Réalisations et projets du Genoscope

6



Le Genoscope est engagé, seul ou en collaboration, dans de nombreux projets portant sur des régions chromosomiques, voire des génomes entiers : plus de 20 "petits" génomes de bactéries ou d'eucaryotes unicellulaires ont été séquencés ou sont en cours de séquençage. Toutefois, certains projets se distinguent par leur ampleur ou leur intérêt.

Le "Projet Génome Humain"

Dès sa création en 1997, le Genoscope s'est engagé dans le "Projet Génome Humain", avec des centres de séquençage aux Etats-Unis, en Grande Bretagne, au Japon, en Allemagne et en Chine. Ce consortium international a tenu le premier de ses objectifs en annonçant, en juin 2000, l'assemblage d'une "ébauche" de la séquence du génome humain, librement accessible à tous les scientifiques. La réalisation de l'objectif final – une séquence presque sans trous et une exactitude de 99,99% – a été annoncée en avril 2003. L'effort de finition consenti depuis 2000 pour obtenir une séquence de qualité est essentiel pour une caractérisation exhaustive des gènes humains, ainsi que pour l'étude de leur implication dans les maladies.
A l'automne 2002, le Genoscope a achevé sa part du projet : le séquençage et l'annotation du chromosome 14 humain. Le travail s'est limité à la partie qui contient les gènes, longue de 87 millions de nucléotides (les 24 types de chromosomes du génome humain représentent 3,2 milliards de nucléotides). Nous avons publié le résultat de l'analyse du chromosome 14 dans la revue *Nature* le 6 février 2003. Il s'agit du quatrième chromosome humain dont la séquence finie a donné lieu à une publication, et du premier publié sans "trous" résiduels.

Un poisson au génome compact

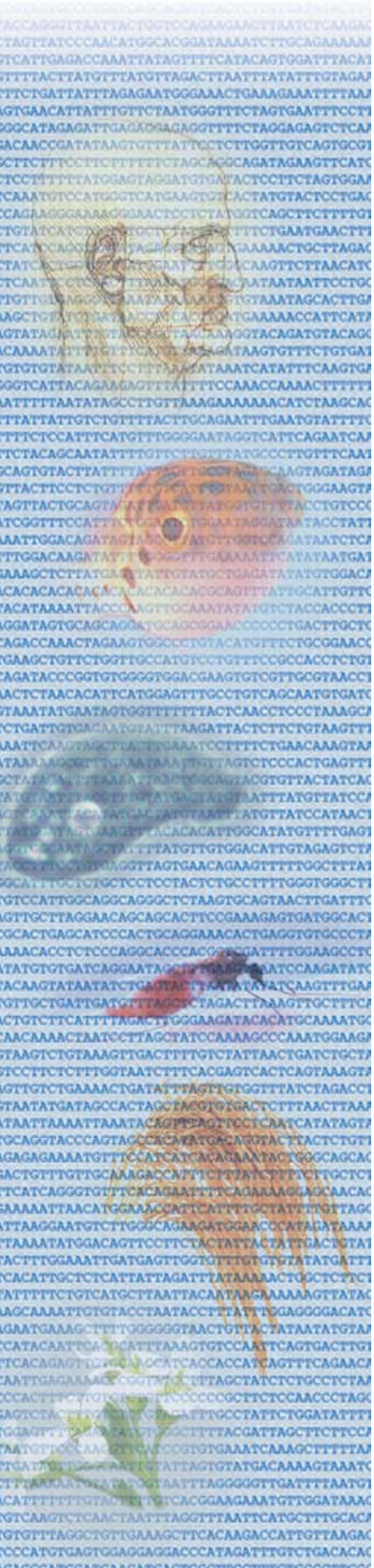
Rechercher les gènes est difficile, en particulier dans la séquence des génomes des organismes supérieurs tels que l'être humain, où ils sont dispersés et morcelés *(voir la fiche* Interpréter les séquences*)*. La comparaison entre séquences génomiques d'organismes plus ou moins apparentés est précieuse dans cette recherche, car elle permet de repérer les séquences codantes, davantage conservées au cours de l'évolution. En particulier, les Vertébrés possèdent un ensemble de gènes très comparable d'une espèce à une autre. Le Genoscope s'intéresse à un petit poisson tropical, *Tetraodon nigroviridis*, dont le génome de 400 millions de nucléotides est l'un des plus petits parmi les Vertébrés. Pourtant, ses gènes ne sont ni tronqués, ni moins nombreux. Cette compacité du génome de *Tetraodon* est due à la réduction des introns et des séquences comprises entre les gènes. Le séquençage du génome entier de ce poisson permet donc d'accéder aux gènes à moindre coût.
Nous avons d'ores et déjà utilisé les séquences d'une partie du génome de *Tetraodon* pour rechercher, dans le génome humain, des séquences conservées au cours de l'évolution entre l'être humain et ce poisson. Nous avons pu de la sorte estimer en mai 2000 le nombre des gènes humains à environ 30 000, bien moins que les valeurs alors communément admises. Les progrès du séquençage du génome humain ont depuis conforté cette estimation qui n'accorde à l'homme que le double environ des gènes de la mouche.

Analyses bioinformatiques

La procédure informatique développée pour la comparaison entre les séquences des génomes de l'homme et de *Tetraodon* a été appliquée à d'autres paires de génomes (anophèle-drosophile, riz-arabette) et sert désormais d'outil pour améliorer et corriger les annotations des génomes. D'autres ressources utilisées à cette fin sont les collections d'ADN complémentaires, copies des ARN messagers issus de la transcription des gènes. Le Genoscope a développé une importante expertise bioinformatique dans l'utilisation de ces ADNc pour délimiter de façon précise les gènes dans la séquence génomique. *(suite au dos)*

Réalisations et projets du Genoscope (suite)

6



L'arabette, plante modèle

La petite crucifère *Arabidopsis thaliana* (arabette) avait plusieurs atouts pour séduire les chercheurs en biologie végétale. Facile à cultiver et à transformer, elle possède en outre un génome de 125 millions de nucléotides seulement, une très petite taille parmi les végétaux. Elle est devenue plante modèle, objet de nombreuses études génétiques. Un consortium international, auquel a participé le Genoscope, a publié la séquence de son génome fin 2000. Il s'agit du premier génome de plante séquencé.

Le génome du riz, sésame pour ceux des autres céréales

De toutes les céréales, le riz possède le plus petit génome : 430 millions de nucléotides. Or l'arrangement des gènes le long des chromosomes est relativement similaire d'une céréale à une autre. Le riz est donc un bon point d'entrée pour caractériser les gènes des céréales, et les associer à tel ou tel trait agronomique. Le génome du maïs est en effet 5 fois plus gros, et celui du blé, 40 fois plus gros ! Mais le génome du riz est aussi intéressant en lui-même, car cette céréale nourrit la moitié de l'humanité. La connaissance du génome sera une aubaine pour les sélectionneurs qui cherchent à créer de nouvelles variétés.
Un consortium international s'est mis en place pour le séquençage du génome du riz. Le Genoscope y participe en séquençant le chromosome 12, long de 30 millions de nucléotides. Alors que deux initiatives privées ont déjà annoncé la production "d'ébauches" incomplètes de ce génome, le consortium s'attelle à la production d'une séquence finie, seule capable de satisfaire l'ensemble des attentes des scientifiques.

L'anophèle, un génome pour lutter contre le paludisme

Le Genoscope, avec l'institut Pasteur, s'était engagé dès 1998 dans le séquençage à grande échelle du génome du moustique *Anopheles gambiae*, principal vecteur du paludisme. Cet effort a été poursuivi par la constitution, en mars 2001, d'un consortium international pour le séquençage de l'anophèle, au sein duquel le Genoscope a assuré 10% du travail de séquençage. Une "ébauche" de la séquence du génome de l'anophèle est librement accessible depuis mars 2002. Avec le séquençage concomitant du parasite majoritairement responsable du paludisme, *Plasmodium falciparum*, et la finition de la séquence du génome humain, on dispose à présent des génomes des trois acteurs de la maladie.
Trois pistes parmi d'autres pour tirer parti du génome de l'anophèle : l'identification des gènes responsables de la résistance du moustique aux insecticides, des gènes codant les récepteurs olfactifs guidant l'insecte vers sa victime, ou encore des gènes intervenant dans l'immunité de certaines souches d'anophèle vis-à-vis de *Plasmodium*.

La diversité bactérienne révélée par la génomique

Notre vision du monde bactérien est fondée sur les bactéries que l'on parvient à isoler et cultiver en laboratoire. Or il ne s'agit que d'une petite fraction des bactéries présentes dans les milieux naturels. Au seuil des années 1990, la mise au point de techniques permettant d'amplifier, sans culture préalable, des séquences d'ADN caractéristiques de chaque espèce a révélé de nombreuses espèces nouvelles. Souvent, elles restent connues par cette seule séquence d'ADN ! Or la diversité métabolique de ces bactéries mystérieuses pourrait être davantage exploitée dans des applications industrielles, environnementales, ... Le Genoscope étudie par ces techniques la biodiversité microbienne de la station d'épuration d'Evry, afin de mieux connaître les mécanismes et les acteurs de l'épuration. Nous entreprenons en outre le séquençage de grands fragments d'ADN génomique afin d'effectuer un inventaires des gènes de l'ensemble des génomes de ces communautés microbiennes, un "métagénome" au moins aussi vaste que le génome humain.