



médecine/sciences 2001 ; 17 : 290-3

Génome : les méandres de la technologie

Lorsqu'on célèbre l'aboutissement d'un programme, la vision rétrospective de son déroulement tend toujours à gommer les incertitudes, les faux départs et les impasses pour privilégier un déroulement harmonieux et logique – mais bien souvent reconstruit *a posteriori*. La saga du Génome humain n'échappe pas à cette tendance. Il est pourtant instructif d'examiner les inflexions qui ont émaillé son histoire, ne serait-ce que pour se persuader que même un projet aussi « routinier » (d'après ses détracteurs) que le séquençage de notre génome ne s'est pas déployé de manière totalement ordonnée et prévisible. Je me limiterai ici au plan de la technologie ; on pourrait sûrement s'attacher avec profit aux aléas politico-organisationnels de ce grand projet (un peu comme l'a fait Robert Cook-Deegan pour ses premières années dans son excellent livre [1]), mais je ne me risquerai pas – du moins cette fois-ci – sur ce terrain.

Les surprises de la cartographie

Bien que ses promoteurs aient beaucoup parlé de séquençage (ne serait-ce que pour emporter l'adhésion des parlementaires nord-américains et débloquent ainsi les financements), le programme Génome, dans ses premières années, fut logiquement centré sur la construction des cartes génétiques et physiques. La cartographie génétique, fondée sur l'analyse de la transmission de marqueurs polymorphes au sein de grandes familles,

allait progresser de manière relativement continue, sur la lancée du concept initial proposé par David Botstein [2]. Les premiers marqueurs employés, les RFLP (*restriction fragment length polymorphisms*) firent progressivement place aux microsatellites (séquences répétées polymorphes du type [GT]_n, n étant variable). Aujourd'hui, ce sont les SNP (*single nucleotide polymorphisms*, prononcé *snips*) qui tiennent le haut du pavé. Mais il s'agit toujours de détecter de petites variations dans la séquence de l'ADN, et l'étude de familles reste d'actualité. Seules la nature des marqueurs polymorphes, la méthode de détection... et l'échelle de l'analyse ont changé. Le projet mené par Jean Weissenbach au Généthon joua de ce point de vue un rôle central, en montrant l'efficacité d'une organisation centralisée et en aboutissant dès 1992 [3] à une carte génétique à base de microsatellites détaillée et fiable – qui reste encore aujourd'hui le fondement de nombreuses études de génétique médicale.

Pour la cartographie physique, l'histoire des technologies est beaucoup plus chaotique. A la fin des années 1980, il était beaucoup question de méthodes nouvelles particulièrement adaptées à l'étude de grandes régions d'ADN : l'analyse par électrophorèse en champs pulsés, le saut le long du chromosome (dont le logo de ces chroniques est un souvenir), et le clonage de très grands fragments d'ADN dans des chromosomes artificiels de levure ou YAC (*yeast arti-*

ficial chromosomes) [4]. La découverte d'enzymes de restriction à site rare, coupant l'ADN humain en moyenne toutes les quelques centaines de kilobases, combinée à la séparation des grands fragments obtenus par électrophorèse en champs pulsés [5], devait permettre de construire des « macro-cartes de restriction » s'étendant sur un chromosome entier, à l'image de ce qui avait été réalisé pour des virus avec des enzymes et des méthodes de séparation conventionnelles. Le saut le long du chromosome, acrobatique procédé fondé sur la circularisation de grands fragments d'ADN génomique et permettant de cloner ensemble deux fragments distants de centaines de kilobases dans le génome [6] devait compléter cette approche *top-down* (du haut vers le bas) en fournissant sondes et points de repère.

En réalité, la cartographie par électrophorèse en champs pulsés allait se révéler très aléatoire, notamment en raison des méthylations partielles de l'ADN génomique aux sites des enzymes de coupure, qui faisaient apparaître ou disparaître ces sites en fonction de l'état physiologique des cellules utilisées pour préparer l'ADN à analyser. A cela s'ajoutaient des difficultés techniques pour séparer de très grands fragments de manière reproductible. Cette méthode rendit certes de grands services pour étudier précisément des régions couvrant une ou plusieurs mégabases autour d'un gène particulièrement intéressant, mais elle échoua en tant que

technique de cartographie générale. Le saut le long du chromosome, malgré son emploi par le groupe de Francis Collins dans le clonage du gène de la mucoviscidose* [7], s'avéra être une technique extrêmement délicate et sujette à artéfacts. Elle disparut progressivement des laboratoires – après avoir fait perdre leur latin à bien des chercheurs.

Restent les YAC. L'article de Burke, Carle et Olson en 1987** [8] fit l'effet d'une bombe : cloner 500 kilobases et même une mégabase d'ADN humain dans la levure, c'était réellement un résultat révolutionnaire, et qui arrivait à son heure. Dans les laboratoires du DOE (*department of energy*), à Lawrence Livermore (Californie, USA) et Los Alamos (Nouveau Mexique, USA), des équipes dotées d'importants moyens techniques et informatiques essayaient de construire les cartes des chromosomes 16 et 19 à partir de la comparaison des cartes de restriction de milliers de cosmides. Tâche héroïque, et sans doute impossible. Le Japonais Yuji Kohara avait bien réussi, presque tout seul, à construire en 1987 une telle carte pour le génome d'*Escherichia coli* à partir d'un jeu de trois mille quatre cents segments clonés dans le phage lambda [9]. Mais *E. coli* possède en tout quatre mégabases d'ADN, alors que les chromosomes 16 et 19 en comptent chacun près d'une centaine... On pouvait logiquement penser que la représentation d'un chromosome par quelques centaines de YAC (au lieu d'une dizaine de milliers de cosmides) allait considérablement simplifier le puzzle et rendre viable cette approche *bottom-up* (du bas vers le haut).

De fait, en septembre 1992, un article du groupe de Daniel Cohen [10] paraissait dans la revue *Cell* et faisait l'effet d'une bombe. Intitulé *Mapping the whole human genome by fingerprinting yeast artificial chromosomes*, émanant du CEPH, de Généthon et de l'INRIA, il étendait l'approche du DOE à l'ensemble du génome hu-

main à partir d'une banque de « méga-YAC » dont les *inserts* approchaient et dépassaient parfois la mégabase. Article fortement médiatisé : le quotidien *Libération* indiquait tout bonnement que cette technique révolutionnaire allait permettre « d'achever d'ici la fin de l'année la cartographie physique des gènes humains » [11]. Ces espoirs devaient être déçus. L'article beaucoup plus modeste paru un an plus tard dans *Nature* [12], et intitulé *A first-generation physical map of the human genome*, s'appuyait en fait sur la carte génétique réalisée à Généthon et utilisait largement les marqueurs déjà placés par l'analyse génétique pour positionner les YAC. L'approche n'avait plus grand-chose à voir avec la méthode annoncée dans l'article de *Cell*. De plus, la carte présentée n'était guère utilisable dans la réalité. Une « vraie » carte physique raisonnablement détaillée et fiable ne devait apparaître que nettement plus tard, notamment grâce aux travaux de l'équipe de Lander (*Whitehead Institute*, Cambridge, États-Unis), fondés sur le repérage des STS (*sequence-tagged sites*) dans les clones mais utilisant toujours les méga-YAC du CEPH [13]. La tentative française avait au moins servi à stimuler les autres laboratoires en leur montrant qu'il était concevable d'établir de manière globale la carte physique du génome humain...

En fait, les YAC avaient trahi leurs utilisateurs par un chimérisme catastrophique : la moitié (peut-être même plus) des clones étaient formés de segments d'ADN provenant de deux régions distinctes du génome. Sans parler des clones instables dont la taille diminuait à chaque analyse, ou de ceux qui ne devenaient stables qu'après avoir subrepticement perdu la moitié de leur *insert*... Finalement, le mérite des YAC fut de prouver qu'il était possible de cloner de grands segments d'ADN, et d'encourager la recherche d'autres systèmes plus fiables : dès 1990, Nat Sternberg dévoilait un vecteur fondé sur le phage P1, qui permettait de propager dans des bactéries des fragments d'au moins 100 kilobases [14]. Aujourd'hui, les cartes physiques « prêtes à séquencer » sont généralement construites à l'aide de BAC

(*bacterial artificial chromosomes*) dont les *inserts* de 200 kilobases sont presque toujours stables et non chimériques.

Finalement, des trois méthodes qui, en 1990, semblaient devoir aboutir aux cartes physiques, seuls subsistent aujourd'hui les BAC, héritiers des YAC. D'autres approches nouvelles, en particulier l'emploi systématique d'hybrides d'irradiation, ont joué un rôle complémentaire important. Si les « cartes intégrées » du génome sont aujourd'hui une réalité, c'est grâce à des méthodes bien différentes de celles sur lesquelles on avait parié (et investi) il y a une dizaine d'années.

La révolution du séquençage se fait attendre

Le séquençage intégral fut, dès le début, le cheval de bataille des promoteurs du Programme Génome Humain. Il semblait pourtant hors de portée des procédés existant à l'époque : technique de Maxam et Gilbert ou de Sanger, mises en œuvre à la main avec des marquages radioactifs ou, avec des réactifs fluorescents, dans les premiers « séquenceurs » développés dans le laboratoire de Lee Hood et produits par la compagnie *Applied Biosystems* dès 1987. Le débit maximal dans les meilleures conditions était estimé à une centaine de kilobases par an et par personne – trop peu pour s'attaquer aux trois milliards de bases de notre génome... Le potentiel d'amélioration de ces approches semblait pour beaucoup très limité***.

On fondait donc beaucoup d'espoirs, notamment aux États-Unis, sur la mise au point de nouvelles méthodes. Il y avait le choix : excision puis détection par cytométrie de flux de bases individuelles (à partir d'une seule molécule d'ADN), lecture directe de la séquence par microscopie à effet tunnel ou à force atomique [15], analyse par spectrométrie de masse, ou encore séquençage par hy-

* Une rumeur invérifiable disait que même pour ce projet son rôle avait été moins important qu'il ne semble à la lecture de l'article.

** On oublie trop souvent le second auteur, le Français Georges Carle.

*** Un document du Department of Energy datant de 1992 (Primer on Molecular Genetics, DOE, juin 1992) les qualifie d'interim sequencing technologies.

bridation. Ces méthodes de déchiffrage « sans gel » devaient augmenter l'efficacité de plusieurs ordres de grandeur, et on considérait généralement que la majeure partie du génome humain serait déchiffrée à l'aide de l'une d'entre elles*. En fait aucune de ces méthodes n'a contribué au séquençage. La première est restée bloquée par des problèmes de sensibilité (il faudrait être capable de détecter à coup sûr une seule molécule de base fluorescente). Les microscopies à champ proche, elles, ont produit beaucoup d'artefacts interprétés de manière optimiste comme de l'ADN jusqu'à ce qu'on se rende compte que la visualisation de cette molécule était effectivement possible dans des conditions opératoires bien précises, mais que la résolution obtenue ne permettait pas de distinguer les bases. La spectrométrie de masse commence à avoir un impact dans certaines méthodes de détection de SNP, mais pas pour le séquençage *a priori*. De même, l'hybridation à des jeux d'oligonucléotides s'avère très performante pour la recherche de mutations (donc la détection de variants par rapport à une séquence connue) mais impraticable pour la détermination de séquences inconnues.

C'est donc la méthode « conventionnelle » de Sanger qui est responsable de la quasi-totalité des trois milliards de nucléotides d'ADN humain maintenant emmagasinés dans EMBL et GenBank. Elle s'est finalement révélée de loin la plus opérationnelle, et a atteint le but requis grâce à de nombreux perfectionnements de détail, à l'automatisation de plusieurs étapes**, à des financements conséquents et à de gros progrès tant dans une organisation de type industriel pour les opérations de mégaséquen-

çage que dans le contrôle, l'assemblage et l'exploitation informatique des données. La surprise, ici, a été l'absence de révolution, l'échec de toutes ces nouvelles approches (je n'ai cité que les principales) dont on pouvait penser qu'au moins une se révélerait efficace, et les potentialités d'une méthode datant de 1977 – autant dire de la préhistoire !

EST : le succès d'un franc-tireur

Le dernier exemple d'inflexion inattendue est bien sûr celui des EST (*expressed sequence tags*). L'idée de s'intéresser préférentiellement aux gènes, donc à l'ARN messager *via* des banques d'ADNc, avait déjà été proposée plusieurs fois, notamment par Sydney Brenner. Plus près de nous, un rapport établi en 1990 par Philippe Kourilsky proposait de centrer le programme français sur le séquençage des ADNc et d'y consacrer au moins 20% du budget***. Aux États-Unis, premier parti dans la course du génome, la doctrine officielle était, cependant, de viser le séquençage intégral de l'ADN génomique humain, et de ne pas se laisser distraire par des projets latéraux. Il faut avouer que les tentatives précédentes de séquençage au hasard d'ADNc n'avaient pas été très fructueuses [16] – sans doute ces projets étaient-ils trop sous-dimensionnés pour être efficaces.

Le génie de Craig Venter fut de définir un mode opératoire à grande échelle adapté aux séquenceurs existants : prendre les clones d'ADNc au hasard dans des banques et se contenter pour chacun d'eux d'une seule réaction de séquence donnant quelques centaines de bases avec un taux d'erreur pouvant approcher 1% [17]. Ce mode opératoire était bien éloigné des habitudes des chercheurs, qui préférèrent choisir amoureusement

quelques ADNc puis les séquencer intégralement avant de les décrire sous toutes les coutures dans un article prévu pour *Nature* – mais qui finit souvent dans une revue bien moins prestigieuse... L'approche de Venter devait démontrer un rapport qualité/prix incomparable. L'approche des EST connut le succès que l'on connaît, attesté aujourd'hui par les presque sept millions de séquences contenues dans dbEST – sans parler des millions que revendiquent *Incyte* ou *Human Genome Sciences*. L'importance scientifique et économique de ces séquences partielles a été amplement démontrée, depuis le clonage *in silico* de gènes responsables de cancers héréditaires [18] jusqu'à la découverte de protéines thérapeutiques comme un facteur de croissance de kératinocytes, KGF-2 [19], rebaptisé « répifermine » par *Human Genome Sciences* et maintenant proche de la mise sur le marché comme accélérateur de cicatrisation. La séquence intégrale du génome humain, aujourd'hui disponible, ne devient interprétable que grâce à la comparaison avec les EST, comme l'expérience des chromosomes 22 et 21 l'a amplement démontré [20].

La morale de l'histoire

Sur le plan technologique (le seul que nous abordions aujourd'hui), le déroulement du Programme Génome a donc été tout sauf un long fleuve tranquille. On s'est beaucoup trompé, bien des certitudes se sont effondrées tandis que des paris osés étaient tantôt gagnés, tantôt perdus****. C'est la règle du jeu : la prévision ne marche vraiment bien qu'après coup ! Il faut en tirer les conséquences : de grands programmes scientifiques comme celui-ci demandent à la fois continuité et souplesse. Continuité : les « coups d'accordéon » financiers qu'a subi le projet français au fil des changements de gouvernement ont très cer-

* Le même document indique « Third-generation gel-less sequencing technologies... are expected to be used for sequencing most of the human genome. »

** Il y aurait aussi beaucoup à dire sur les méandres de la robotique associée aux projets Génome : les échecs ont été nombreux, même au Japon (pourtant a priori bien parti), et encore aujourd'hui l'atelier de séquençage entièrement robotisé (dont la mise en place fut tentée à plusieurs reprises) reste une vue de l'esprit.

*** Ce rapport (« Rapport sur le Génome Humain », élément d'un projet « Génomes » par Philippe Kourilsky, 5 juillet 1990) demandé par le ministre de la Recherche, Hubert Curien, et établi après consultation de nombreux chercheurs, reste très intéressant à lire aujourd'hui. Son auteur avait bien vu les enjeux et esquissé des stratégies (scientifiques et organisationnelles) tout à fait adéquates. Il est fort dommage que ses recommandations n'aient guère été suivies...

**** Pas question de me poser en censeur sentencieux : je me suis trompé au moins aussi souvent que les autres, couvrant d'éloges les « méthodes mégabasse » (gels pulsés, saut et YAC) ou exprimant mon scepticisme sur la possibilité d'obtention de la séquence humaine dans un délai raisonnable...

tainement nuit à l'efficacité de recherches qui demandent à être planifiées sur deux ou trois ans. Mais aussi souplesse: les inflexions technologiques peuvent rendre très vite obsolète un programme ou des installations*, il faut les percevoir à temps et ajuster le tir en conséquence. Et naturellement cette souplesse doit aussi caractériser les structures et les personnes. De ce point de vue, le statut d'EPST, les règles de la comptabilité publique, le code des marchés et aussi la qualité de fonctionnaire ne constituent sans doute pas le cadre le mieux adapté... ■

* La salle des Mark II au Généthon en est un bel exemple. Ces machines automatisant la technique du Southern blot, mises au point (assez lentement) par l'entreprise Bertin, entrèrent en production au moment où la cartographie génétique basculait des RFLP (détectés par Southern) aux microsatellites (analysés sur gel de séquence). Cet « éléphant blanc » (2 MF par machine) fut quelque temps recyclé dans le fingerprinting des YAC avant d'être démantelé.

Bertrand Jordan

Coordinateur de Marseille-Géropole, Parc scientifique et technologique de Luminy, Case 901, 13288 Marseille Cedex 9, France.

RÉFÉRENCES

1. Cook-Deegan R. *The gene wars*. New York : WW Norton and Co, 1994.
2. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980 ; 32 : 314-31.
3. Weissenbach J, Gyapay G, Dib C, et al. A second-generation linkage map of the human genome. *Nature* 1992 ; 359 : 794-801.
4. Jordan BR. Megabase methods : a quantum jump in recombinant DNA techniques. *Bioessays* 1988 ; 8 : 140-5.
5. Schwartz DC, Cantor CR. Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 1984 ; 37 : 67-75.
6. Collins FS, Weissman SM. Directional cloning of DNA fragments at a large distance from an initial probe : a circularization method. *Proc Natl Acad Sci USA* 1984 ; 81 : 6812-6.
7. Collins FS, Drumm ML, Cole JL, Lockwood WK, Vande Woude GF, Iannuzzi MC. Construction of a general human chromosome jumping library, with application to cystic fibrosis. *Science* 1987 ; 235 : 1046-9.
8. Burke DT, Carle GF, Olson MV. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 1987 ; 236 : 806-12.
9. Kohara Y, Akiyama K, Isono K. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* 1987 ; 50 : 495-508.
10. Bellanne-Chantelot C, Lacroix B, Ougen P, et al. Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* 1992 ; 70 : 1059-68.
11. Bensimon C. Percée décisive dans la recherche génétique. *Liberation*, 18 septembre 1992.
12. Cohen D, Chumakov I, Weissenbach J. A first-generation physical map of the human genome. *Nature* 1993 ; 366 : 698-701.
13. Hudson TJ, Stein LD, Gerety SS, et al. An STS-based map of the human genome. *Science* 1995 ; 270 : 1945-54.
14. Sternberg N. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc Natl Acad Sci USA* 1990 ; 87 : 103-7.
15. Jordan BR. Le tunnel séquencera-t-il le génome? *Med Sci* 1990 ; 6 : 1007-8.
16. Putney SD, Herlihy WC, Schimmel P. A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* 1983 ; 302 : 718-21.
17. Adams MD, Kelley JM, Gocayne JD, Olde B, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991 ; 252 : 1651-6.
18. Papadopoulos N, Nicolaidis NC, Wei YF, et al. Mutation of a *mutL* homolog in hereditary colon cancer. *Science* 1994 ; 263 : 1625-29.
19. Jimenez PA, Rampy MA. Keratinocyte growth factor-2 accelerates wound healing in incisional wounds. *J Surg Res* 1999 ; 81 : 238-42.
20. Jordan B. ADNc: les incontournables. *Med Sci* 2001 ; 17 : 81-4.

TIRÉS À PART

B. Jordan.