



Les technologies de séquençage du génome humain – 2

Dans l'épisode précédent de cette série estivale sur le séquençage du génome humain, nous avons fait le point sur la manière dont l'ADN était construite en la décortiquant comme une poupée russe. Nous avons notamment décrit ses "bases", les primitives de la molécule, puis les séquences codantes et non codantes qui conditionnent les gènes et leur expression dans les cellules.

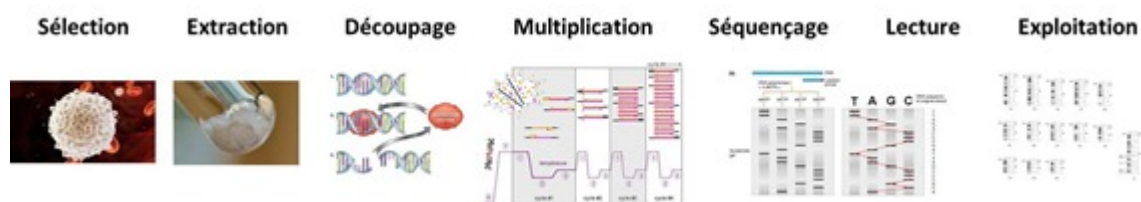
Dans cette seconde partie, nous allons étudier les techniques et appareillages mis en œuvre pour séquencer l'ADN en nous focalisant sur le cas de l'ADN du génome humain.

Il existe plusieurs méthodes de séquençage de l'ADN, la plus courante au départ étant celle de Sanger. Le séquençage de l'ADN est un processus chimique complexe qui comprend de nombreuses étapes. Cela explique pourquoi Sanger et Gilbert qui est avec Maxam créateur d'une autre méthode de séquençage maintenant abandonnée ont tous les deux obtenu le prix Nobel de chimie et pas de médecine !

Le séquençage de l'ADN s'appuie sur des étapes communes aux différentes méthodes employées que je résume et vulgarise comme suit sachant que le séquençage proprement dit commence à la cinquième étape de ce processus :

- 1) La **sélection** des cellules destinées au séquençage, en général, les lymphocytes du sang pour ce qui est des recherches cliniques et biologiques classiques.
- 2) L'**extraction** de l'ADN des cellules et leur purification, qui passe par le broyage des cellules et l'élimination de ce qui ne nous intéresse pas, à savoir principalement les lipides et protéines.
- 3) Le **découpage** des brins d'ADN à différents endroits connus à l'avance qui permettent de sélectionner les portions à étudier et qui s'appuie sur des enzymes de restriction.
- 4) La **multiplication** de ces brins pour augmenter leur nombre par millions et faciliter la lecture des résultats dans les phases suivantes. Celle-ci s'appuie le plus souvent sur de la PCR.
- 5) Le **séquençage** proprement dit s'appuie sur une réplication des brins d'ADN s'arrêtant de manière aléatoire sur chacune des bases de l'ADN. Elle génère une soupe avec des brins de différente taille dont les extrémités sont marquées, souvent par fluorescence, afin que l'on puisse identifier les quatre bases de l'ADN qui sont à leur extrémité. On trie alors par taille ou longueur les brins rognés et marqués, ce se faisait au départ sur des gels mais exploite maintenant différentes techniques à base de nanotechnologies dans les séquenceurs de dernière génération.
- 6) La **lecture** des résultats, historiquement faite sur un gel sur quatre colonnes avec des traits indiquant la présence des quatre bases.
- 7) Leur **exploitation** pour recoller les morceaux et reconstituer le fragment d'ADN ou l'ADN complet que l'on

souhaite séquencer, sachant que le processus est généralement redondant pour détecter et supprimer les erreurs de séquençage.



La variantes de processus de séquençage concernent en particulier le moment où l'on fait une PCR pour démultiplier les brins et sur la parallélisation du traitement de chaque brin issu du premier découpage de l'ADN.

Nous allons examiner ce processus étape par étape sachant que c'est dans la cinquième partie que nous départagerons les grandes méthodes de séquençage et notamment celles qui font appel à du séquençage massivement parallèle. Le tout dans le troisième article de cette série d'été.

Etape 1 : sélection des cellules

Le poids de l'ADN représente environ 1% de celui des cellules. Un corps humain adulte comprend donc environ 600g d'ADN... ce qui n'est pas rien ! En principe, on peut séquencer l'ADN à partir de n'importe quelle cellule, y compris une cellule d'un organisme cliniquement mort, voire même fossilisé. L'ADN le plus ancien qui ait été récupéré était celui d'un coléoptère du crétacé vieux de plus de 120 millions d'années ([source](#)) ! Tout dépend du matériau vivant dont on dispose et aussi de l'usage qui est fait du séquençage : identification d'un gène, séquençage complet de l'ADN, test de paternité, reconnaissance d'individus dans le cadre de la police scientifique, etc. Dans ce dernier cas, tout est bon à prendre : cheveux, salive ou sperme. Mais le processus d'extraction de l'ADN est facilité par l'usage de cellules issues de liquides biologiques, et en particulier du sang.

Pour le séquençage d'ADN complet, on passe généralement par le sang qui est prélevé de manière classique dans le pli du coude comme dans toute prise de sang. Les cellules sanguines présentent la particularité de ne pas se diviser – elles sont générées dans les os – et donc d'avoir un ADN qui ne se présente pas sous forme de chromosomes diploïdes (avec l'ADN répliqué autour du centromère) mais de simple brin autour d'un centromère. Le sang prélevé subit un traitement chimique et de centrifugation pour en extraire les lymphocytes, un des types de globules blancs ([voir ici](#)). Ceux-ci ne représentent que 1% de la composition du sang. On a donc les cellules. Etape suivante : récupérer l'ADN qu'elles contiennent.

Etape 2 : extraction de l'ADN des cellules

L'ADN est confiné dans le nucléole au sein du noyau des cellules. Comment fait-on pour se débarrasser non seulement de tout ce qu'il y a autour du nucléole dans les cellules mais de tous les autres composants, notamment protéiniques, des cellules ?

En pratique, chacun peut tenter l'expérience avec les moyens du bord et des oignons ou des bananes ([cf ici](#)). On détruit la paroi cellulaire (lyse) par traitement mécanique (écrasement dans un mortier) en présence de sel. Un solvant organique comme du liquide vaisselle est utilisé pour éliminer les lipides, protéines et autres acides des cellules. On filtre alors le résultat et on récupère l'ADN par précipitation dans l'alcool. L'ADN peut être coloré pour être rendu visible à l'œil nu sous forme de pelotes de filaments.

En laboratoire, c'est un peu plus compliqué car l'objectif est d'obtenir un ADN purifié vraiment débarrassé de

tout le reste des cellules. Après broyage des cellules, on dissout le résultat avec une solution de lyse comme le **Buffer P2** de Qiagen (37€ le demi-litre). Certaines solutions de lyse vont jusqu'à 2500€ le litre (le **Trizol LS Reagent**). Cette solution va dissoudre les lipides et dénaturer les protéines (déplier leur structure moléculaire tridimensionnelle). Cela va notamment dissoudre les protéines de la chromatine comme les histones autour desquelles l'ADN est enroulé.

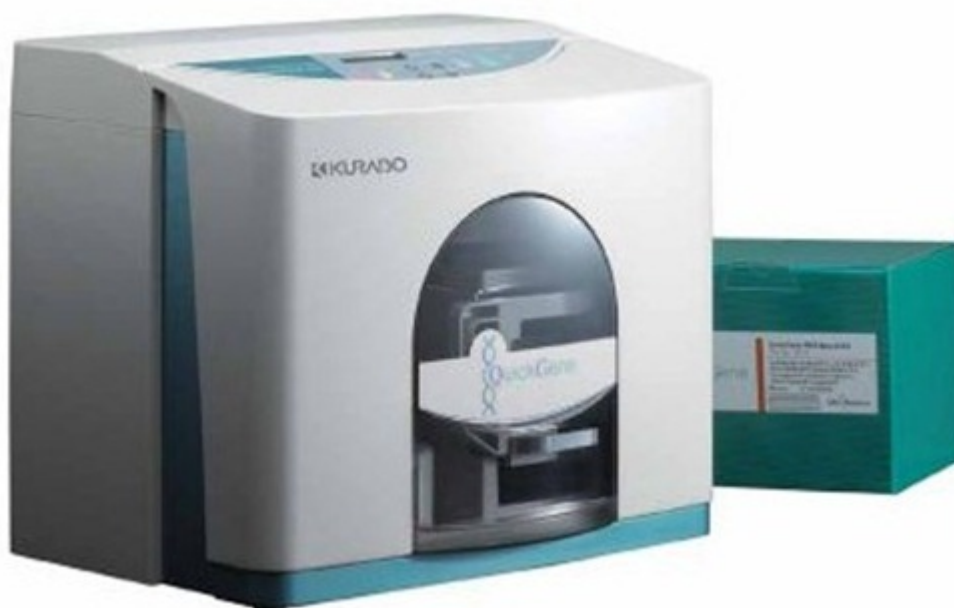
Il existe des machines spécialisées dans la préparation des cellules comme la **Qiagen TissueLyser II** (*ci-dessous*) qui optimise en broyage des cellules en associant la centrifugation et l'usage de microbilles de 0,1 microns. Au passage, cela permet de vendre du consommable, le business model de base de tous les fabricants de ces instruments de laboratoire, même si ces derniers restent assez chers (5700€ pour le Qiagen).



On élimine ensuite les protéines avec différents solvants comme le dodécylsulfate de sodium (**SDS**, utilisé dans les détergents domestiques ou le dentifrice) et avec de la protéinase K, une enzyme qui les digère. Dans l'éprouvette, l'ADN est récupéré par décantation ou centrifugation (*cf centrifugeuse eppendorf ci-dessous*). Enfin, on la précipite avec de l'éthanol et on la nettoie ensuite avec du phénol et du chloroforme pour la débarrasser de tous les solvants utilisés.



Tout ce processus peut durer plusieurs heures lorsqu'il est réalisé "à la mano". Et on n'a pas encore démarré le séquençage proprement dit ! En fait, le processus d'extraction d'ADN a lui aussi bénéficié de progrès technologiques ces dernières années. Il en va ainsi du système **QuickGene 810** (*ci-dessous*) créé par Fujifilm Life Science et distribué par Kurabo qui permet de gérer tout ce processus en 6 minutes pour extraire l'ADN du sang. Au lieu de la centrifugation, il utilise un système de filtration sous pression à base de nano-membranes.



Il existe même des solutions encore plus simples à base de seringues d'extraction d'ADN comme le **Pinpoint Slide DNA Isolation System** de Zymo Research. Mais la digestion des protéines à base de Protéinase K dure tout de même 4 heures. Dans la même veine, il y a aussi le **LeukoLOCK** de la société Ambion qui est spécialisé sur l'extraction d'ADN de cellules sanguines. Ainsi que le **MELT** également de Ambion, utilisable pour les autres types de cellules et qui utilise une combinaison propriétaire d'enzymes. Enfin, citons le kit **Agencourt Genfind v2** de Beckman Coulter qui fonctionne avec du sang et du sérum et utilise une technologie à base de magnétisme, la technique de "Solid Phase Reversible Immobilization", qui permet une extraction en moins de 3 heures avec un haut rendement. Cette forme d'extraction est utilisée à la fois au début du processus, au niveau de l'extraction, et ensuite, après PCR.

On n'en a pas encore terminé car il faut maintenant purifier l'ADN ainsi obtenu. On peut aussi le faire par précipitations successives dans l'éthanol ou par chromatographie. En ajoutant du sel (NaCl), on fait apparaître l'ADN sous la forme d'un "nuage de méduse" qui est une pelote de longs filaments d'ADN dépliés. Et cela ressemble à ça quand on en a une grande quantité (photo prise au Généthon à Evry) :

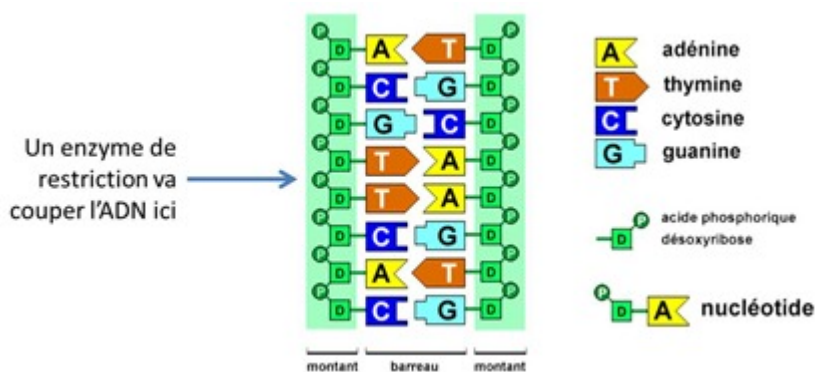


D'un point de vue pratique, l'ADN extrait peut-être conservé à sec après séchage sous vide ou être dissous dans de l'eau pure ou dans un mélange trishydroxyméthylaminométhane et acide éthylène diamine tétra-acétique, communément appelé Tris-EDTA.

Etape 3 : découpage

Après l'étape 2, nous avons obtenu des microgrammes d'ADN purifié. Alors, on le séquence directement ? Bien non. On va en général d'abord le découper en morceaux, puis "amplifier" ces morceaux en les répliquant. Tout du moins pour la plupart des méthodes de séquençage de première et de seconde génération.

Le découpage s'effectue à l'aide de protéines spécialisées, les enzymes de restriction, qui vont littéralement couper l'ADN en deux. Elles opèrent au niveau des liaisons phosphodiester qui relient les nucléotides entre elles (entre phosphate et sucre). Ces ciseaux chimiques coupent l'ADN à des endroits bien précis appelés "sites de restriction" identifiés par de courtes séries de 4 à 10 paires de bases. Dans le milieu naturel, ces enzymes de restriction sont utilisées par les bactéries pour se protéger des virus bactériophages qui les attaquent. En gros, elles taillent en pièces l'ADN de ces virus.



La coupure opérée par les enzymes de restriction touche les deux brins de l'ADN. Elle a lieu entre liaisons du même type (A-G et G-A par exemple) au milieu ou pas de la séquence de paires de bases et sur des séquences palindromiques, à savoir des séquences où les deux brins sont symétriques l'un de l'autre comme illustré ci-

dessous. L'exemple ci-dessous à gauche présente le ciseau de l'enzyme *EcoRI* qui provient de la bactérie *Escherichia Coli*. Celui de droite correspond à l'enzyme *SmaI* issue de la bactérie pathogène *Serratia marcescens*, une habituée des infections nosocomiales.



La contraposée des enzymes de restrictions (endonucléases) sont les ADN ligases qui peuvent raccomoder des brins d'ADN cassés. Elles opèrent lors du processus de réplication naturel de l'ADN ou bien dans les techniques de génie génétique.

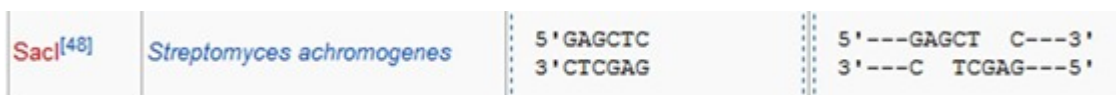
Les enzymes de restriction ont été découvertes à la fin des années 1970 par HindII et leur inventaire réalisé dans les années 1970 par Aber, Nathans et Smith (tous prix Nobel). Parmi les applications de ces découvertes, on compte la fabrication d'insuline humaine synthétique à partir de bactéries *Escherichia Coli*, en lieu et place de l'insuline à base de pancréas de porc qui ne pouvait pas être produite en volume et qui n'était pas parfaitement biocompatible avec le corps humain.

A ce jour, on a identifié environ 3000 enzymes de restriction s'appliquant à l'homme et 600 d'entre elles sont couramment exploitées dans la préparation de l'ADN pour son séquençage. Mais elles ont plein d'autres utilités dans l'industrie, la chimie – comme avec les enzymes glutons des lessives –, l'agriculture et la santé.

On distingue trois types d'enzymes de restriction. Les enzymes de type I et de type III reconnaissent des sites de restriction mais coupent l'ADN en un endroit aléatoire pouvant être à 25 jusqu'à 1000 paires de base du repère (type I et III). les enzymes type II coupent l'ADN à l'endroit de la séquence reconnue.

Pour le séquençage de l'ADN, on essaye de couper celle-ci pour créer des morceaux faisant moins que 1 mbases (1 million de paires de bases). Comme il y a environ 3 millions de bases dans l'ADN haploïde (hors phase de duplication) et plus de 23000 gènes identifiés, cela veut dire que les enzymes de restriction couvrent une partie seulement du génome et qu'elles permettent de séquencer des morceaux d'ADN qui comprennent parfois plusieurs gènes.

D'un point de vue pratique, le découpage de l'ADN pour son séquençage va s'appuyer sur l'emploi de cartes de restrictions qui font l'inventaire des sites de restriction connus dans l'ADN humain. On trouve d'une part des bases de données de "ciseaux" qui permettent de découper un fragment d'ADN identifié à un endroit donné comme la **NEBCutter V2** ou le **WebCutter** qui partagent comme de nombreux sites web de génétique un look très années 1990. Une fois qu'on a choisi ses ciseaux, il faut trouver les enzymes associées. On peut aller taper dans une base d'enzymes de restriction comme **Rebase**. Et après ? On achète les enzymes chez le marchand d'enzymes ! Il y en a des dizaines dans le monde et ils sont notamment répertoriés [ici](#). Par exemple, chez New England Biolabs, on pourra trouver la *SacI* qui coupe la séquence symétrique GAGCTC comme indiqué ci-dessous. On trouve aussi des grossistes et revendeur comme le français **Ozyme**. Chaque pack d'enzyme coute environ 50€.



On fait agir les enzymes avec l'ADN dans de l'eau purifiée complétée d'une solution tampon (buffer) et dans

des petits tubes placés dans un bain marie ou un four à la température de 37°... celle du corps humain (*exemple ci-dessous à droite*) ! On a alors comme résultat “un” morceau d’ADN qui va passer à l’étape suivante de l’amplification, le plus souvent par PCR.

Dans certaines méthodes de séquençage massivement parallèle, on découpe aléatoirement l’ADN en morceaux de taille variable pour créer une “bibliothèque de brins d’ADN” (DNA template library) qui seront ensuite séquencés simultanément. Nous verrons cela plus loin dans notre investigation du sujet.



Les enzymes sont conservées réfrigérées à des températures comprises aux alentours de -20°C dans des frigos de labo qui présentent plusieurs particularités : ils sont régulés par microprocesseurs, dotés d’alarmes sonores lorsque la température n’est pas la bonne, leur compresseur est en haut du frigo pour faciliter la circulation de l’air et ils n’ont pas de système automatique de dégivrage (qui fait remonter la température à 0°C toutes les 24h dans les congélateurs domestiques). *Exemple à gauche ci-dessus*.

Etape 4 : amplification

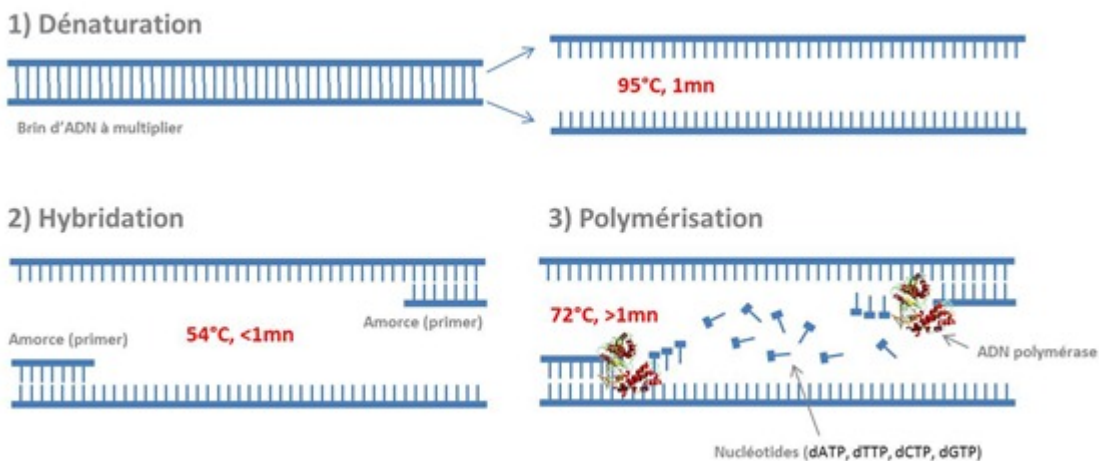
Dernière étape que nous allons aborder dans cet article, l’amplification. Elle consiste à prendre un brin d’ADN qui a été obtenu par découpe via enzyme de restriction et à le démultiplier jusqu’à des millions de fois pour en permettre le séquençage. Il existe plusieurs méthodes de multiplication mais je vais me focaliser ici sur la plus connue et la plus utilisée : la polymerase chain reaction (PCR) sachant qu’il en existe plein de variantes.

La PCR est un procédé inventé au milieu des années 1980 et qui a révolutionné le génie génétique. Avec la découverte des ADN polymérases qui interviennent dans la réplication de l’ADN et des enzymes de restriction, c’est l’un des maillons clés qui a permis le décodage du génome humain en 2000.

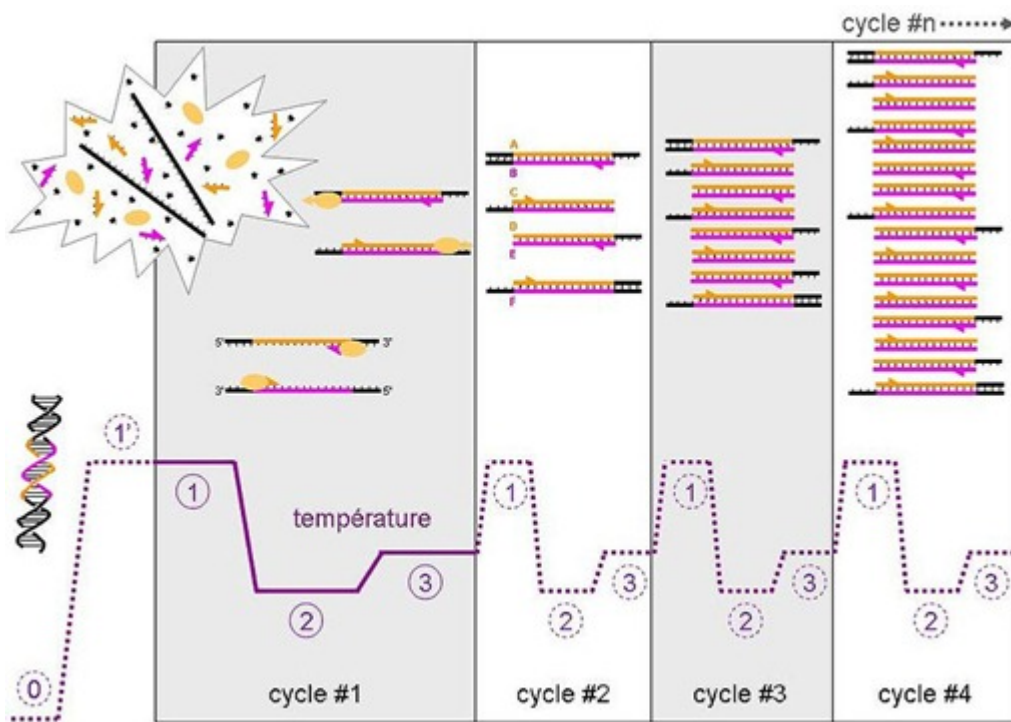
La PCR est une sorte de photocopieuse d’ADN qui permet de reproduire rapidement le processus de réplication semi-conservative de l’ADN qui intervient dans celui de la division cellulaire. Il permet de multiplier les brins d’ADN par 2, puis 4, puis 8, etc en peu de temps et de générer un échantillon de plusieurs dizaines de millions de brins d’ADN identiques que l’on pourra ensuite exploiter dans le séquençage. Cette amplification est rendue nécessaire du fait des techniques de lecture optique du résultat du séquençage que nous verrons plus loin. La PCR s’est aussi révélée très utile pour élucider des crimes lorsque l’on ne dispose que d’un tout petit nombre de cellules, comme un bout de cheveux.

La PCR fonctionne sur un cycle répété n fois qui comporte trois étapes :

- La **dénaturation**, qui sépare les deux brins de l'ADN à séquencer par une montée à une température de 95°C pendant une minute. On avait vu dans la première partie de cette série que ceci serait facilité par les liaisons hydrogènes qui relient les bases dans l'hélice d'ADN, ces liaisons étant à faible énergie.
- L'**hybridation**, où l'on ajoute de courtes amorces nucléotidiques de synthèse d'entre 18 et 24 bases qui sont complémentaires des deux extrémités du brin d'ADN à amplifier. L'opération a lieu à 50°C-60°C pendant moins d'une minute.
- La **polymérisation** où l'on fait agir des ADN polymérases (*exemple ci-dessous*) pour créer un brin complémentaire de chacun des deux brins séparés par la dénaturation. Ces ADN polymérase savent identifier les amorces et déclencher la "photocopie" de l'ADN par ajout de nucléotides complémentaires des paires de bases. On ajoute pour ce faire dans la solution les quatre nucléotides (dATP, dTTP, dCTP, dGTP) qui vont se lier à chacune des nucléotides complémentaires des deux brins. L'élongation commence à partir des amorces créées lors de l'hybridation et se poursuit jusqu'à la génération de deux copies à l'identique du brin d'ADN initial. On obtient alors quatre brins d'ADN. L'opération a lieu à 72°C pendant deux minutes.



On obtient ainsi un nombre de copies qui est une puissance de 2 du nombre de cycles qui durent chacun trois minutes. Il faut au moins une trentaine de cycles pour obtenir un échantillon convenable pour la suite des événements. Mais le nombre de cycles dépendra bien entendu de la taille de l'échantillon initial d'ADN. Une PCR durera entre 1h et 2h. Il existe des méthodes de PCR dites "**temps réel**" avec plein de variantes qui permettent de réduire le nombre de cycle grâce à une mesure temps réel du nombre de cycles avec un marqueur fluorescent. Cela permet d'arrêter la PCR lorsque l'on dispose d'un échantillon suffisant alors qu'en PCR classique, on n'obtient cette mesure qu'à la fin des cycles.



Quid des appareils de PCR et des consommables ? Les appareils sont une autre sorte de fours que ceux que nous avons vu pour le découpage de l'ADN à séquencer. Il s'agit de thermocycleurs. Dans les systèmes PCR temps réel devenus standards, ils exploitent un spectrofluorimètre qui mesure la fluorescence permettant de mesurer l'avancement de la PCR et de l'arrêter lorsque l'échantillon obtenu est suffisant. Ces appareils ne servent plus principalement au séquençage, mais à l'identification de gènes dans des échantillons d'ADN. Ils sont capables de mener des PCR en parallèle sur des dizaines d'échantillons d'ADN différents sélectionnés en fonction des gènes recherchés. La ABI Prism 7900HT ci-dessous est ainsi capable de mener 384 analyses en parallèle !



Il existe une variante de la PCR : l'emPCR ou "émulsion PCR" qui exploite une bibliothèque de fragments d'ADN générée par fragmentation aléatoire d'ADN génomique. Chaque fragment est attaché à une sorte de perle (bead) dont la surface est recouverte d'oligonucléotides. L'ensemble est intégré dans une émulsion eau-lipide. Chaque perle est enfermée dans sa bulle de lipide au sein de laquelle aura lieu une mini-PCR qui va répliquer le brin d'ADN qui y est rattaché, les répliquats étant attachés à la perle. On verra comment cette technique est utilisée dans certains séquenceurs massivement parallèles.

Côté consommables, les packs de nucléotides en solutions salines valent la peau des fesses : plus de 200€ par milli-litre mais cela s'utilise heureusement en toute petite quantité par PCR ! L'ADN polymérase est tout aussi chère tout comme les amorces. Nous sommes déjà passés par une bonne demi-douzaine d'appareils différents et nous n'avons toujours pas démarré le séquençage ! Ce qui me laisse songeur quand on parle de séquençage à \$100. Ce prix correspondra peut-être au consommable du séquençage à proprement parler. Mais il n'est pas évident qu'il intègre aussi toutes ces étapes que nous venons de traiter, sans compter l'amortissement de toutes ces machines et la main d'œuvre.

Je vais m'arrêter ici car le quota de signes et d'illustrations est dépassé. Prochain épisode : le **séquençage**, oui, lui, le vrai...

Cet article a été publié le 2 août 2012 et édité en PDF le 22 décembre 2021.
(cc) Olivier Ezratty – “Opinions Libres” – <https://www.oezratty.net>