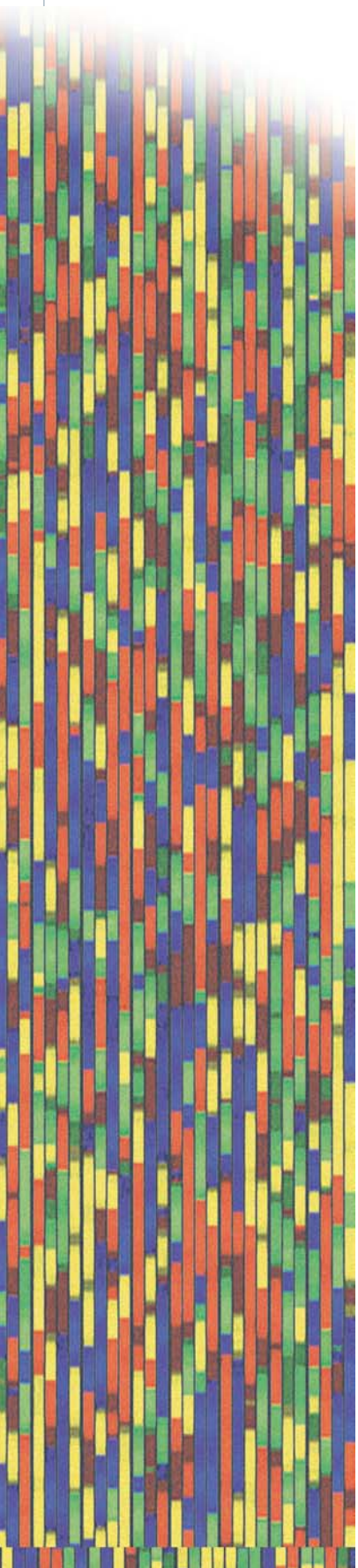
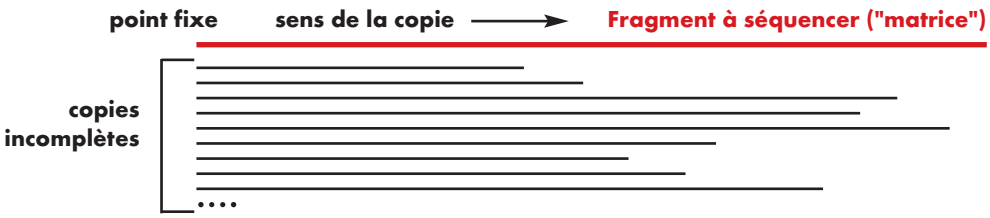
Le séquençage (lectures)

2

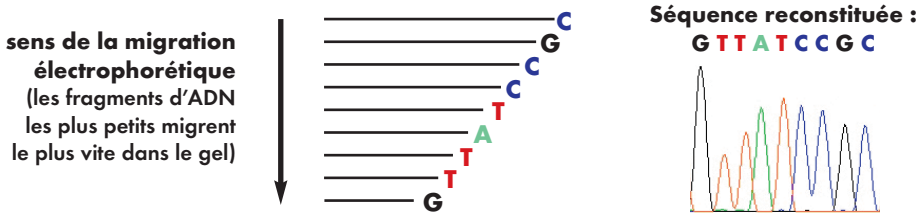


Les nucléotides, constituants élémentaires de l'ADN, sont de 4 types distincts désignés par les lettres A, C, G et T. Dans un fragment d'ADN à séquencer, des centaines de nucléotides s'enchaînent selon un ordre défini. Séquencer le fragment d'ADN consiste à déterminer cet ordre, ou séquence, un long mot écrit dans un alphabet à quatre lettres seulement.

Le principe du séquençage est de réaliser, à partir d'un point fixe, des copies incomplètes de la molécule d'ADN (la "matrice"), interrompues au hasard. L'interruption est provoquée par l'incorporation aléatoire, lors de la copie, d'un analogue de nucléotide qui bloque la réaction. Comme l'on travaille sur un très grand nombre de molécules de matrice, on obtient toutes les copies partielles possibles à partir du point fixe.



On sépare ensuite ces divers fragments selon leur taille en les faisant migrer dans un gel poreux sous l'action d'un courant électrique. Un tel gel permet de séparer deux copies incomplètes consécutives qui ont une différence de taille d'un seul nucléotide. Si l'on peut identifier le nucléotide au niveau duquel la copie s'est interrompue (nucléotide terminal) pour chacune des copies incomplètes, de la plus petite à la plus grande, on est alors en mesure de reconstituer la succession des nucléotides tout au long de la copie.



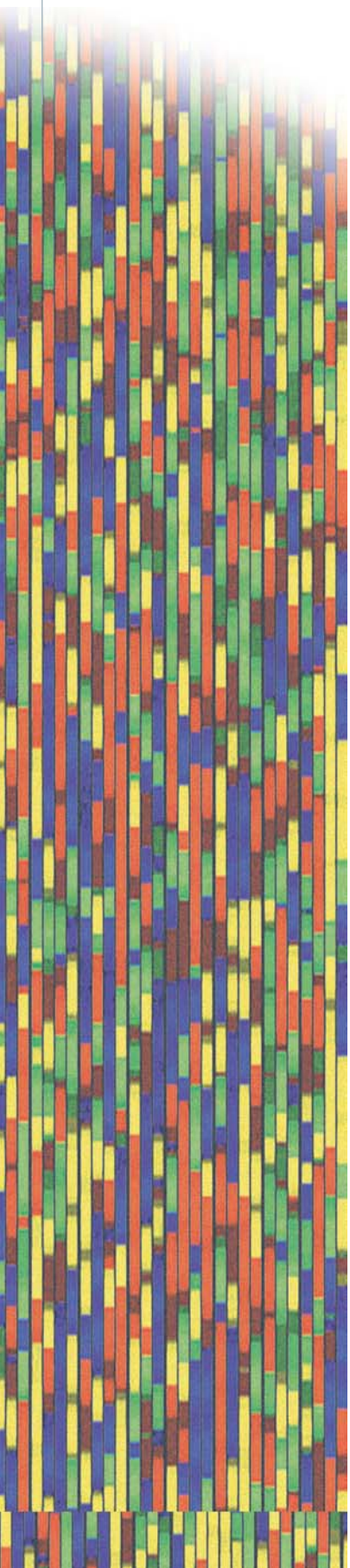
Mais comment identifie-t-on, en pratique, les nucléotides terminaux ? On réalise 4 séries de copies en parallèle, chacune incluant un analogue d'un des 4 types de nucléotides A, C, G ou T. Par exemple, toutes les copies intermédiaires d'une série seront terminées par un A. En outre, le composé provoquant l'interruption est fluorescent : le fragment copié peut ainsi être détecté automatiquement par un système optique à la fin de sa migration dans le séquenceur. Le signal obtenu à mesure que les copies partielles achèvent leur migration est interprété par un programme informatique qui reconstitue la séquence originale du fragment d'ADN analysé. Par opération unitaire, ou lecture, un séquenceur automatique peut déterminer l'enchaînement de 500 à 1000 nucléotides.

Quelques chiffres

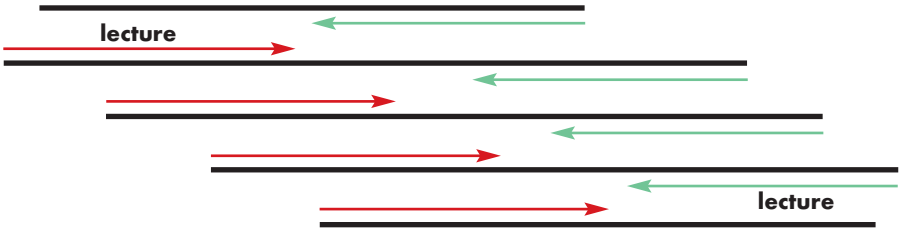
Chacun des 9 séquenceurs automatiques du Genoscope peut réaliser 96 lectures en parallèle. Les séquenceurs fonctionnent 24 heures sur 24, 7 jours sur 7, ce qui confère au centre une capacité moyenne de près de 14 000 lectures par jour. Ces lectures ont une longueur moyenne de 800 nucléotides. Le Genoscope séquence donc en moyenne 11 millions de nucléotides par jour. *(suite au dos)*

Le séquençage (assemblage)

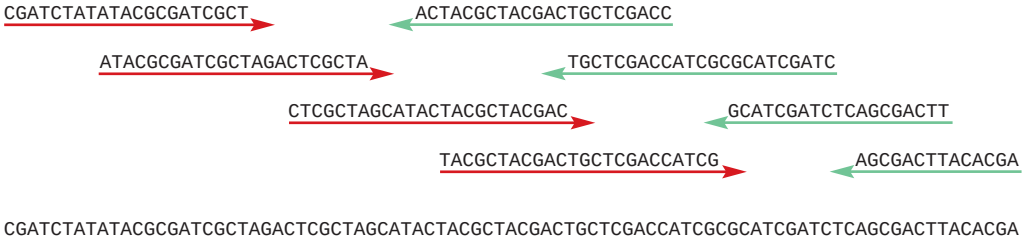
2



Comme les molécules d'ADN des génomes sont beaucoup plus longues que les lectures réalisées à l'aide des séquenceurs, il est nécessaire de pouvoir raccorder les lectures les unes aux autres. En pratique, on commence par casser de façon aléatoire le fragment d'ADN à séquencer en sous fragments de quelques milliers de nucléotides. En procédant à la lecture des extrémités d'un grand nombre de ces sous-fragments pris au hasard, on obtient des séquences qui se recouvrent en partie.



À l'aide de ces parties communes entre les séquences lues, on peut reconstituer la séquence de la totalité du fragment de départ. Cette opération d'assemblage, réalisée par des programmes informatiques, permet *in fine* de déterminer la séquence de molécules longues de plusieurs millions à plusieurs dizaines de millions de nucléotides.



L'assemblage rencontre plusieurs difficultés. Il faut notamment éviter l'écueil des séquences répétées, présentes à l'identique en plusieurs endroits du génome. Elles sont particulièrement abondantes dans les génomes de mammifères, et représentent 50% du génome humain. Par ailleurs, l'assemblage peut laisser des "trous" dans les régions pour lesquelles on n'a pas obtenu de lecture. Diverses méthodes existent pour combler ces trous.