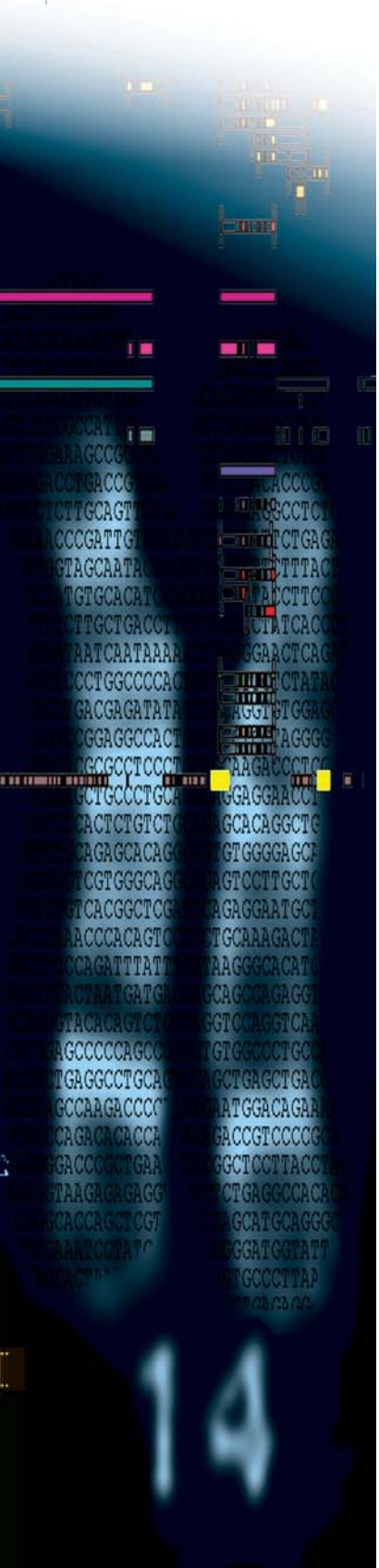
Le génome humain

7



Au début des années 1990, l'on avait déjà identifié plusieurs gènes humains impliqués dans des maladies génétiques, mais cela avait été au prix d'efforts individuels épars et désorganisés. Dans la communauté scientifique internationale, une idée avait fait son chemin : si les progrès techniques se poursuivaient en matière de séquençage, l'inventaire des gènes humains pourrait être obtenu à moindre coût par le séquençage de l'intégralité du génome humain. Le "Projet Génome Humain" était né.

Les promoteurs du projet s'engagèrent à séquencer la totalité du génome humain pour le début du troisième millénaire, aux environs de 2005. En raison de la taille de ce génome – 3,2 milliards de nucléotides, soit, en caractères, le contenu de 2 000 livres de 500 pages –, les grands centres de séquençage, réunis en un consortium international, ont convenu de diviser le travail. Chacune des 20 institutions du consortium s'est chargée de régions chromosomiques ou de chromosomes particuliers, parmi les 24 types de chromosomes humains. Chaque centre s'est également engagé à déposer les séquences, dès leur obtention, dans des bases de données publiques. Les efforts se sont répartis entre les six nations participantes de la façon suivante :

Etats-Unis	Royaume Uni	Japon	France	Allemagne	Chine
60,8 %	28,9 %	4,9 %	2,8 %	1,5 %	0,7 %

C'est ainsi que le Genoscope, seul contributeur français, a pris en charge la majeure partie du chromosome 14 (environ 3 % du génome humain). Comme indiqué dans la fiche *Cartographie*, le séquençage d'un chromosome, qui est une molécule d'ADN de très grande taille, commence par une étape cartographique : la molécule d'ADN chromosomique est coupée au hasard en grands fragments dont l'ordre est reconstitué. Ce sont ces grands fragments, et non le chromosome entier, que l'on casse alors individuellement en morceaux plus petits destinés à être séquencés *(voir la fiche* Stratégies de séquençage*)*. La nécessité d'un travail préalable de cartographie explique que le séquençage proprement dit n'ait commencé qu'en 1998.

Or, cette même année, une nouvelle société nommée Celera Genomics déclara qu'elle avait l'intention de séquencer le génome humain avant la fin de l'année 2001, sans établir au préalable de carte de grands fragments ordonnés *(stratégie du "séquençage aléatoire global", voir la fiche* Stratégies de séquençage*)*. En réponse à cette annonce, et afin de faire face aux menaces d'appropriation du génome humain, les organismes publics de financement anglo-saxons (NIH, DOE et Wellcome Trust) ont annoncé une augmentation importante des budgets et un nouvel objectif : l'assemblage, pour le printemps 2000, d'une "ébauche" préliminaire de la séquence du génome humain. Les autres pays (Allemagne, Chine, Japon et France) se sont joints à cette initiative.

Les délais ont été tenus, puisque le consortium international a annoncé le 26 juin 2000 l'achèvement de cette "version de travail" (*working draft*) couvrant près de 90 % du génome. En dépit de "trous" et d'incertitudes, l'ébauche, librement accessible, permettait d'ores et déjà d'identifier de nombreux gènes humains. Grâce à l'effort de cartographie et à la nature hiérarchique de l'ébauche, le travail résiduel pouvait être focalisé sur des régions bien localisées le long des chromosomes.

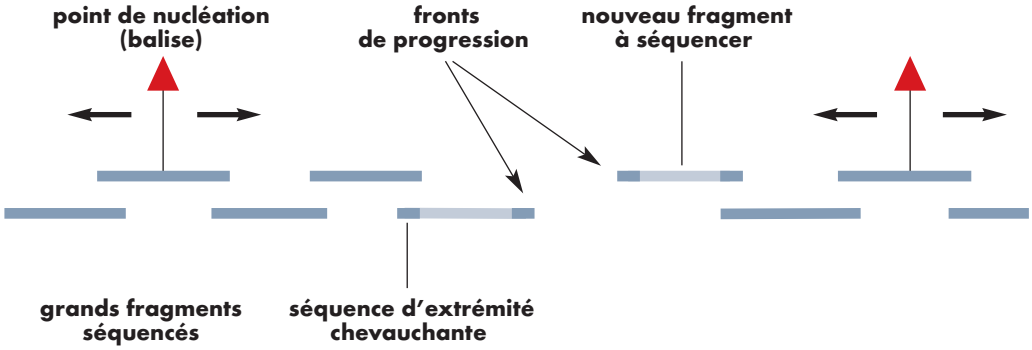
(suite au dos)

Le génome humain (suite)

7



Au Genoscope, les grands fragments du chromosome 14 ont été recherchés par criblage d'une collection, ou "banque", de près d'un demi-million de grands fragments (longs de 150 000 à 200 000 nucléotides) issus de la totalité du génome humain. On a utilisé pour ce criblage 1 500 balises préalablement cartographiées sur le chromosome 14. Le travail de séquençage a ensuite été entrepris à partir d'une centaine de points régulièrement espacés sur le chromosome (points de nucléation), en progressant dans les deux directions à partir de chaque point. Cette méthode présuppose le séquençage préalable des extrémités des grands fragments. Chaque fois qu'un grand fragment est à peu près complètement séquencé, on recherche d'autres fragments dont une des séquences d'extrémités est chevauchante, et l'on retient celui qui s'étend le plus loin. Une fois ce nouveau fragment séquencé, le processus recommence, jusqu'à la jonction de deux fronts de progression en directions opposées. Dans cette méthode, le séquençage progresse donc de front avec la cartographie des grands fragments : plus de 650 grands fragments ont ainsi été ordonnés le long du chromosome 14.



La seconde phase du "Projet Génome Humain", l'étape de finition, a été engagée dès l'obtention de l'ébauche, et s'est officiellement achevée en avril 2003, un demi-siècle après la découverte de la structrure en double hélice de l'ADN. L'objectif initial du projet était atteint : la mise à disposition d'une séquence presque dépourvue de trous et d'une exactitude de 99,99% (une erreur toutes les 10 000 bases). Contrairement à ce que suggère le caractère quasi complet de l'ébauche, la finition est la partie la plus fastidieuse d'un tel projet, car le comblement de certains trous exige un travail localisé, à façon. Certaines parties des chromosomes resteront toutefois absentes de la séquence "finie" : ces régions, monotones et très difficiles à séquencer, sont en général laissées de côté car elles sont à peu près vides de gènes.

Le Genoscope, pour sa part, a achevé le séquençage et l'analyse du chromosome 14 à l'automne 2002, et a publié ses conclusions dans la revue *Nature* le 6 février 2003. La partie séquencée du chromosome 14 est longue de 87 millions de nucléotides. C'est le quatrième chromosome humain dont la séquence finie a donné lieu à une publication, et le premier pour lequel cette séquence ne comporte aucun trou résiduel. Au moment de la publication, il s'agissait même de la plus longue molécule d'ADN dont la séquence soit connue d'un seul tenant. Le séquençage du chromosome 14 a coûté environ 10 millions d'euros, auxquels s'ajoutent plusieurs millions d'euros pour l'analyse et l'annotation. Quant au coût total du Projet Génome Humain, il est de l'ordre de 2,8 milliards de dollars (en dollars de l'année fiscale 1991).