



par Bertrand JORDAN

*Génome : la caverne d'Ali Baba...  
ou le supplice de Tantale ?*

Nombre de lecteurs de ces *Chroniques génomiques* les parcourent avec attention (du moins je l'espère...), mais sans se sentir personnellement concernés : le monde du génome est loin de leurs préoccupations quotidiennes, et si leur curiosité les porte à vouloir savoir ce qui s'y passe, ils n'y voient pas de conséquence directe pour le déroulement quotidien de leurs travaux. Pourtant les recherches menées dans le cadre des programmes Génome, si elles impliquent souvent de grands laboratoires fortement équipés et des projets importants, produisent beaucoup d'outils et d'objets présentant un intérêt réel pour les équipes « normales ». Le but de cette chronique est de le rappeler à l'aide de quelques exemples, et de donner des indications propres à en faciliter l'accès. Les connaisseurs me pardonneront le manque d'informations nouvelles pour eux dans ce texte qui, à vrai dire, ne leur est pas destiné...

**Les retombées des programmes Génome**

Une équipe venant d'isoler un morceau du gène qui est l'objet de son étude a, en général, pour premier désir d'en disposer dans son intégralité, de préférence sous la forme d'un segment cloné contenant aussi les séquences de régulation qui gouvernent l'expression et qui peuvent parfois se trouver assez loin des séquences codantes. Plutôt que d'explorer à cet effet une banque génomique construite dans des phages ou des cosmides, il peut donc être avantageux d'isoler un chromosome artificiel de levure ou YAC (*yeast artificial chromosome*) : la taille probable du segment inséré (quelques centaines de kilobases) rend très vraisemblable l'obtention du gène complet accompagné des différents promoteurs, *enhancers*, *CAT*

*boxes* et autres séquences pertinentes. Mais, on le sait [1], la construction de banques de YAC reste une entreprise très délicate et de longue haleine : il est hors de question de se lancer dans ce travail pour un besoin ponctuel. La réaction habituelle en ces cas pour un biologiste moléculaire est de demander un exemplaire de la banque au laboratoire qui l'a construite ; mais une banque YAC ne se transfère pas comme une librairie de phages ou de cosmides, il ne suffit pas d'envoyer un tube contenant quelques centaines de milliers de clones que le laboratoire destinataire pourra utiliser à sa guise. Les banques de chromosomes artificiels de levure se présentent sous forme d'un jeu de clones ordonnés dans des plaques à microtitration : une librairie représentative du génome humain (50 000 clones) occupe ainsi 500 plaques à 96 puits. La duplication et le transport d'un ensemble aussi volumineux ne sont évidemment concevables que dans un nombre très limité de situations\*. Le chercheur pourrait, à défaut, envoyer sa sonde à un laboratoire central qui se chargerait du criblage : c'est ce qu'a fait le groupe de David Schlessinger (Saint-Louis, MO, USA) au début, mais il a vite croulé sous le nombre des sondes et reculé devant l'irresponsabilité des utilisateurs.

Depuis un ou deux ans, l'accès à ces banques a été organisé, grâce en grande partie aux financements des programmes Génome. La modalité la plus courante consiste en un criblage par PCR à deux étapes. Dans un

premier temps, le laboratoire demandeur séquence sa sonde et définit des oligonucléotides autorisant l'amplification d'un fragment spécifique. Il reçoit, des détenteurs de la banque, des échantillons d'ADN correspondant à des *pools*, c'est-à-dire contenant chacun l'ADN extrait d'un mélange de mille ou deux mille clones YAC, la banque entière étant ainsi représentée par quelques dizaines d'échantillons. Il effectue alors les réactions PCR sur chacun des *pools* afin d'identifier celui qui répond positivement à l'amplification (en donnant le fragment de la taille prévue) et qui doit donc contenir le clone positif recherché. Il envoie alors au laboratoire central cette information ainsi que les oligonucléotides permettant l'amplification afin que ce dernier effectue la deuxième étape du criblage. Cette procédure partage le travail et y implique le demandeur ; et la preuve est maintenant faite qu'elle constitue une façon viable de permettre à la communauté l'accès à une banque YAC. C'est en effet sous cette forme que sont exploitées plusieurs d'entre elles, en France dans le cadre du CEPH ou en Grande-Bretagne au *Resource Centre du Human Gene Mapping Programme* ; les financements nécessaires à la mise en place d'une telle organisation ont été fournis pour l'essentiel par des contrats « Génome » nationaux ou européens, et mettent — moyennant un certain travail ! — cet outil remarquable à la disposition de l'ensemble des laboratoires.

Comme je l'ai déjà mentionné, d'autres modalités sont envisageables, entre autres l'élégante solution des « librairies de référence », défendue par Hans Lehrach (*Imperial Cancer Research Fund*, Londres). Une banque de cosmides du chromosome X a

\* On peut à cet égard signaler qu'il est possible d'acheter une banque YAC humaine (diffusée par la firme Clontech) ; mais le prix de cette librairie sous sa forme développée (340 plaques à microtitration contenant chacune 96 clones) est en France de près de 100 000 F.

ainsi été exploitée par de multiples équipes à l'aide de filtres à haute densité distribués par ce groupe ; pour les YAC, le procédé n'est pas encore complètement opérationnel (en tant que système de distribution), ce qui est hommage car le criblage direct par hybridation présente beaucoup d'avantages pour l'utilisateur, surtout si ce dernier recherche de nombreux clones... Mais, en tout état de cause, l'accès aux clones YAC est maintenant possible à tout laboratoire — à condition qu'il sache où s'adresser et qu'il fasse preuve d'une certaine patience. Rappelons que les techniques d'analyse des YAC sont maintenant, pour la plupart, bien codifiées, à la portée de tout laboratoire de biologie moléculaire normalement constitué et que ces clones peuvent aussi être transférés dans des cellules et, sans doute bientôt, dans des animaux. Notons aussi que les YAC se prêtent tout à fait à la localisation chromosomique par hybridation *in situ* après marquage non radioactif : une simple amplification sur l'ADN total de la cellule de levure contenant le clone avec des amorces Alu (spécifiques de l'ADN humain) produit un mélange de sondes directement utilisable.

Autre cas de figure : une équipe vient d'isoler, chez la souris, une série de gènes apparentés par des homologies de séquence, qui pourrait représenter une nouvelle famille de récepteurs de neurotransmetteurs. Compte tenu des précédents, il est concevable que ces gènes soient regroupés dans le génome, ce qui pourrait indiquer qu'ils sont issus d'une duplication récente ou qu'ils sont soumis à une régulation commune. Il est donc essentiel de savoir très vite — et à peu de frais si possible... — si cette hypothèse est tenable. L'hybridation *in situ* n'est pas la solution car les sondes dont on dispose (des ADNc de petite taille) ne donnent pas de résultats suffisamment rapides et fiables par cette technique. Mais il existe, dans le cas de la souris, une alternative : les systèmes de croisement interspécifiques qui, grâce à la divergence de séquence entre les ADN des deux souches, autorisent la localisation de toute sonde qui se révèle polymorphique dans ces con-

ditions... pour autant qu'un travail préalable ait permis d'effectuer les croisements, de préparer les ADN des centaines de souris qui en résultent, et de repérer les points de recombinaison sur chacun des chromosomes à l'aide d'une étude avec un jeu de sondes de référence. Une entreprise de cette nature est en voie d'achèvement dans le cadre du *Backcross* européen qui associe les équipes de Jean-Louis Guenet à l'Institut Pasteur et le *HGMP Resource Centre* britannique ; et cela devrait aboutir, avant la fin de cette année, à un jeu de filtres sur lesquels il suffira d'hybrider une sonde quelconque (sous réserve qu'elle révèle un fragment différent sur les ADN des deux parents) pour déterminer son chromosome d'origine et sa position à une dizaine de centimorgans près. Il sera ainsi possible, pour reprendre l'exemple discuté, de savoir rapidement si les gènes en cause proviennent d'une même région chromosomique... sans se livrer à de nombreuses et délicates hybridations *in situ* ni se lancer dans une longue série d'analyses en champ pulsé.

A ce stade, d'ailleurs, peut apparaître un troisième type d'utilisation des « ressources » produites par les programmes Génome. Il s'agit cette fois de ces banques de données de plus en plus complètes qui emmagasinent et rendent accessibles les informations obtenues par les laboratoires spécialisés [2]. L'une d'elles, GBASE, est principalement consacrée à la souris et contient un état à jour des correspondances (synténies) connues entre les chromosomes humains et murins. On sait que notre génome est étonnamment proche de celui de ce rongeur familier, que pratiquement tout gène humain existe sous une forme très proche chez la souris ; mais on constate aussi que la disposition de ces gènes le long des chromosomes est conservée sur des distances importantes. Le chromosome 11 de la souris, par exemple, contient une grande zone homologue au 17 humain, une autre correspondant au 5. Dans chacune de ces régions, les gènes sont dans le même ordre chez les deux espèces. Ces synténies sont maintenant bien explorées ; de sorte que, de la position d'un gène

chez la souris, on peut déduire presque à coup sûr sa localisation chez l'homme, surtout si l'on dispose des données les plus récentes. L'exploitation décrite ci-dessus du système de « *backcross* interspécifique » permettra donc d'obtenir non seulement l'information chez la souris, mais aussi une position très vraisemblable chez l'homme. On pourra alors se demander si des syndromes héréditaires pouvant être dus à un dysfonctionnement des récepteurs en question n'auraient pas par hasard été « localisés » (au sens de la localisation génétique, première phase de la « génétique inverse ») dans la région chromosomique humaine ainsi désignée. La consultation des données génétiques et cliniques contenues dans OMIM (*On-line Mendelian Inheritance in Man*, la version électronique du fameux atlas des maladies génétiques humaines de Victor McKusick) ou dans la base française Genatlas sera à même de répondre à cette interrogation... et peut-être d'indiquer ainsi une voie de recherche très féconde !

#### De la coupe aux lèvres...

C'est maintenant que l'on en vient au supplice de Tantale : les choses, en effet, ne sont pas aussi aisées. Pour employer les « ressources » (au sens anglo-saxon du mot) décrites ci-dessus, il faut déjà connaître leur existence : or l'information ne circule pas toujours bien et elle n'est pas parfaitement fiable. Certains services restent parfois confidentiels, et seuls les initiés peuvent, du coup, en bénéficier ; d'autres entreprises d'intérêt commun, au contraire, très largement annoncées (et quelquefois financées dans ce but par les programmes Génome nationaux ou européens), s'avèrent ne pas fonctionner concrètement. Il faut reconnaître que la technologie évolue très rapidement, mais aussi que la mise en place d'une telle organisation se révèle en général nettement plus difficile que prévu... Ce qui fait que les délais envisagés sont rarement tenus. De plus, la tension entre recherche et service, discutée dans une précédente chronique [3], joue à plein ici ; elle aboutit parfois à des situations confuses où le scientifique « extérieur » ne sait pas ce à quoi il a droit et ce

qui relève de la collaboration à négocier (avec signature d'articles à la clef...). Cela traduit un souci compréhensible de la part d'équipes qui doivent assurer à leurs membres les éléments d'une carrière ; mais on ne peut que souhaiter que les contrats « Génome » nationaux ou européens spécifient clairement les conditions dans lesquelles l'activité de service s'effectue, et que le cahier des charges soit net, précis... et public, même s'il doit être fréquemment révisé.

Tout cela fait que de nombreux laboratoires (surtout ceux qui ne sont pas directement dans le « circuit » Génome) ne profitent pas de ces ressources considérables, alors même qu'elles pourraient considérablement accélérer leur travail.

### Quelques pistes...

Nous resterons donc résolument pratiques en indiquant pour finir quelques pistes pour ceux qui voudraient, eux aussi, faire usage de ces possibilités. L'information est, on l'a compris, une denrée précieuse. Elle est abondante, mais inégalement diffusée, et n'atteint pas obligatoirement ses destinataires.

Une première mesure très simple consiste à faire en sorte de recevoir quelques petits journaux qui rassemblent une foule d'informations utiles sur les programmes Génome et sur les services qu'ils proposent. Il s'agit notamment de *Human Genome News*, le journal du programme Génome des États-Unis (HGMS, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, TN 37831-6050, USA) et de *G-ome News*, son homologue britannique (*HGMP Resource Centre*, CRC, Watford Rd, Harrow, HA1 3UJ, UK). L'abonnement à ces deux gazettes sans prétention est gratuit, et elles renferment toutes sortes de renseignements : noms (et numéros de télécopieur !) des responsables, ressources offertes, nouvelles des « Genome Centers », annonce de colloques spécialisés...

Plus évoluées, plus complètes, mais aussi plus techniques sont les bases de données (voir pour plus de précisions le numéro « spécial séquences » de la revue *Nucleic Acids Research* [4]). Les principales, en ce qui nous concerne, sont GDB (Genome Data Base) et

Genatlas pour les données génomiques hors séquence, GBASE pour la souris et les synténies, et OMIM pour les maladies génétiques. Compte tenu de la masse d'informations produites par la recherche génomique, dont la plupart ne seront jamais publiées au sens classique du mot, elles se révèlent de plus en plus indispensables. Le problème est qu'il ne s'agit pas de bases de données « simples » comme celles concernant les séquences d'ADN, qui peuvent être distribuées sur disquette ou sur disque optique. Ces dernières ont en effet une structure peu complexe (malgré le volume de leur contenu) car elles contiennent essentiellement des objets de même espèce (les séquences), pourvus de quelques annotations et exploités par un système informatique qui n'a pas à gérer de relations compliquées entre ces éléments. Au contraire, une base génomique comme Génome Data Base (ou Genatlas) renferme des données de nature très diverse (clones, sondes, polymorphismes, maladies, cartes génétiques, cartes physiques...) dont les interrelations sont multiples. Il faut donc les emmagasiner dans un système informatique à structure relationnelle, qui gère les relations entre les objets autant que les objets eux-mêmes. De tels systèmes ne sont pas « portables » sur le micro-ordinateur de chaque chercheur : il faudrait importer non seulement les informations, mais aussi le programme de gestion qui les manipule et ne « tourne » que sur station de travail de type Sun. Bref, dans la pratique, il faut y accéder par voie de réseau, et de réseau à haute vitesse (au moins 9 600 bauds) sinon elles sont quasiment inexploitable.

Comment faire en pratique ? Il est facile d'être enregistré comme utilisateur par ces organisations et de disposer d'un mot de passe : il suffit pour cela de remplir un formulaire (l'adresse de GDB et d'OMIM est donnée dans la référence [4], celle de GBASE est : The Jackson Laboratory, Bar Harbor, Maine 04609, USA, et Genatlas est bien connu des lecteurs de *médecine/sciences*, ne serait-ce que par les cartes qui paraissent dans les dernières pages de chaque numéro). Mais, pour employer réel-

lement ces banques, il faut pouvoir se raccorder à un réseau à grande vitesse (Internet, de préférence par une liaison par fibre optique), disposer d'un terminal (un Macintosh ou un PC peuvent suffire, mais seule une station de travail pourra exploiter pleinement l'interface graphique), et savoir s'y retrouver dans ces ensembles où la navigation est parfois complexe. Là encore, il a été créé des structures destinées à faciliter l'entrée dans ces systèmes informatiques, à diminuer les coûts de connexion et à proposer des formations spécialisées aux utilisateurs : le HGMP britannique offre un tel service et, dans le cadre d'un contrat européen, le DKFZ de Heidelberg fait de même (*European Data Resource for Human Genome Analysis*, DKFZ, Im Neuenheimer Feld 280, D 6900 Heidelberg, Allemagne). On peut souhaiter que, dans notre pays aussi, des relais de ce genre soient établis, au niveau national ou régional ; en attendant l'accès à ces données est possible mais exige une bonne dose de patience et un environnement informatique solide. C'est d'ailleurs un domaine dans lequel les organismes de recherche devraient faire un sérieux effort, car il en va de notre compétitivité...

Sur un plan plus matériel, il faut citer les nombreuses banques de cellules, de clones ou de bibliothèques dont les services sont soit gratuits, soit peu onéreux. Tout le monde connaît l'ATCC (*American Type Culture Collection*), mais certains ignorent qu'elle distribue aussi des banques spécifiques de chromosomes ou d'ADNc, y compris une des premières banques « égalisées », celle qui émane du laboratoire de Sherman Weissman à Yale (New Haven, USA). Une autre structure, le *Corriell Cell Repository* (401 Haddon Ave, Camden, NJ 08103, USA), diffuse un précieux jeu d'hybrides monochromosomiques : vingt-quatre lignées de cellules de hamster contenant chacune un seul chromosome humain. Plus près de nous, des banques de sondes assez complètes sont gérées par le *Resource Centre* du HGMP britannique ; notons au passage que ce centre (dont une partie du financement est européenne) est ouvert à tout chercheur de la CEE. En France, le

CEPH et surtout le Genethon (1, rue de l'Internationale, BP 59, 91002 Evry Cedex, France. Tél. : 1.69.47.28.00) assurent eux aussi certains services : criblage de la banque de YAC bien sûr, mais aussi établissement de lignées lymphoblastoïdes, extractions d'ADN et préparation de *Southern blots*. Les prestations ne sont pas gratuites, mais les prix sont modérés et très inférieurs à ceux du secteur commercial : moins de 300 F pour établir une lignée lymphoblastoïde, de 300 à 400 F pour un *Southern blot* de 18 échantillons, y compris les digestions enzymatiques. C'est donc une aide précieuse, d'autant que ces montants peuvent être inclus dans une demande de financement déposée auprès de l'Association française contre les myopathies...

#### En guise d'envoi

Cette chronique résolument concrète aura peut-être déçu certains de mes lecteurs habituels parmi les spécialistes : ils n'y auront pas appris grand-chose, ils auront peut-être même relevé quelques approximations. Mais si elle a pu convaincre d'autres chercheurs d'aller voir de plus près en quoi les ressources des programmes Génome pourraient leur être profitables, si elle leur a donné quelques pistes pour commencer à y accéder... elle aura rempli son objectif ! ■

---

#### Bertrand Jordan

*Directeur de recherche au Cnrs, responsable du groupe de génétique moléculaire humaine. CIML, Inserm/Cnrs, case 906, 13288 Marseille Cedex 9, France.*

---

#### RÉFÉRENCES

1. Jordan BR. La montée en puissance des YAC. *médecine/sciences* 1990 ; 6 : 470-2.
2. Jordan BR. Génome et informatique : condamnés à s'entendre ? *médecine/sciences* 1991 ; 7 : 726-8.
3. Jordan BR. Les contradictions du génome. *médecine/sciences* 1992 ; 8 : 476-82.
4. Sequence supplement. *Nucleic Acids Res* 1991 ; 19 : 2221-49.