

Chroniques génomiques

Séquence : le juste prix...

Bertrand Jordan



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

Depuis l'apparition des nouvelles techniques de séquençage, l'idée de la « séquence personnelle » a fait du chemin, et les annonces se succèdent, entretenant une certaine confusion. Certes, chacun a compris que le coût avait considérablement baissé, et qu'il n'était plus question, comme lors des temps héroïques du programme génome, de dépenser un ou deux milliards de dollars pour obtenir une telle information... mais où en est-on exactement ?

Rappelons d'abord qu'il y a séquence et séquence : le tarif n'est pas du tout le même s'il s'agit d'une lecture un peu approximative, séquence brute ou *raw sequence*, comportant un ou deux pour cent d'erreurs et un bon nombre de « trous », ou d'une séquence « finie » dont le taux probable d'erreurs (jamais nul !) est inférieur à un pour dix mille, et dans laquelle les trous sont rares. Ce qui sort des machines - qu'il s'agisse de séquenceurs *Applied 3730* à 96 capillaires, dernière incarnation de la bonne vieille technique de Fred Sanger, ou des systèmes *454 (Roche)*, *Solexa/Illumina* ou *SOLiD/Life Technologies* -, c'est de la séquence brute qui, même une fois assemblée, gardera son caractère approximatif. Les machines actuelles [1] sont effectivement capables de lire plusieurs milliards de bases (gigabases) en une session (une ou quelques journées), à un coût de quelques milliers de dollars : un génome humain, en somme (trois gigabases), mais un génome « brut ». Il est à peu près inutilisable : si on veut le comparer à celui d'une autre personne, sachant que le nombre de différences ponctuelles entre deux humains est de l'ordre de trois millions, soit un pour mille, il est clair qu'un taux d'erreur de un pour cent est

rédhibitoire... Pour en arriver à la séquence « finie », il faudra répéter la lecture dix, vingt ou trente fois : comme les erreurs se produisent de manière à peu près aléatoire, un point donné du génome sera lu dix-neuf fois comme un G, une fois comme un A, et l'on considèrera donc qu'il s'agit d'un G... La redondance est ainsi essentielle, son niveau exact dépend un peu des qualités intrinsèques de la technique : dix ou quinze passages suffisent pour Sanger ou *454*, alors que pour *Illumina* ou *SOLiD* on visera plutôt trente déterminations. Bien entendu, cela multiplie d'autant le coût, et il faut éventuellement y ajouter des expériences complémentaires pour lever certaines ambiguïtés. Notons aussi que l'on parle ici de séquence humaine, pour laquelle on dispose d'une référence qui va faciliter l'alignement de toutes les petits fragments lus (longs de 200 à 400 bases pour *454*, de trente à cinquante pour *Illumina* ou *SOLiD*). S'il s'agissait de lire l'ADN d'un organisme totalement nouveau (séquençage *de novo*) l'assemblage serait plus difficile et la redondance nécessaire tout comme la difficulté des calculs monterait en flèche...

Disons aussi, au risque de décevoir le lecteur en quête d'informations gravées dans le marbre, que tout ce qui peut être affirmé ici est éminemment provisoire et révisable. Pour le plus grand bonheur des chercheurs, le marché du séquençage de deuxième génération comporte trois acteurs principaux (cités ci-dessus), qui se livrent une concurrence effrénée et améliorent constamment les performances de leurs appareils. Nouveaux kits de réactifs, cellules de réaction plus sophistiquées, programmes informatiques plus performants se succèdent



de trimestre en trimestre, et deux au moins des fabricants assurent qu'il atteindront un chiffre de cent gigabases par session d'ici la fin de l'année... Les frais (amortissement des systèmes, et réactifs) se répartissent donc sur une quantité plus importante de séquence, diminuant d'autant le coût unitaire : on est déjà au voisinage de un dollar par mégabase de séquence brute, et ce n'est pas fini. D'autant plus que la troisième génération pointe son nez [2] et promet des performances encore plus étonnantes - et, en attendant, stimule (s'il en était encore besoin) les acteurs actuellement bien implantés. Il faut aussi prendre garde aux affirmations publicitaires. Quand IBM annonce le « génome à 100 dollars » grâce à une technologie utilisant les nanopores [3], il s'agit d'un espoir, d'un but, d'une projection à terme et non d'une réalité actuelle : un de ces *Forward looking statements* trouvés dans les communiqués de presse des entreprises, et dont un paragraphe final en petit caractères souligne la nature spéculative...

Quelles sont donc les offres réelles sur ce marché que certains estiment promis à un grand avenir ? Disons d'abord, pour fixer les idées, que le coût de la lecture d'un génome humain par la technique Sanger de dernière génération s'élève à quelques dizaines de millions de dollars. La première entreprise à se lancer sur ce marché en utilisant les nouvelles méthodes rapides a été, fin 2007, *Knome* (anciennement *Cambridge Genomics*), dont l'un des fondateurs est George Church, un des « gourous » du nouveau séquençage. Son tarif de 350 000 dollars était relativement dissuasif, et son mode de fonctionnement précis tout comme ses clients sont restés un peu obscurs ; *Knome* relance actuellement une proposition à 99 500 dollars, nettement plus compétitive [4]. L'offre d'*Illumina*, que nous avons évoquée récemment [5], semble aujourd'hui être la plus sérieuse (<http://www.everygenome.com/>), et son coût (48 000 dollars) est presque abordable. Le fait qu'il s'agisse du plus important fournisseur mondial de séquenceurs « nouvelle génération » inspire aussi confiance... Le premier client, Hermann Hauser, a reçu ses données fin août 2009. La redondance annoncée est de 30 fois, *a priori* suffisante pour un résultat de bonne qualité. Le total de séquence brute est dans ce cas de 110 gigabases, soit un montant de 0,4 dollars par mégabase. On peut imaginer qu'*Illumina* « casse » un peu les prix - et peut-être même vende à perte au début - afin de prendre position sur ce marché... L'entreprise a tout intérêt à assurer ses positions, puisqu'un concurrent, *Complete Genomics*, prend déjà des commandes pour des génomes à 5 000 dollars... Cette firme, dont le responsable scientifique Radoje Drmanac est un des pontes du séquençage par hybridation depuis les années 1990 [6], voit très grand : elle compte « produire » 20 000 génomes en 2010 ! Elle entend jouer sur une technique originale (et dont on ne connaît pas encore tous les détails) mais aussi sur les économies d'échelle : le prix de 5 000 dollars par génome s'applique à des commandes d'au moins deux cents génomes... Il ne s'agit donc pas d'individus isolés souhaitant « se faire séquencer », mais de programmes publics comme le *1 000 genomes programme* [7], ou encore de nouvelles études d'association à grande échelle s'intéressant non plus à 500 000 Snip mais carrément aux séquences complètes de chaque personne étudiée. Et bien sûr, à moyen terme, le PDG de *Complete Genomics*, Clifford Reid, espère que l'on séquencera l'ADN de chaque enfant à la naissance¹ : il verrait bien son entreprise jouer un rôle majeur dans ce secteur prometteur. En attendant, il est à prévoir que les firmes qui proposent d'établir votre

profil génomique personnel grâce à l'analyse de centaines de milliers de Snip (Navigenics, 23and Me... [8]) vont bientôt ajouter la séquence à leur catalogue²...

On peut affirmer qu'aujourd'hui le tarif d'*Illumina* (48 000 dollars) est sans doute le plus réaliste : il correspond à un service effectivement offert et à une technique qui fonctionne sur des centaines de machines installées dans de nombreux laboratoires. À ce prix, le marché est encore restreint, tant du côté des individus que de celui des structures de recherche (la plupart des projets imaginables supposant de lire des centaines sinon des milliers de génomes). Mais la course effrénée à laquelle se livrent les acteurs actuels du marché, l'émergence de nouvelles méthodes de troisième génération encore plus efficaces, et la demande de grands programmes de recherche financés par les crédits dégagés dans le cadre de la stimulation de l'économie américaine³ se conjuguent pour assurer que l'on arrivera très vite au génome à 5 000 et, pourquoi pas, à 1 000 dollars. Ce changement d'échelle, qui va se produire dans les toutes prochaines années, aura des implications considérables pour nos systèmes de santé et posera des questions éthiques d'une intensité inédite... ♦

How much for a sequence?

CONFLIT D'INTÉRÊTS

L'auteur déclare n'avoir aucun conflit d'intérêts concernant les données publiées dans cet article.

RÉFÉRENCES

- Jordan B. Une révolution longuement attendue. *Med Sci (Paris)* 2008 ; 24 : 869-73.
- Jordan B. La génération suivante, déjà... *Med Sci (Paris)* 2009 ; 25 : 649-50.
- Bourzac K. Big blue's DNA-reading chips: IBM scientists are developing microchips for genome sequencing. *MIT Technology Review*, October 6, 2009 (<http://www.technologyreview.com/biomedicine/23589/>)
- <http://www.knome.com/service/knomecomplete.html>
- Jordan B. Chacun sa séquence ! *Med Sci (Paris)* 2009 ; 25 : 647-8.
- Drmanac R, Drmanac S, Chui G, et al. Sequencing by hybridization (SBH): advantages, achievements, and opportunities. *Adv Biochem Eng Biotechnol* 2002 ; 77 : 75-101.
- <http://www.1000genomes.org/page.php>
- Jordan B. « Génome personnel », gadget ou révolution ? *Med Sci (Paris)* 2008 ; 24 : 91-4.
- Maher B. Exome sequencing takes centre stage in cancer profiling. *Nature* 2009 ; 459 : 146-7.

² Je n'ai pas discuté ici l'approche alternative qui consiste à purifier « en masse » l'ADN qui correspond aux séquences exprimées et à séquencer cet « exome » [9]. Ce procédé qui mériterait un article à lui seul n'a sans doute pas un très grand avenir compte tenu du coût rapidement décroissant du séquençage intégral.

³ Aux États-Unis, sur les *stimulus funds* reçus par le *National Institutes of Health* pour l'année fiscale 2009, les trois quarts (soit 145 millions de dollars) vont au séquençage...

¹ Dans les pays riches, soit environ 20 millions de naissances par an.

TIRÉS À PART

B. Jordan