

Un être vivant dont le génome vient d'être entièrement séquencé : *Haemophilus influenzae* Rd

L'ensemble du génome chromosomique de la bactérie *Haemophilus influenzae* Rd vient d'être séquencé par un groupe américain conduit par Craig Venter (Institute for Genomic Research, Gaithersburg) et Hamilton Smith (Johns Hopkins University, Baltimore) [1]. Il s'agit de la première publication d'un génome complet d'organisme vivant; seuls des génomes de virus avaient été séquencés en entier à ce jour. Ce travail va permettre d'avoir un dictionnaire de toute l'information nécessaire à la vie [2]. L'approche utilisée repose sur le séquençage et l'assemblage de fragments d'ADN pris au hasard à partir du chromosome entier de *H. influenzae* (1 830 137 paires de bases). Deux banques d'ADN génomique contenant des inserts d'environ 2 kb (plasmide) ou de 15 à 20 kb (phage) ont été réalisées à partir du génome de *H. influenzae*. Les deux extrémités des inserts ont été séquencées à partir d'amorces s'hybridant avec les vecteurs (soit une séquence d'à peu près 450 pb lue par extrémité). 24 304 fragments de séquences ont ainsi été obtenus puis ordonnés grâce à un programme informatique. Avec ce programme, 140 séquences continues ont été ordonnées en 42 groupes physiques et donc autant de «trous physiques» (pas de matrice d'ADN pour ces régions). Ces régions manquantes ont été obtenues pour la plupart en criblant la banque d'ADN phagique à l'aide d'amorces oligonucléotidiques produites à partir des extrémités des séquences continues, puis séquencées. Le génome contient 1 743 gènes avec, ce qui est classique chez les bactéries, très peu d'ADN non codant. Plus de 40 % de ces gènes (736) n'ont pas de contrepartie chez d'autres organismes. Maintenant que ces gènes sont identifiés, il va être possible de les inactiver chez la bactérie et d'établir leur fonction. Les 1 007 gènes restants ont une contrepartie dans d'autres organismes et incluent, par exemple, les gènes co-

dant pour les enzymes métaboliques, les facteurs de transcription et de nombreux autres équipements de la cellule vivante [2]. Cette même équipe rapporte également, dans le même numéro de *Science*, la fréquence et la distribution des séquences signal de capture de l'ADN (5' AAGTGCAGT). *H. influenzae* est en effet une bactérie à Gram⁻ naturellement transformable qui capture l'ADN de bactéries mortes de la même espèce en reconnaissant cette séquence de 9 pb. Dix-sept pour cent de ces 1 465 sites sont en groupes de paires inversées. Nombre de ces tandems sont localisés à l'extrémité 3' des gènes et sont capables de former des structures en épingle à cheveux dans l'ARN messager qui pourraient fonctionner comme des signaux de terminaison de transcription [3]. *H. influenzae* est un commensal des voies respiratoires supérieures chez l'homme mais peut entraîner des otites ou infections respiratoires, voire des méningites, surtout chez l'enfant, qui sont gravissimes avec survenue de séquelles neurologiques dans 50 % des cas. La souche séquencée (Rd) est inoffensive mais la comparaison de souches virulentes à des souches commensales devrait permettre d'identifier des gènes impliqués dans la virulence.

L'équipe de Venter a également séquencé (mais non publié) la séquence de *Mycoplasma genitalium*, une bactérie associée à des infections du tractus génital. La séquence d'une archaéobactérie, *Methanococcus jannaschii*, une bactérie vivant à 100 °C, est prévue pour la fin de l'année [2]. Cette équipe a reçu une trentaine de demandes de collaboration pour séquencer différentes bactéries (agent de la tuberculose, de la syphilis...). «Les gens pensaient que ce type de projet représentait de nombreuses années et de nombreux millions de dollars» dit Venter. «Nous venons de montrer que cela peut être fait en moins d'un an pour moins de

50 cents (2,50 F) par base» [2]. Le génome d'*Escherichia coli* est attendu pour 1997.

Le séquençage de bactéries appartenant aux groupes des eubactéries et des archaéobactéries devrait permettre de mieux comprendre les données de l'évolution et donc d'établir plus clairement les relations entre les trois grands règnes vivants – eubactéries, archaéobactéries, eucaryotes [4]. De plus, le séquençage d'archaéobactéries devrait permettre d'expliquer comment des enzymes peuvent fonctionner à des températures de 100 °C avec des débouchés industriels pour la catalyse à haute température.

Le séquençage complet d'un organisme vivant, soit-il une simple bactérie, est le premier pas vers le séquençage du génome humain.

E.D.

1. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, McKenney K, Sutton G, Hugh WF, Fields C, Gocayne JD, Scott J, Shirley R, Liu LL, Glodek A, Kelley JM, Weidman JF, Phillips CA, Spriggs T, Heldblom E, Cotton MD, Utterback TR, Hanna MC, Nguyen DT, Saudek DM, Brandon RC, Fine LD, Fritchman JL, Fuhrmann JL, Geoghagen NSM, Gnehm CL, McDonald LA, Small KV, Fraser CM, Smith HO, Venter JC. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269: 496-512.

2. Nowak R. Bacterial genome sequences bagged. *Science* 1995; 269: 468-70.

3. Smith HO, Tomb JF, Dougherty BA, Fleischmann RD, Venter JC. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. *Science* 1995; 269: 538-40.

4. Denamur E, Picard B. De la génétique des populations bactériennes à l'épidémiologie des maladies infectieuses. *médecine/sciences* 1995; 11: 1399-406.