

Séquençage

1- Définition

Depuis la fin des années 1970 et l'avènement des techniques de la biologie moléculaire, il est possible de séquencer un brin d'ADN, c'est-à-dire de lire l'enchaînement, ou séquence, des nucléotides constitutifs de cette molécule. Cela se ramène en fait à déterminer la succession des bases, la seule partie variable des nucléotides.

Cependant, les techniques actuelles ne permettent de lire, à chaque opération de séquençage, qu'un millier de bases au plus. Or la partie « séquençable » du génome humain comprend 2,9 milliards de paires de bases ! Il est donc impossible de lire l'ensemble d'un génome en une fois. Il est, de toute façon, impossible de manipuler des molécules d'ADN de plusieurs dizaines, voire centaines de millions de bases (l'ordre de grandeur de celles qui constituent les chromosomes humains).

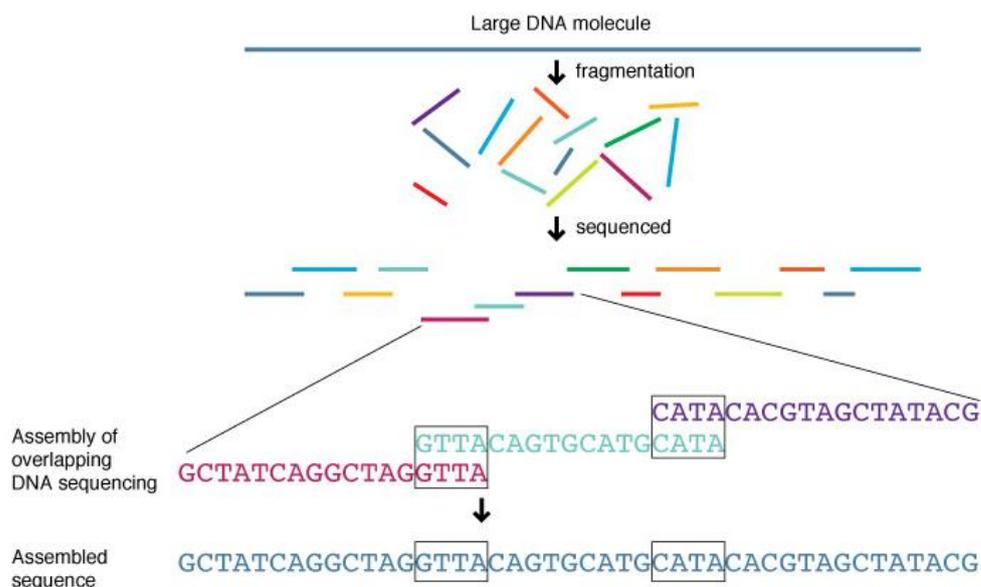
On parle de « partie séquençable » du génome humain (2,9 Gb, pour un total de 3,2 Gb). En effet, il n'est pas possible, techniquement, de déterminer la séquence de certaines régions presque exclusivement constituées de séquences répétées, telles que les centromères, les télomères ou les bras courts de certains chromosomes. Il y a deux raisons à cela : d'une part, il est difficile d'isoler des fragments d'ADN de taille convenable issus de ces régions ; d'autre part, il n'est pas possible de reconstituer la séquence complète à partir de morceaux de séquences pratiquement identiques. De ce fait, seule la séquence de la partie dite euchromatique du génome peut être effectivement déterminée.

2- Principe

Le principe de base, dans tout séquençage d'un génome, consiste à fragmenter de façon aléatoire ce génome – ou de grands morceaux d'ADN dérivés du génome – pour obtenir des morceaux d'ADN de quelques milliers de paires de bases, faciles à

manipuler. Les extrémités d'un grand nombre de ces petits fragments sont alors séquencées. La séquence complète du génome – ou du grand morceau de génome – est ensuite reconstruite à partir de ces séquences unitaires, ou lectures, sur la base des chevauchements entre les séquences (si les séquences sont chevauchantes, c'est que les fragments d'ADN dont elles dérivent ont une partie de leur longueur en commun ; la cassure étant aléatoire, les molécules d'ADN de l'échantillon ne sont pas toutes cassées aux mêmes endroits).

Rodger Staden, invente le premier « logiciel » de séquençage de l'ADN. En 1982, Sanger l'utilise pour assembler la totalité des 48 502 pb du génome du bactériophage Lambda.



3- Applications

La lecture des séquences d'ADN permet d'étudier l'information biologique contenue par celle-ci. Étant donné l'unicité et la spécificité de la structure de l'ADN chez chaque individu, la séquence de l'ADN permet de nombreuses applications notamment dans le domaine de la médecine, pour le diagnostic et le traitement de nombreuses maladies humaines (exemples : cancers, maladies infectieuses, maladies héréditaires...). Il fournit aussi des informations sur le génome (structure, fonction,

évolution) et permet l'étude des variations du génome (polymorphismes bialléliques, insertions, délétions, insertions/délétions (appelées aussi indéls), réarrangement de gènes, variation du nombre de copies de gènes, duplication (ou plus). Le séquençage est actuellement utilisé pour la recherche de variants génétiques associés à une pathologie (par exemple, le diabète), l'analyse de méthylation du génome (études épigénétiques et méthylome), l'analyse microbiologique (identification d'espèces, taxonomie, études épidémiologiques, génotypage à but pronostique et/ou thérapeutique)...en plus des applications en médecine légale et en anthropologie.

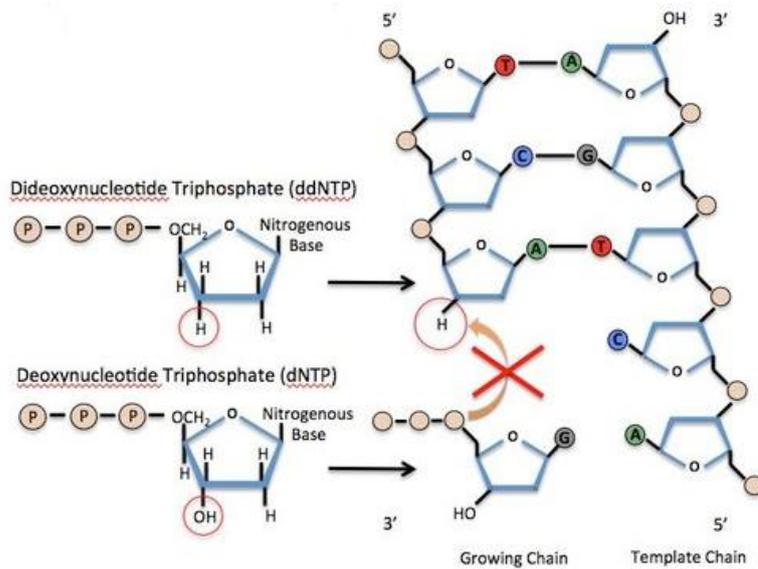
4- Technique

Les deux premières techniques de séquençage de l'ADN, celle de Maxam-Gilbert et celle de Sanger ont été décrites en 1977. La technique de Maxam-Gilbert est pratiquement abandonnée de nos jours, tandis que celle de Sanger reste à ce jour la principale méthode de séquençage utilisée dans les laboratoires.

- **Principe:** Dans un premier temps, il est nécessaire d'amplifier l'ADN cible par PCR, puis de le dénaturer afin d'obtenir un ADN simple brin. À l'aide d'une amorce spécifique et complémentaire du brin étudié (sens ou antisens), une ADN polymérase effectue alors la synthèse de l'ADN complémentaire à partir de cette amorce. L'amorce dite universelle, est complémentaire à l'extrémité 3' connue en amont du fragment à séquencer (sur le vecteur).

De l'extrémité 5' vers l'extrémité 3', l'ADN polymérase ajoute les désoxyribonucléotides-triphosphates (dNTP) complémentaires et de manière aléatoire et inconstante des didéoxyribonucléotides triphosphates (ddNTP), par exemple un ddGTP sera parfois ajouté à la place d'un dGTP. La réaction se faisant dans un seul tube, les ddNTP (ddATP, ddGTP, ddCTP et ddTTP) sont marqués à l'aide de fluorophores différents pour chaque ddNTP (fluorophores « quatre couleurs »). Lorsqu'un ddNTP est incorporé à la place d'un dNTP, l'ADN polymérase ne peut plus continuer sa polymérisation. La réaction d'extension s'arrête (en effet, le

didéoxynucléotide ne possède pas de groupe 3'-hydroxyle indispensable à la réaction de polymérisation de l'enzyme).



Statistiquement, au cours de la réaction, pour chaque « base » de l'ADN cible, au moins une fois, un ddNTP complémentaire sera incorporé à la place d'un dNTP. Par conséquent, à la fin de la réaction, nous obtiendrons des fragments de tailles différentes. L'analyse de la réaction est ensuite effectuée par électrophorèse.

L'ADN est séparé en fonction de sa taille et chaque bande peut être détectée à l'aide d'un détecteur de fluorescence.

