

Le séquençage d'un génome

Le chemin est encore long pour le blé

Première étape dans la caractérisation génétique d'un organisme vivant, le séquençage du génome permet de lire à livre ouvert son code génétique. Charles Poncet, responsable de la plate-forme de génotypage haut débit de l'INRA de Clermont-Ferrand, nous explique les techniques de décodage de l'ADN et la complexité rencontrée sur blé tendre.



© N. Cornec



▲ Charles Poncet (INRA de Clermont-Ferrand) : « *Maintenant que le génome humain est séquencé, les compétences acquises par le Génoscope et les progrès technologiques vont accélérer le séquençage d'autres espèces.* »

« L'information génétique des organismes vivants s'écrit avec quatre lettres (A, C, G et T) correspondant aux quatre bases azotées (adénine, cytosine, guanine et thymine) qui constituent l'ADN. Leur enchaînement détermine le code génétique responsable des caractères propres à un individu.

Le séquençage du génome, le génome étant l'ADN de l'ensemble des chromosomes, consiste à déterminer la nature et l'ordre des bases azotées d'une espèce.

La taille d'un génome se mesure en nombre de bases. On parle de kilo-bases (Kb) pour une séquence de 1 000 bases d'ADN et de méga-bases pour 1 000 000 de bases

d'ADN. Les génomes peuvent aller de quelques kilo-bases pour les virus à des centaines de milliers de méga-bases pour certains organismes (tableau 1). La quantité d'ADN

d'un génome n'est pas proportionnelle à la complexité de l'espèce: le blé a par exemple un génome presque six fois plus grand que celui de l'homme.

Décoder l'ADN de « A à Z »

Le séquençage du génome ne veut pas dire séquençage des gènes: l'ADN n'est pas fait que de gènes codant pour la synthèse de protéines. Il comprend également une fraction non codante, composée essentiellement de séquences répétées, qui intervient dans la régulation des gènes et l'évolution du génome. Le génome humain possède 38 % d'ADN non codant tandis que le blé a plus de 80 % de séquences répétées dans son génome.

Taille du génome de différentes espèces (en méga-bases) (tab. 1)

Levure	13 Mb
Nématode	100 Mb
Drosophile (mouche)	180 Mb
Maïs	2500 Mb
Souris	3000 Mb
Homme	3000 Mb
Blé tendre	17000 Mb

La taille du génome n'est pas proportionnelle à la complexité de l'espèce. Elle est souvent liée à son histoire évolutive.

Pour être séquencé, l'ADN est d'abord découpé en de multiples morceaux de quelques bases à quelques kilobases. Une fois séquencés, ils sont ré-organisés les uns par rapport aux autres afin de retrouver le message initial.

Pour les génomes volumineux, il faut au préalable établir une carte physique des chromosomes avant de se lancer dans le séquençage pur. Cette opération consiste à mettre des « étiquettes » sur certaines parties du chromosome à l'aide de petites séquences connues. De cette manière, il sera plus facile de reconstituer l'ordre des fragments pour retrouver la séquence complète de l'ADN.

L'ADN se dévoile par fluorescence

Le séquençage en tant que tel s'appuie sur la méthode « chain-termination » mise au point par Sanger, prix nobel de chimie en 1980.

Les fragments d'ADN sont soumis à une enzyme, l'ADN polymérase, capable de dupliquer l'ADN. L'enzyme se fixe au brin codant d'ADN et synthétise un nouveau brin (une « copie ») à partir de nucléotides présents dans la cellule.

Pour séquencer l'ADN, on introduit dans le milieu réactionnel des nucléotides marqués

par fluorescence et capables de bloquer cette synthèse.

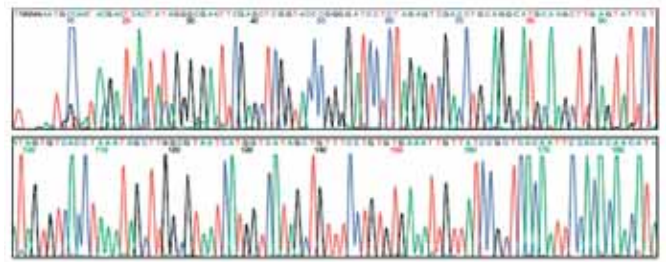
Dès que l'ADN-polymérase les utilise pour fabriquer le nouveau brin d'ADN, elle est bloquée et la duplication s'arrête, libérant un fragment d'ADN plus court que l'original.

▶ Des nucléotides marqués par fluorescence stoppent la synthèse d'ADN et libèrent des fragments de toutes les tailles.

Cette réaction ayant lieu sur des milliers de copies du fragment d'ADN à séquencer, le milieu compte au final des brins marqués de différentes longueurs.

Ces fragments sont ensuite déposés dans un séquenceur qui sépare les molécules selon leur taille. Elles passent ensuite devant une fenêtre de lecture où un laser excite la fluorescence des marqueurs. La couleur émise (rouge, vert, jaune ou bleu) renseigne alors sur le type de base ayant bloqué la synthèse d'ADN (*graphique*). Chaque base correspond à une couleur. Par exemple, si le fragment de dix bases est vert, il se termine par une guanine; si le fragment de onze bases est rouge, il se termine par une cytosine, et ainsi de suite.

Il y a suffisamment d'exem-



▲ Lecture de séquence sortie d'un séquenceur automatique d'ADN: chaque courbe de fluorescence correspond à une base d'ADN. Les pics renseignent sur le type et l'ordre des bases d'ADN.

plaires dans le mélange pour obtenir la séquence complète du fragment.

Pourquoi est-ce compliqué sur le blé ?

Le séquençage du génome du blé représente un vaste chantier: 17 000 méga-bases à séquencer dont 80 % de séquences répétées.

Le blé tendre est issu de la fusion de trois génomes de blés ancestraux A, B et D. Ces trois génomes comprennent chacun 7 paires de chromosomes: le blé possède 42 chromosomes en tout.

Un consortium international (International wheat genome sequencing consortium) a vu le jour en 2005. Il regroupe des partenaires d'une quinzaine de pays (instituts, sélectionneurs, laboratoires de génomique...) autour du séquençage du génome du blé. Chaque partenaire s'est réparti les 21 chromosomes. L'INRA de Clermont-Ferrand a la charge de séquencer le chromosome 3B (le troisième du génome B) dans le cadre du programme européen « Triticeae genome ».

Avec une taille de 995 méga-bases, ce chromosome est le plus grand des chromosomes du blé et représente à lui seul 2,5 fois le génome complet du riz.

▶ Le blé tendre devrait être séquencé d'ici 5 à 10 ans.

Après trois ans et demi de travail, l'INRA de Clermont-Ferrand vient d'établir la première carte physique d'un chromosome du blé. Ce travail



▲ Avant d'être séquencé, l'ADN est découpé en petits morceaux qui sont ensuite « photocopiés » en milliers d'exemplaires par PCR.

préalable ouvre maintenant la porte au séquençage en tant que tel qui devrait prendre entre 5 et 10 ans. Compte-tenu de l'importance des séquences répétées, les machines capables de séquencer des fragments d'ADN longs (800 à 1 000 bases) sont mieux adaptées sur blé, mais leur débit est moindre (100 kg-bases en 2 heures).

Le séquençage n'est qu'une première étape. Une fois que le livre du blé sera décrit, il restera à le comprendre pour mieux le valoriser! » ■

Nicolas Bousquet
n.bousquet@perspectives-agricoles.com

Le séquençage a évolué au fil des technologies. Aujourd'hui, on parle de séquençage à très haut débit avec des machines capables de séquencer 100 mégabases en 7 heures par fragments de 30 à 400 bases. Il y a 2 ans, le séquenceur de référence analysait 100 kilo-bases en 2 heures par fragments de 1000 bases. ▼

