acides aminés qui constitue la molécule d'insuline (bovine) et l'autre en 1980 pour le procédé de séquençage qui porte son nom, inventé en 1977.

Amorçage

Le séquençage s'applique à un brin d'ADN de taille moyenne, de quelques centaines de bases, qui a été obtenu par différentes méthodes. On utilise couramment la PCR que nous avons décrite dans l'**article précédent** de cette série, qui est la méthode la plus récente et qui permet de démultiplier plusieurs dizaines de millions de fois un petit échantillon de brin d'ADN, lui-même généré par une extraction d'un ADN complet avec une enzyme de restriction. Mais on peut aussi utiliser d'autres méthodes comme le clonage de brins d'ADN avec des vecteurs de clonage divers comme l'ADN du bactériophage M13.

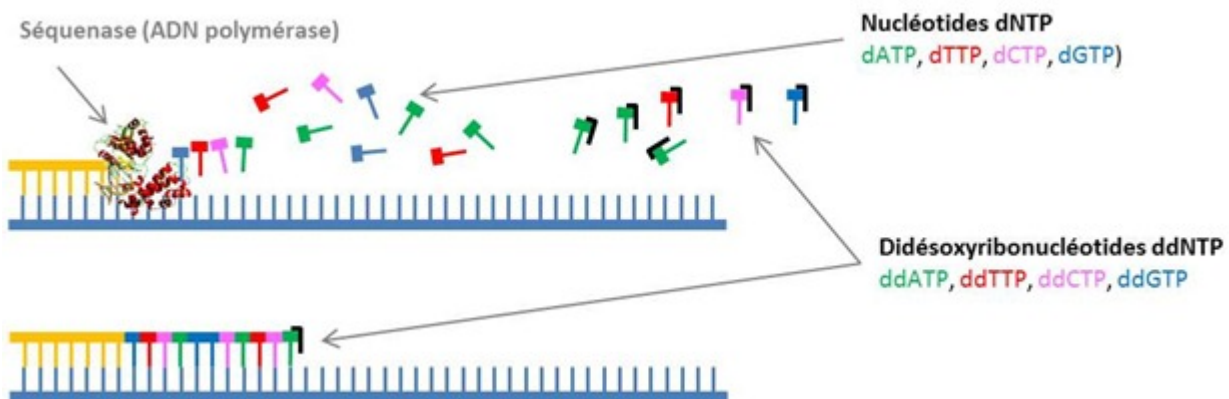
Comme pour la PCR, on commence par dénaturer par la chaleur (séparer les deux brins) d'un échantillon d'ADN en solution. L'autre brin d'ADN résultant de la dénaturation n'est pas utilisé. Le brin d'ADN doit être en quantité suffisante, à savoir que l'on en dispose de plusieurs dizaines de millions. La quantité est rendue nécessaire par le moyen de détection des séquences de base utilisé dans la méthode. On ajoute au brin une amorce de 18 bases et quelques qui ne fonctionne que dans un sens (dit "montant") en s'accrochant au bout des brins d'ADN à répliquer et qui servira de point de départ pour la copie du reste du brin. L'amorce est le "négatif" au niveau des bases du début du brin d'ADN à répliquer. Dans la PCR, on utilise des amorces "montantes" et "descendantes" pour répliquer les deux brins d'ADN dénaturé.



Polymérisation

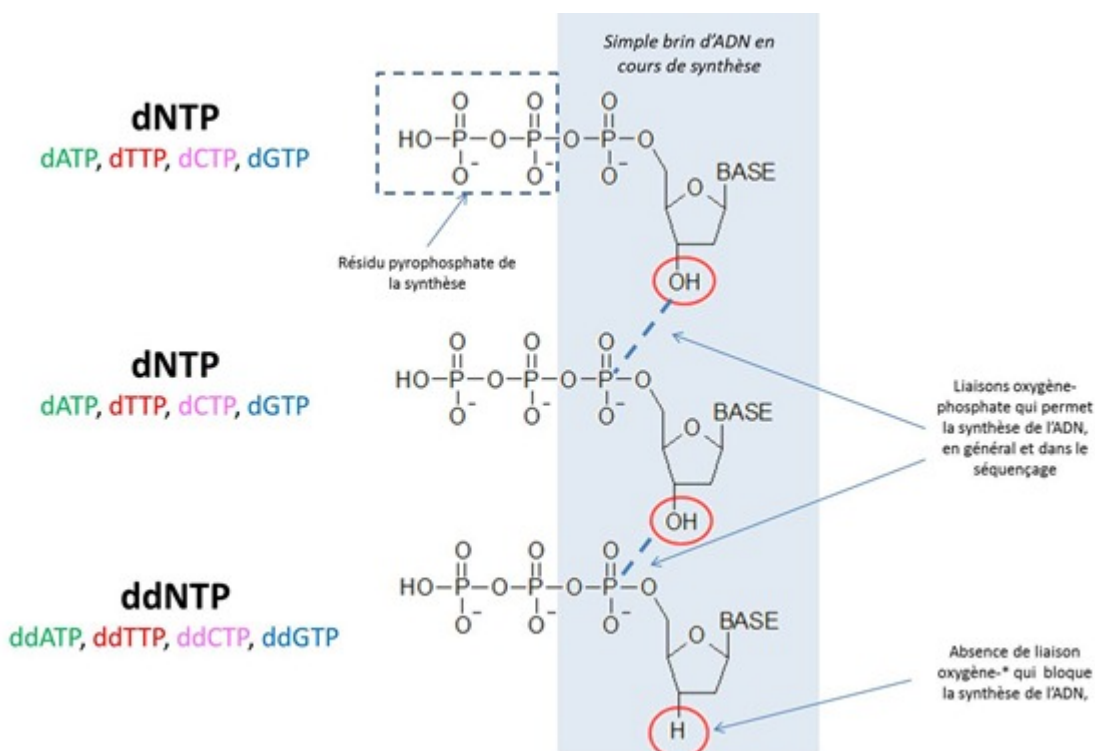
C'est la réaction chimique la plus importante du processus : elle permet le clonage partiel du brin d'ADN à qui on a ajouté son amorce.

La polymérisation s'appuie sur l'emploi d'une ADN polymérase qui assemble des nucléotides au brin à compléter. Ces nucléotides sont des molécules dNTP ou désoxyribonucléotides qui associent une des quatre bases de l'ADN (d'où les appellations dATP, dTTP, dCTP et dGTP), un sucre (pentose) et un triphosphate. Ces dNTP s'associent aux nucléotides complémentaires du brin d'ADN à séquencer sous l'effet de l'ADN polymérase. On appelle cette phase du processus la SBS, "Sequencing by Synthesis". Elle utilisée dans tout un tas de méthodes différentes de séquençage.



Fin de polymérisation

La méthode consiste à arrêter cette réaction de polymérisation à chaque position de nucléotide dans le brin. On utilise pour cela des molécules ddNTP ou didésoxyribonucléotides (ddATP, ddTTP, ddCTP, ddGTP). Ces nucléotides bloquent la polymérisation de l'ADN du fait d'une liaison OH qui est remplacée par une liaison H par rapport aux nucléotides équivalentes. La liaison OH s'associe au premier P (phosphore) du dNTP suivant lors de la polymérisation. Dans le O, pas de polymérisation. On met beaucoup moins de ddNTP que de dNTP pour que la polymérisation se poursuivre tant que faire ce peut. On contrôle aussi sa vitesse avec la température.



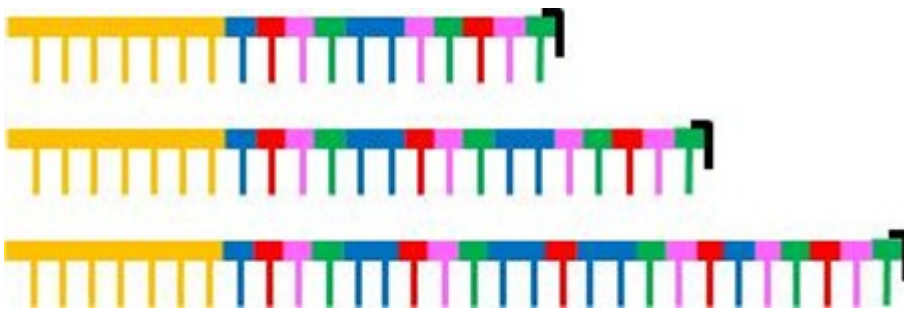
Les ddNTP sont marqués par “fluorescence” pour que l’on puisse les détecter plus loin dans le processus. Le label fluorescent est du S-35 thiophosphate, ou 35S-dATP, qui s’associe aux ddNTP du côté des phosphates sans empêcher la liaison OH-P d’opérer avec le dNTP précédant dans le processus de polymérisation. La proportion ddNTP sur dNTP est généralement de 1%.

Dans ce processus, la polymérisation dure une quinzaine de minutes.

Dénaturation

Une fois encore, on va procéder par chauffage pour séparer les brins d’ADN répliqués et les répliqués partiels obtenus par polymérisation que l’on souhaite conserver. Il reste dans la solution une soupe avec : nos brins d’ADN négatifs non répliqués, les brins entiers répliqués, le relicat de dNTP et de ddNTP ainsi que l’ADN polymérase sans compter les reliquats de phosphores issus de la polymérisation (les dNTP libèrent un pyrophosphate lors de la polymérisation, cf le schéma précédent, ce pyrophosphate ne sert à rien ici, mais on l’utilise dans le pyroséquençage dont nous parlerons plus loin).

On se retrouve avec une soupe dont des brins d’ADN synthétisés avec une extrémité fluorescente qu’il va falloir trier par taille pour reconstituer leur séquence de bases (A, T, C, G). Mais auparavant, on aura éliminé par traitement chimiques divers les résidus cités ci-dessus.



Tri sur gel

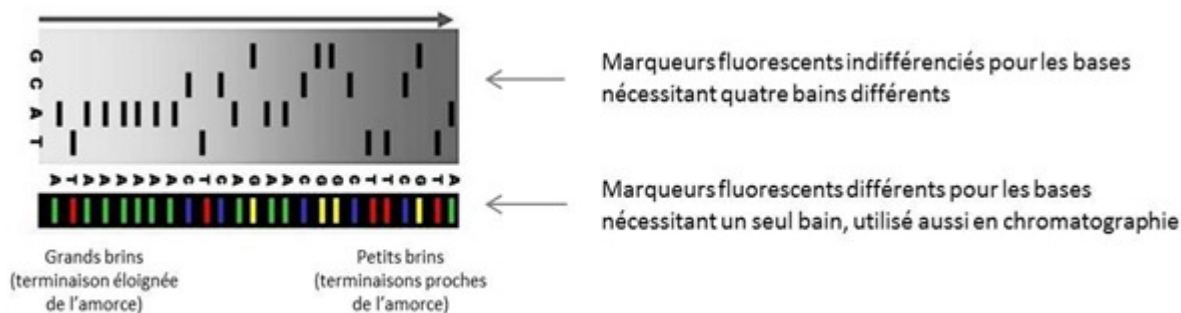
Dans la méthode de Sanger, on procède généralement au tri avec une électrophorèse sur une solution contenant 1% de gel d’agarose (extrait de l’agar-agar, déjà évoqué dans cet article...). Il se trouve que les molécules d’ADN sont chargées négativement et en fonction de leur longueur. On applique une tension électrique continue dans une solution contenant ce gel d’agarose et à plat sur un filtre en nylon. La tension va faire migrer les molécules d’ADN vers le + au niveau des phosphates et en fonction de leur taille, le tout à 50°C. Les petites vont aller le plus loin. Le gel d’agarose est une sorte d’amortisseur ou de régulateur de ce processus de migration.

Les molécules les plus courtes correspondront aux bases qui étaient proches de l’amorce (primer) du processus. Le gel d’agarose est associé à différentes autres molécules dont de l’urée qui évite que les molécules d’ADN se replient sur elles-mêmes. Une fois la migration terminée, on va fixer l’ensemble par ultraviolet.

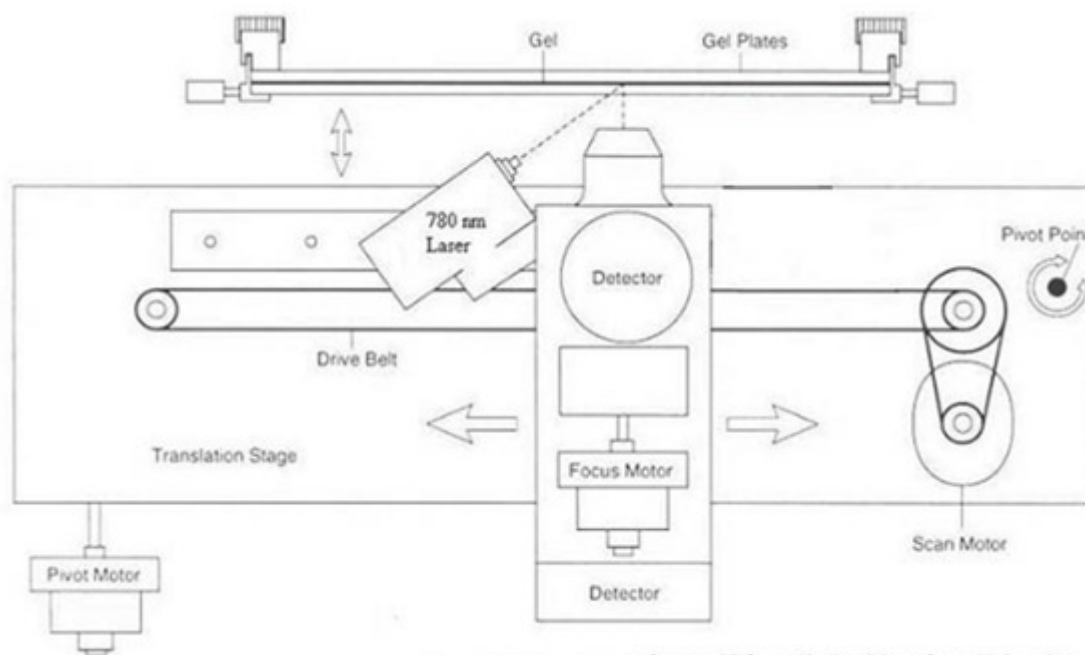
Les molécules fluorescentes de ddNTP ne sont pas visibles à l’œil nu. On va les révéler à l’aide d’un laser qui va les exciter et les faire émettre de la lumière dans une couleur différente. Une caméra enregistre alors l’ensemble du gel.

Un gel fait quelques dizaines de centimètres de longueur et permet d’identifier quelques centaines de bases. C’est lié à la résolution du papier ! D’où le fait que la méthode de Sanger est assez lente. Cette phase d’électrophorèse sur gel dure plusieurs heures ! Elle a bien été utilisée pour séquencer différents bactéries et cellules ainsi que l’ADN humain dans le Human Genome Project, mais a alors nécessité beaucoup de temps et

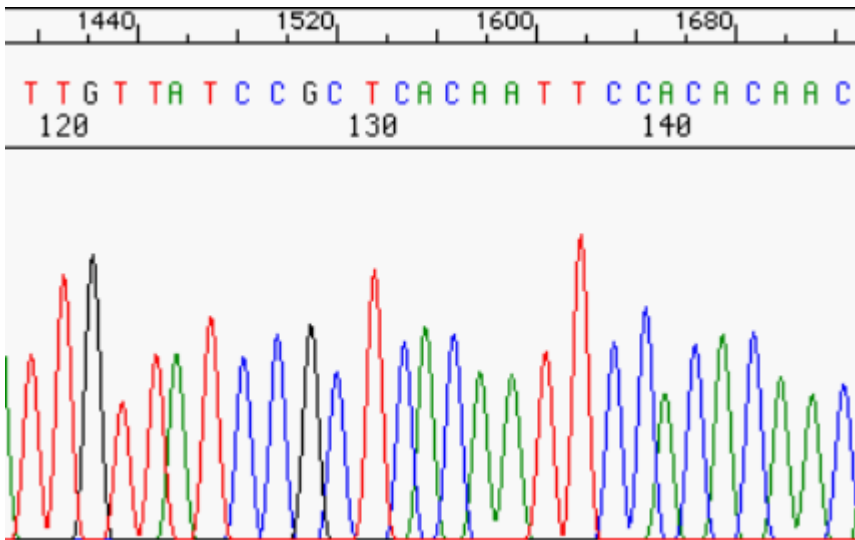
de machines. Il a aussi fallu automatiser au maximum toutes les étapes de ce processus.



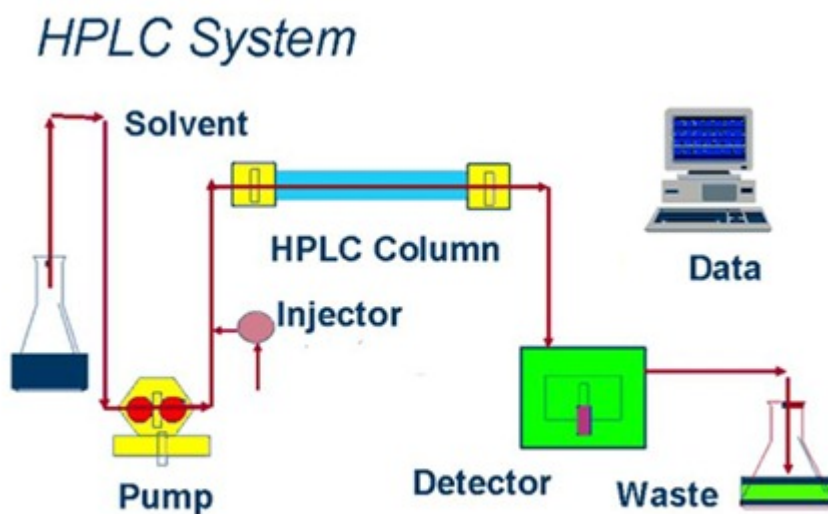
Dans les années 1980/90, on utilisait quatre préparations, une avec chaque type de ddNTP car si on savait les marquer par radioactivité ou par fluorescence, on ne savait pas les distinguer. Le tri des brins d'ADN synthétisés se faisait donc en quatre parties, une par base. Depuis, on utilise des marqueurs fluorescents distincts pour chacun des quatre ddNTP et ils sont activés par laser. Ce qui permet de ne faire la polymérisation que dans un seul « bain » avec l'ensemble des ddNTP mélangés. La lecture sur gel se fait sur une seule colonne « en couleur » et en général avec une caméra motorisée qui se déplace le long du gel.



Dans la pratique, on obtient en fait une courbe qui associe le continuum de luminosité sur les quatre couleurs analysées dans la fluorescence. On en déduit la succession de bases. Mais le processus est entaché d'erreurs et on les élimine en faisant plusieurs mesures redondantes et avec une analyse statistique (la base la plus fréquente gagne en cas de doute).



Dans les séquenceurs automatiques de seconde génération, on a remplacé l'électrophorèse sur gel par une électrophorèse capillaire dans un tube extrêmement fin. Le résultat est bien plus rapide et permet de lire jusqu'à un millier de séquences de bases. Ce procédé appelé HPLC (high-performance liquid chromatography, ou high-pressure lc) utilise en lieu et place du filtre plat de gel d'agarose un tube très fin de 4 mm de diamètre qui met sous pression le liquide à analyser.



Source: College of Arts and Sciences, New Mexico State University

Le liquide sort à l'autre extrémité du tube et est analysé par un chromatographe utilisant des LED. Le procédé est amélioré avec la UHPLC (Ultra-High Performance Liquid Chromatography). Voici quelques exemples de systèmes de HPLC et UHPLC, sachant que leur usage dépasse de loin le séquençage d'ADN. La spectrographie en phase liquide a plein d'applications dans la santé et la chimie. On peut l'utiliser pour identifier les composants chimiques d'un aliment industriel, mesurer divers niveaux de toxicité, etc.



Voilà pour la méthode de Sanger et ses évolutions.

Passons maintenant aux autres méthodes et technologies de séquençage apparues ces dix dernières années, dans un rythme d'innovation qui s'est sans cesse accéléré au point d'aller plus vite que la fameuse loi de Moore en termes de capacité de traitement.

Où sont les variantes ? De plusieurs ordres :

- La **synthèse analysée en temps réel** : au lieu de bloquer la synthèse aléatoirement à la hauteur de n'importe quelle base comme dans la méthode de Sanger, un grand nombre de nouvelles méthodes vont gérer la synthèse base par base de manière contrôlée et analyser le résultat au fur et à mesure, d'où l'appellation "temps réel". On retrouve notamment ce principe dans le pyroséquençage et le séquençage par terminateur réversible. On identifie les bases intégrées par fluorescence ou des résidus de cette intégration, les pyrophosphates (avec la méthode 454, par fluorescence) ou les ions hydrogène (avec la méthode de Ion Torrent).
- La **synthèse réalisée en temps réel** : variante des méthodes précédentes, il s'agit de synthèse réalisée à la même vitesse que dans le vivant et observée au fur et à mesure par différents moyens (fluorescence ou captation électrique). On trouve des méthodes de ce genre chez Pacific Biosciences et BioNano Genomics.
- Le **besoin d'amplification** ou pas : certaines méthodes impliquent l'amplification des brins d'ADN avant le séquençage, d'autres non, comme le SMRT de Pacific Bioscience. D'où un gain de temps.
- Le **séquençage parallèle** : au lieu de séquencer un brin unique d'ADN, on va en créer un grand nombre sous la forme d'une "bibliothèque" (DNA library) généré par "shotgun" (fragmentation aléatoire d'ADN en petits brins de quelques centaines de bases) et les analyser en parallèle. Le tout s'appuie sur des matrices avec des nano-cuves ou des plaques de verre dans ou sur lesquelles on isole chacune des réactions de transcription contrôlée.
- Les **domaines d'applications** : tous les systèmes de séquençage ne sont pas dédiés au séquençage complet du génome humain. Ce domaine d'application est d'ailleurs relativement marginal. Certaines applications visent à détecter le polymorphisme dans l'ADN, à savoir les variations par rapport à une cartographie de référence. On cherche à mesurer les variations dans les gènes liées aux différents allèles (zones codantes des gènes qui varient d'un individu à l'autre) tout comme le processus d'expression des gènes qui contrôle la manière dont les gènes génèrent les protéines via tout le processus de transformation qui comprend notamment l'ARN messager, l'ARN de transfert ainsi que les ribosomes.

Côté **performances**, les données clés sont le nombre de bases séquencées par heure, la taille moyenne des brins

d'ADN séquencés d'une seule traite et le pourcentage d'erreurs de séquençage. Mais on peut y intégrer également le nombre d'agents et de solvants nécessaires dans le processus. Et in-fine, le cout d'ensemble des opérations, de la machine aux consommables. Sachant que tout cela progresse plus vite que la loi de Moore, il est assez difficile de comparer clairement les technologies en présence.

Séquençage par terminateur réversible (Illumina, Helicos)

Le séquençage par terminateur réversible (ou CRT pour Cyclic Reversible Termination) est commercialisé par **Illumina** sous l'appellation TruSeq dans ses machines HiSeq et aussi dans une autre variante chez Helicos.

Dans cette **méthode**, on séquence l'ADN par polymérisation avec des dNTPs marqués par fluorescence en mode quatre couleurs (une par type de base) et un terminateur qui arrête la polymérisation. Après chaque ajout d'une base, on détecte laquelle a été ajoutée par un capteur d'image couleur de plusieurs millions de pixels (Illumina, où on ajoute toutes les bases d'un coup) ou noir et blanc (**Helicos** BioSciences, où l'on ajoute les bases une par une).

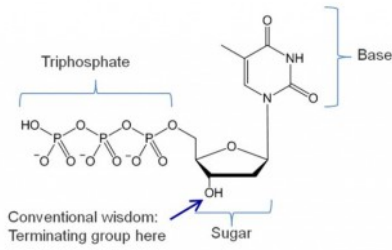
On enlève ensuite aux ultra-violets le terminateur de la dernière base ajoutée, du nitrophényl, qui bloque la liaison OH qui permet l'enchaînement des bases (cf le schéma plus haut dans cet article qui explique comment fonctionne cette liaison OH). On peut alors procéder à l'ajout de la base suivante. C'est donc un séquençage dit temps-réel comme ceux que nous verrons plus loin. Le temps de séquençage pour une base est environ de 2 minutes.

Ces étapes sont réalisées sur une plaque de verre et de manière parallèle avec des millions de brins d'ADN étalés dessus (pour faire simple) dont la préparation à partir d'un échantillon d'ADN dure environ trois heures. La technologie permet de séquencer 10 millions de brins d'ADN par centimètre carré. La machine la plus récente, la HiSeq 2500 (*ci-dessous*) peut analyser 120 gigabases en 27 heures, auxquelles il faut ajouter 20h de traitement informatique. Le taux d'erreurs de lecture semble relativement élevé, de plusieurs %, mais il est compensé par la redondance de celles-ci. Ainsi, un run de 120 Gbases correspond-il à 40 fois le nombre de base d'un ADN humain.



On trouve aussi cette technologie de terminateur réversible (CRT : Cyclic Reversible Terminator) chez **LaserGen** (Houston, \$5m levés). La startup planche avec National Instruments sur un système qui exploite un autre type de terminateur réversible. Ils visent le séquençage d'une gigabase par jour ce qui n'a rien d'extraordinaire sauf si le processus est lui-même hautement parallélisé.

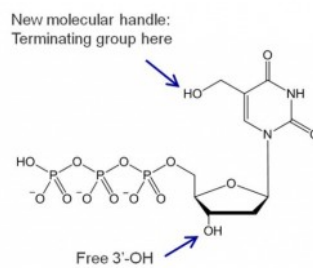
Contemporary Reversible Terminators:



Problems:

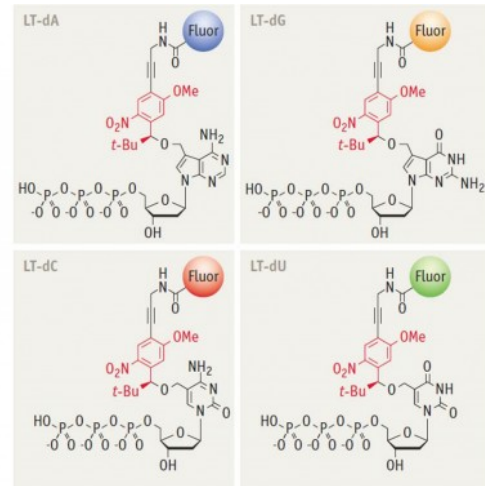
- Requires mutant enzymes
- Less accurate
- Shorter read-lengths

LaserGen® Reversible Terminators:



Solutions:

- Uses natural enzymes
- More accurate
- Potentially longer read-lengths

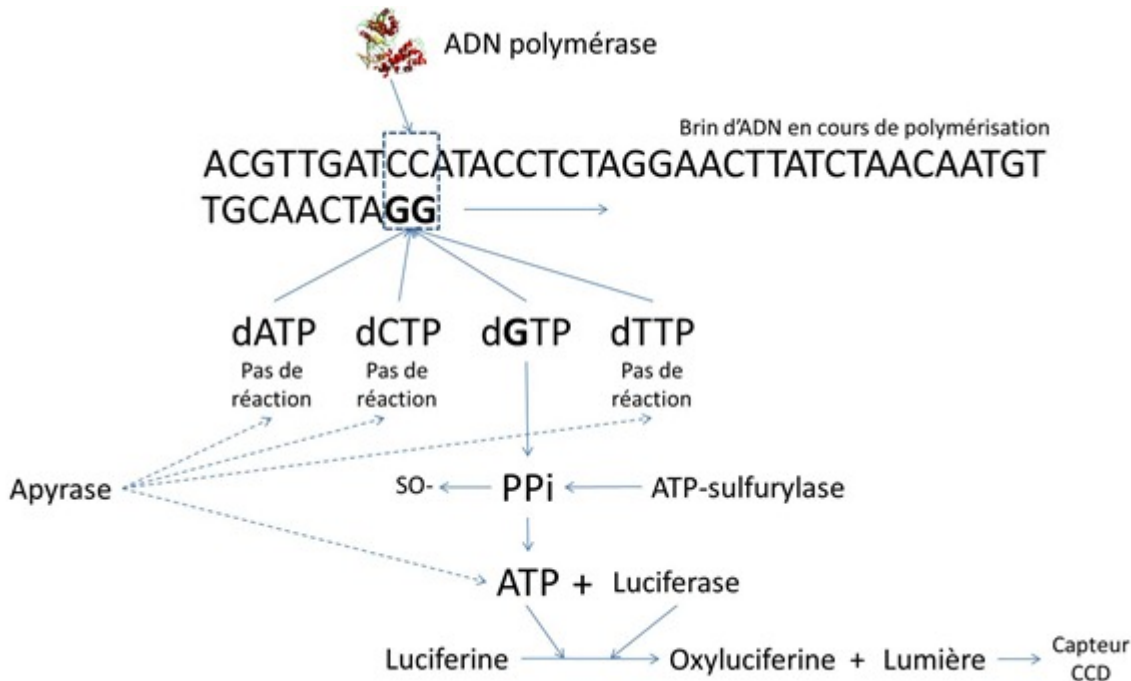
**Pyroséquençage (Roche/454, Qiagen, ...)**

Le pyroséquençage utilise une méthode de séquençage créée en 1988 par Hyman et perfectionnée ensuite entre 1996 et 1998 par l'intégration de la PCR dans le processus. Elle est différente du séquençage sauce Sanger. Nous l'examinons ici même si son domaine d'application concerne plutôt les courtes séquences d'ADN ou d'ARN et n'est pas spécialement adapté au séquençage complet du génome humain. Mais certaines techniques de séquençage massivement parallèles s'appuient sur le principe du pyroséquençage.

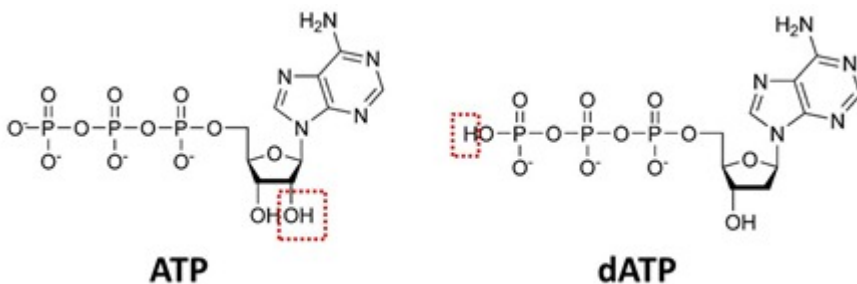
Comme dans la méthode de Sanger ou la PCR, on part aussi d'un brin d'ADN préalablement sélectionné et complété par une amorce et on lance une polymérisation avec de l'ADN polymérase et des nucléotides dNTP. Par contre, on ne va pas arrêter cette polymérisation avec l'emploi de ddNTP marqués par fluorescence qui la stoppent de manière aléatoire dans la méthode de Sanger.

On va au contraire en provoquer une seule en alimentant successivement le "bain" avec les quatre nucléotides et on va détecter celle qui va s'accrocher au brin d'ADN en cours de polymérisation. Pour cela, on va utiliser un résidu de cette polymérisation : le pyrophosphate (PPi, molécule à deux phosphates) qui se détache des dNTP lorsque le reste (1 base, 1 ribose et 1 phosphate) s'intègre au brin d'ADN en cours de construction. Le pyrophosphate a donné son nom à la technique... du pyroséquençage. Pourquoi pyro ? Parce que les pyrophosphates sont préparés industriellement en chauffant des phosphates.

Pour chaque base synthétisée, il y aura donc un dégagement de PPi. On va alors le convertir en dégagement de lumière après une triple réaction chimique contrôlée qui passe par les étapes suivantes :



- Un ATP sulfurylase transforme la molécule de PPi en ATP en générant un résidu de SO₂ (sulfure). L'ATP ou adénosine triphosphate sert sinon au transport de l'énergie dans les cellules. C'est une molécule qui se différencie de la dATP (l'une des quatre dNTP qui sert à la synthèse de l'ADN) par un atome d'oxygène en moins à l'extrémité du triphosphate et par un groupement OH en plus sur le ribose.



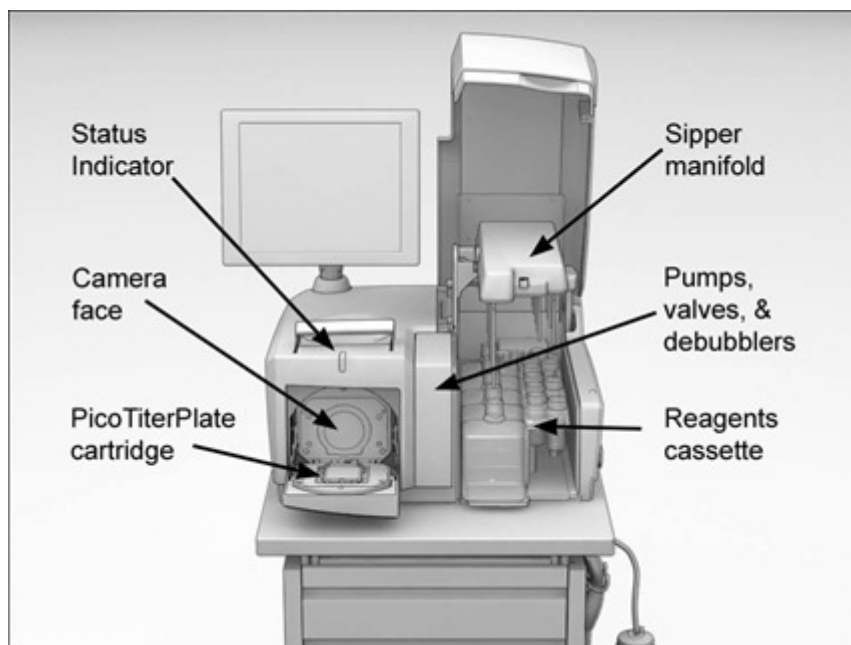
- La molécule d'ATP générée va elle-même convertir de la luciférine en oxyluciférine à l'aide d'une luciférase ce qui va générer de la lumière que l'on va détecter avec une caméra CCD.
- Une apyrase va dégrader les dNTP qui n'ont pas été utilisés tout comme les résidus d'ATP. Et on peut recommencer le cycle. Sachant que l'on pourrait en principe passer à la base suivante de la polymérisation sans passer par toutes les bases une fois qu'on a détecté la bonne et avoir en moyenne 2,5 dNTP à tester par base à séquencer. Mais il ne semble pas que ce soit le cas, histoire d'être plus rigoureux d'un point de vue statistique. Le pyroséquençage présente en l'effet de générer beaucoup moins d'erreurs pour le séquençage que la méthode de Sanger.

A la fin de ce processus, un autre dNTP est ajouté pour passer à la base suivante dans la synthèse. Sachant que la réaction n'a donc lieu statistiquement qu'une fois sur quatre, lorsque l'on a injecté la base qui correspond à celle qui manquait dans la chaîne d'ADN en cours de polymérisation (ou copie). Si il y a plusieurs bases du même type qui se suivent dans l'ADN à polymériser, plusieurs dNTP du même type seront intégrés par l'ADN à polymériser et l'émission de lumière sera multipliée d'autant.

Voici un exemple de pyroséquenceur : le PyroMark Q96 MD Automated de l'allemand **Quagen** (*ci-dessus*) qui automatise comme son nom l'indique toutes les étapes de ce processus répétitif. Il peut traiter en parallèle 10 échantillons identifiés par un code barre.



Mais c'est technologie "454", datant de 2005, qui est la plus avancée dans l'usage du pyroséquençage. Elle a été créée par la société américaine 454 Life Science qui a été acquise par le groupe suisse **Roche** en 2007. Elle permet de séquencer environ un demi-milliard de base en 23 heures grâce à du pyroséquençage parallèle. Le séquenceur associé le plus élaboré est le GS FLX Titanium XL+ (*ci-dessous*).

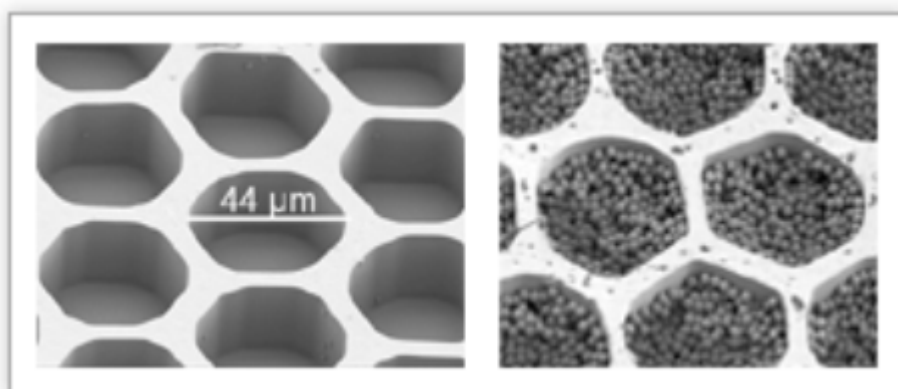


Le procédé 454 fonctionne en trois étapes clés :

- L'ADN à séquencer est d'abord dénaturé et puis fragmenté aléatoirement en petits brins de quelques centaines de bases auxquels on adjoint des ligateurs aux deux extrémités sur lesquels vont s'accrocher les amorces du séquençage. Cette opération est réalisée par "nébulisation", un procédé consistant à souffler un courant d'air sur un petit flux de liquide un peu comme on vaporise l'essence dans l'injecteur du piston d'un moteur à explosion. On obtient des brins d'ADN dont la taille moyenne est de 500 à 700 bases. Mais

avec une répartition qui va de 100 à 1000. Comme d'habitude, ceci ne s'effectue pas sans préparation. L'échantillon d'ADN à nébuliser doit être préparé : avec un mélange d'eau, de glycérol, de gel d'agarose (encore lui), du glycogène, de l'éthanol et d'isopropanol. Le résultat : des centaines de milliers de brins d'ADN simples qui vont maintenant jusqu'à 1000 bases, le tout à basse température. Cette partie du processus dure plus de 4 heures.

- Les fragments sont attachés un par un à une microbille de polystyrène entourée de streptavidin, une protéine complexe qui résiste bien aux divers solvants utilisés dans la suite des opérations. L'association est réalisée dans une émulsion eau-lipide qui est ensuite fixée dans une "micro-cuve" de 44 microns de large sur un support appelé PicoTiterPlate. Ce support contient 1000000 micro-cuves. On en déduit que c'est un carré de 4,4 cm de côté, ce qui est compatible avec l'illustration ci-dessus.
- Le fragment d'ADN est alors amplifié par emPCR (vue dans l'épisode précédent) autour de la bille. Ce processus dure 8 heures à 23 heures selon la machine utilisée et la taille moyenne des brins d'ADN séquencés.



- Enfin, le pyroséquençage base par base est lancé en parallèle dans chacune des microcuves du système à partir d'un système de microfluidique qui alimente le MicroTiterPlate avec les quatre nucléotides (séparément) et les différents agents de l'opération. La micro-cuve est couplée à un grand capteur CCD qui permet la lecture des résultats. C'est la taille du capteur qui semble limiter le nombre de bases qui peuvent être lues en parallèle dans cette machine. En consultant le site de **Dalsa**, l'un des fabricants des plus grands capteurs CCD et CMOS, on constate que **leur plus grand capteur** fait 50mmx50mm, donc peut couvrir la surface en apparence rectangulaire du MicroTiterPlate. Les pixels de ce CCD font 24 microns de côté, ce qui correspond à peu près à la moitié de l'espacement entre les microcuves. Ces capteurs sont monochromes car on n'a besoin de détecter qu'une émission de lumière et pas sa couleur. Ces capteurs sont d'ailleurs aussi utilisés en radiographie, elle aussi noir et blanc. Cela veut donc dire que si c'est le capteur ou le type de capteur que la machine utilise, quatre pixels du CCD suffisent à détecter la réaction lumineuse du pyroséquençage. A moins que le capteur ne soit plus petit et qu'une simple optique ne réduise l'image des deux plaques du pyroséquençage. Ceci étant, on trouve de plus grands capteurs, comme chez Agilent dont le **Titan** fait 165 mm de côté, mais 2Kx2K pixels !

Le processus de pyroséquençage dure 23 heures ce qui veut dire que pour une moyenne affichée de 700 bases par séquençage, la machine est capable de générer un cycle complet de détection d'une base en environ 2 minutes, ce qui correspond aux 2 minutes habituelles du pyroséquençage classique. Une bonne règle de trois permet toujours de faire quelques vérifications !

Alors, on lit 700 millions de bases en une passe, donc on peut séquencer le génome humain en 5 passes sachant que celui-ci comprend 3 milliards de bases dans sa forme haploïde, c'est-à-dire non dupliquée ? Non, pas si simple ! Un bon séquençage de génome nécessite d'avoir beaucoup de redondance. On génère habituellement 20 à 30 fois plus de séquences que n'en contient un génome. Après, un logiciel reconstruit le génome par recouplement entre les séquences obtenues. La redondance permet à la fois le recouplement et de détecter et supprimer les erreurs de séquençage qui subsistent dans tous les procédés. Dans le procédé 545, il y a 1% d'erreurs dans le séquençage !

La suite...

Ce petit tour est loin d'être terminé. Dans l'**article suivant** de cette série, nous aborderons d'autres technologies de séquençage dites de seconde et troisième génération, et dont certaines sont vraiment fascinantes.

Cet article a été publié le 10 août 2012 et édité en PDF le 24 décembre 2021.
(cc) Olivier Ezratty – “Opinions Libres” – <https://www.oezratty.net>