

Chroniques génomiques

La génération suivante, déjà...

Bertrand Jordan



L'Histoire s'accélère

Après une stagnation technologique qui a duré près de trente ans, le séquençage d'ADN a fait un bond en avant avec l'arrivée de nouveaux systèmes (454/Roche, puis Solexa/Illumina et SOLiD/Life Technologies (ex-ABI)) [1]. Ceux-ci sont maintenant bien implantés et produisent en quelques jours des gigabases de séquence (brute) à un coût de quelques dollars par mégabase - soit des performances supérieures d'au moins deux ordres de grandeur à celles de la technique classique publiée par Fred Sanger en 1977¹. Mais à peine la communauté scientifique a-t-elle apprivoisé ces nouvelles machines et commencé à explorer leurs multiples applications, que déjà pointent à l'horizon des méthodes de « troisième génération » qui constituent un indéniable saut conceptuel et promettent des performances encore inédites [2].

Plus de sous-clonage, plus de librairies, plus d'amplification...

Une des innovations majeures des techniques de deuxième génération était l'élimination du clonage des fragments d'ADN avant séquençage, grâce à d'astucieux procédés d'amplification locale suivis d'un séquençage simultané en masse [1]. Les méthodes de troisième génération, elles, font l'impasse sur l'amplification et lisent directement des molécules d'ADN individuelles, sans avoir besoin d'en produire préalablement des millions d'exemplaires identiques. Elles réalisent en fait le vieux rêve, formulé dès les années 1980, de lire les bases une à une le long d'un segment d'ADN, et effectuent cette lecture simultanément sur des millions de fragments. Du coup, il devient possible de travailler à partir d'une toute petite quantité d'ADN (quelques nanogrammes), mais aussi d'envisager la lecture ininterrompue de milliers, de dizaines de milliers de bases, ce qui bien sûr facilite énormément l'assemblage des séquences



Marseille-Nice Génopole,
case 901,
Parc Scientifique de Luminy,
13288 Marseille Cedex 9, France.
brjordan@club-internet.fr

obtenues au cours d'un projet de séquençage - tout en atteignant des débits jusqu'ici hors de portée.

Une machine déjà opérationnelle, le Helicoscope, et une autre dans les starting-blocks

L'entreprise Helicos commercialise déjà depuis 2008 son Helicoscope, qui réalise la lecture d'ADN sur une seule molécule. Sans entrer dans les détails, le mode opératoire repose sur la synthèse séquentielle d'un deuxième brin pratiquée sur des « milliards » de molécules individuelles d'ADN fixées à la surface d'une chambre de flux². La séquence est donc lue au fur et à mesure sur chaque fragment ; l'instrument pratique des lectures courtes (35 bases) mais atteint un débit élevé, une trentaine de gigabases par session. Comme on pouvait s'y attendre, la mise au point finale du système n'a pas été sans aléas : il exige une très grande sensibilité de détection et un fonctionnement impeccable des différentes réactions biochimiques. Le planning très ambitieux d'Helicos pour 2009 a dû être revu à la baisse. L'entreprise a néanmoins réussi à construire et à vendre deux ou trois machines (à un prix de l'ordre du million de dollars pièce), a publié un article de « preuve de concept » [3], et a depuis réalisé avec le Broad Institute (le plus puissant centre de séquençage outre-Atlantique) une intéressante étude - non encore publiée - sur les origines de réplication de l'ADN chez plusieurs espèces de levures fissipares (*fission yeast*).

¹ Et largement améliorée et automatisée depuis, bien que son principe n'ait pas changé.

² Le site de l'entreprise (<http://www.helicosbio.com/>) comporte une très jolie vidéo montrant le principe de la méthode. Evidemment, tout y paraît très simple...

Le système mis au point par *Pacific BioSciences* repose, lui, sur des « nanopuits » qui forment des « *zero mode waveguides* »³ permettant l'observation d'une seule molécule de polymérase en train d'ajouter un nucléotide fluorescent à un brin d'ADN⁴. Lorsqu'il s'agit du « bon » nucléotide, celui dont la base est complémentaire au brin à séquencer, il reste au contact de la polymérase durant plusieurs millisecondes, alors qu'une base différente ne fait que passer. La technique a elle aussi fait l'objet d'un article [4], et l'entreprise prévoit la mise sur le marché courant 2009 d'un instrument capable de séquencer plusieurs gigabases par heure, avec des lectures de plus de mille bases. Notons que les polymérases employées par *Helicos* comme par *Pacific BioSciences* sont largement « améliorées » par mutagenèse dirigée afin de présenter des caractéristiques adaptées au séquençage...

Les espoirs des nanopores

De nombreuses équipes, enfin, fondent leurs espoirs sur l'emploi de nanopores, minuscules ouvertures pratiquées dans une membrane biologique ou un substrat de silicium. L'idée est de faire passer la molécule d'ADN à travers un tel canal et de détecter de subtiles modifications électriques lors de ce passage - et, encore plus subtilement, de différencier les modifications produites par chacune des quatre bases... La chose n'est nullement chimérique, et il existe déjà tout un *corpus* de résultats montrant les potentialités de cette approche [5]. Le problème principal semble être actuellement de trouver une manière de ralentir la molécule d'ADN lorsqu'elle traverse le nanopore sous l'influence d'un champ électrique, afin de pouvoir effectivement enregistrer le signal qui donnera la séquence (Figure 1). Parmi les firmes les plus avancées, on peut citer *Oxford Nanopore* (Royaume-Uni), *Nabsys* et *Sequenom* (États-Unis), ainsi que plusieurs laboratoires académiques principalement aux États-Unis (voir [5]).

La deuxième génération n'a pas dit son dernier mot

En dépit de ces perspectives mirifiques, le succès commercial du séquençage de troisième génération n'est pas assuré. Les trois mousquetaires de la deuxième génération sont maintenant solidement installés, et ne cessent d'améliorer les performances de leurs systèmes : la dernière mouture du *SOLiD* (version 3) est donnée pour 20 gigabases par session (une semaine), et *Illumina* annonce 3 gigabases/jour. Et les centaines de laboratoires qui viennent d'investir un demi-million de dollars pour l'une de ces machines ne sont sans doute pas prêts à en



Figure 1. Molécule d'ADN et nanopore.

Nanotechnology today

<http://nanotechnologytoday.blogspot.com/2007/01/dna-sequencing-technologies.html>

changer, d'autant plus qu'ils ont déjà bien du mal à « suivre » au niveau de l'informatique et de la bioinformatique [1]... Il est certain en tous cas que les données de séquence, quel que soit l'organisme (ou l'individu) envisagé, deviennent de plus en plus accessibles, ce qui va changer la manière d'aborder bien des questions en recherche fondamentale. On commence même à voir poindre le séquençage ultra-rapide pour le diagnostic médical - mais il faudra bien une autre chronique pour en détailler les applications ! ♦

Next-next gen sequencing, already...

RÉFÉRENCES

1. Jordan B. Une révolution longuement attendue. *Med Sci (Paris)* 2008 ; 24 : 869-73.
2. Check Hayden E. Genome sequencing : the third generation. *Nature* 2009 ; 457 : 768-9.
3. Harris TD, Buzby PR, Babcock H, et al. Single-molecule DNA sequencing of a viral genome. *Science* 2008 ; 320 : 106-9.
4. Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science* 2009 ; 323 : 133-8.
5. Branton D, Deamer DW, Marziali A, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol* 2008 ; 26 : 1146-53.

TIRÉS À PART

B. Jordan

³ Guides d'onde fonctionnant sur le mode zéro, ce qui permet de limiter la détection à un très petit volume de (20 zeptolitres, 20×10^{-21} litres) et donc d'observer la molécule de polymérase sans interférence de la solution environnante.

⁴ Encore une belle vidéo, plus détaillée (et plus réaliste) que la précédente : <http://www.pacificbiosciences.com/>