

par Bertrand JORDAN

*Les yeux plus gros
que le ventre ?*

L'indispensable carte génétique.

Arriver à un centimorgan, ce n'est pas évident : obstacles organisationnels, financiers et techniques.

Un espoir, les microsatellites.

Le typage de spermatozoïdes aura-t-il un impact ?

L'automatisation, toujours l'automatisation. Un optimisme mesuré...

Dans le cadre du « projet génome humain », on parle plus souvent de séquence ou de carte physique que de carte génétique. Pourtant cette dernière est un élément indispensable du projet, et les difficultés récemment rencontrées dans son établissement sont la cause de sérieuses inquiétudes [1, 2]. La carte génétique représente en effet le premier balisage du génome, qui sert de cadre et de point de départ aux études plus précises mais aussi beaucoup plus laborieuses que sont la cartographie physique et son aboutissement ultime, la séquence. La première carte générale du génome humain, publiée en 1987 par l'équipe de Helen Donis-Keller [3], qui travaillait à l'époque pour l'entreprise Collaborative Research aux États-Unis, comportait des jalons espacés en moyenne de dix centimorgans (dix pour cent de recombinaison). Ces jalons sont en fait des sondes d'ADN qui correspondent chacune à un point du génome présentant un polymorphisme de séquence : cela se traduit par l'apparition ou la disparition

de sites de coupure pour certaines enzymes de restriction, de sorte que la sonde va révéler des fragments de restriction différents chez différents individus, d'où le terme de RFLP (*restriction fragment length polymorphism*). Pour établir la carte génétique, il faut donc isoler un nombre suffisant de sondes (plusieurs centaines) pour étudier la façon dont les polymorphismes correspondants sont transmis au cours des générations. Avec beaucoup de sondes, et de grandes familles très bien caractérisées (d'où le rôle primordial du centre d'étude du polymorphisme humain qui a constitué et mis à la disposition de la collectivité une telle collection de familles, voir article de Jean Dausset, *m/s* n° 3, vol. 6, p. 286), on arrivera ainsi à siter les uns par rapport aux autres les *loci* révélés par ces sondes et à évaluer leurs distances en termes de fréquence de recombinaison, un centimorgan correspondant par définition à une fréquence de recombinaison de un pour cent.

Le groupe de Helen Donis-Keller avait donc posé un jalon en moyenne

tous les 10 centimorgans. Cela représentait une étape importante, mais le balisage était encore trop lâche par rapport aux possibilités des techniques capables de prendre le relais. Prenons par exemple le cas de la recherche du gène responsable d'une maladie par l'approche dite de « génétique inverse » : l'étude génétique de familles dans lesquelles cette maladie est présente permet — grâce justement à l'existence de la carte génétique et de la collection de sondes mentionnée plus haut — d'associer, de lier une ou deux sondes et la maladie : cela permet d'affirmer que le gène responsable (lorsqu'il est défectueux) de la maladie doit se trouver quelque part entre la sonde A et la sonde B. Il ne reste plus alors — comme cela a été fait avec succès pour le gène responsable de la mucoviscidose — qu'à partir de A ou de B (ou des deux), qu'à cloner et analyser l'ADN situé dans l'intervalle, y rechercher des gènes et tester chacun d'eux pour voir s'il est en cause... jusqu'à l'identification du gène recherché. Le problème est

qu'en moyenne une distance génétique d'1 centimorgan correspond à une distance le long de l'ADN de 1 000 kilobases, un million de nucléotides ; si les deux jalons sont distants de 10 centimorgans c'est donc une zone de 10 000 kilobases d'ADN qu'il faudra cloner, disséquer, analyser, ce qui reste encore impossible ou du moins extrêmement laborieux même avec les méthodes d'analyse de l'ADN les plus modernes [4].

Il n'est donc pas étonnant que l'obtention d'une carte génétique à 1 centimorgan (c'est-à-dire dans laquelle les jalons ne soient distants en moyenne que d'1 centimorgan, environ 1 000 kilobases) ait été affichée comme une priorité. Aux États-Unis, le *National Research Council* (NRC) avait fixé à cinq ans la période nécessaire à l'obtention d'une telle carte, et ce but comme ce délai avaient été réaffirmés par James D. Watson lorsqu'il prit la tête du projet génome du *National Institutes of Health* (NIH) courant 1988. Mais de très sérieux doutes ont surgi sur la possibilité d'arriver effectivement à un tel résultat dans un délai raisonnable.

Pourquoi ces problèmes ? Il faut d'abord comprendre qu'à technologie constante, la construction d'une carte à 1 cM représente environ cent fois plus de travail qu'une carte à 10 cM (il faut étudier dix fois plus d'individus avec dix fois plus de sondes). La quantité de travail que cela implique est donc colossale... et ce travail est particulièrement ingrat : il consiste pour l'essentiel à digérer l'ADN de centaines d'échantillons (isolé de cultures de cellules établies à partir des individus des familles étudiées) par différentes enzymes de restriction ; à séparer les fragments d'ADN par électrophorèse et à les transférer sur des membranes ; à marquer les sondes et à les hybrider l'une après l'autre sur ces membranes ; à examiner, collationner et interpréter les résultats après autoradiographie des membranes... Travail répétitif et ingrat s'il en est, surtout lorsqu'il doit être effectué avec des centaines de sondes sur des centaines d'échantillons. Travail qui, dans l'organisation actuelle de cette recherche et même aux États-Unis, est effectué

pour l'essentiel manuellement par des étudiants, des thésards ou de jeunes post-doctorants : on comprend qu'ils ne résistent pas très longtemps, d'autant plus que pour être répétitif et très codifié, ce travail n'en réclame pas moins une très grande attention si l'on veut que ses résultats soient fiables. Dans ces conditions, il est difficile d'arriver à poursuivre cette cartographie systématique et — apparemment — routinière, et la tentation est grande, par exemple, de s'intéresser plutôt à une région particulière dans laquelle on sait que doit se trouver un gène impliqué dans une maladie et de concentrer ses efforts sur cette région — travail louable bien sûr, mais pendant ce temps-là la carte générale n'avance plus. C'est un peu comme si des cartographes chargés de faire un relevé général de la France ne s'intéressaient qu'aux sites où l'on pense pouvoir trouver du pétrole !

L'analogie est loin d'être gratuite, et elle pose le problème de la difficulté à effectuer durablement un travail systématique et parfois ingrat dans un domaine caractérisé par l'individualisme des chercheurs, la course au résultat publiable (et médiatisable) et les jugements de valeur sur la base de « modes » parfois contestables. On peut penser, par exemple, que l'énorme effort investi par une ou deux dizaines de groupes en compétition sérieuse sur l'identification du gène impliqué dans la mucoviscidose aurait, investi différemment, pu permettre la cartographie complète d'un chromosome ou d'une partie significative d'un chromosome : c'est justement l'une des justifications d'un projet global et systématique comme le projet Génome humain. Mais comment motiver durablement des équipes à s'investir dans ce travail sans l'abandonner pour se précipiter sur le premier filon qui apparaît ? C'est un problème important et qui se reposera, avec plus d'acuité, pour les travaux plus lourds comme le séquençage systématique.

La motivation des chercheurs n'est pas la seule denrée rare dans ce domaine : il semble que les moyens aient aussi fait défaut. Aux États-Unis, même les organismes comme le NIH qui ont affiché dans leurs priorités l'établissement de cette carte à

1 cM ont semblé, à en croire Helen Donis-Keller [1], peu motivés pour financer ce travail systématique : plusieurs de ses demandes de contrat ont été rejetées par les comités du NIH pour « manque d'originalité », et cette équipe travaille maintenant à *Washington University* (St.Louis, USA) avec principalement des fonds privés. Comme elle le dit elle-même : « Je n'ai jamais prétendu que ce travail soit innovant. Mais il est important et faisable. »

Un dernier obstacle, qui était connu au départ mais n'en est pas moins sérieux, est celui du manque d'informativité des sondes. Pour être vraiment utile dans ce type de travail, une sonde ne doit pas seulement être polymorphique, elle doit être extrêmement polymorphique. C'est une question de degré, en effet : à la limite, toute sonde prise au hasard peut être considérée comme polymorphique puisque, si l'on étudie un nombre suffisant d'individus, on finira toujours par en trouver un qui présente une différence de séquence changeant la taille d'un fragment de restriction. Mais évidemment un polymorphisme dans lequel un allèle est présent chez 99,9 % des individus et l'autre chez 0,1 % ne présente guère d'intérêt pour l'étude génétique, puisque la sonde en question ne sera pratiquement jamais informative dans les familles étudiées. Les polymorphismes pratiquement utilisables et le plus souvent utilisés pour l'établissement de la carte génétique ont en général deux ou, mieux, plusieurs allèles, avec des fréquences si possible du même ordre de grandeur. Mais il arrive néanmoins souvent qu'un tel RFLP ne soit pas informatif dans une famille donnée, c'est-à-dire par exemple que les deux parents soient homozygotes pour l'allèle « a » présent dans 40 % de la population générale : le RFLP en question est alors inutilisable dans cette famille..

Sur ce point, un nouveau type de polymorphisme est en passe d'améliorer considérablement la situation et sans doute de permettre une reprise de la progression vers la carte génétique à 1 cM. Il s'agit de ce que l'on appelle les « microsattellites » (*figure 1*) : ce sont des séquences particulières constituées par la répétition d'un

motif très simple, par exemple, pour la plus répandue chez l'homme, la séquence $(GT)_n$ où n peut aller de 5 à 50. L'intérêt de ces microsatellites est qu'ils se révèlent extrêmement polymorphiques et que si, en un point donné du génome existe une telle séquence, elle prendra la forme $(GT)_{17}$ chez un individu, $(GT)_{18}$ ou $(GT)_{15}$ chez un autre et ainsi de suite. Or ces séquences sont extrêmement répandues dans le génome, puisque l'on estime qu'il en existe une en moyenne tous les 10 kilobases : elles fournissent donc des jalons parfaits pour une carte génétique très fine. Bien sûr il y a un revers à cette médaille : la détec-

tion des polymorphismes de microsatellites fait appel à une technologie plus sophistiquée que le simple *Southern blot*. Il faut d'abord trouver un microsatellite dans la zone où l'on recherche un marqueur, par exemple en criblant des cosmides de cette région avec un oligonucléotide $(GT)_{10}$ (on est quasiment assuré d'en trouver vu leur espacement moyen). Après avoir identifié un fragment d'ADN contenant le microsatellite, il faut le séquencer pour définir des zones d'ADN de séquence unique de part et d'autre ; c'est à partir de cette information que l'on définira des amorces utilisables en PCR (*polymerase chain reaction*) et

ciblant spécifiquement ce point du génome. L'analyse de l'ADN d'une série d'individus (*figure 1*) comportera alors l'amplification par PCR sur chaque échantillon à l'aide de ces amorces, puis l'analyse du produit d'amplification sur un gel d'acrylamide très résolutif du type gel de séquence puisqu'il faut pouvoir distinguer des variations de longueur de deux nucléotides sur une longueur totale (selon les amorces choisies) de 100 à 200. L'analyse est donc un peu compliquée (sans doute automatisable en bonne partie), mais la puissance du système est telle, car ces marqueurs sont presque toujours informatifs et très fréquents, que le jeu en vaut largement la chandelle. On peut donc compter sur les microsatellites, et sur l'ingéniosité des chercheurs qui vont sans doute trouver des façons supplémentaires de les utiliser, pour résoudre une bonne partie des problèmes évoqués plus haut.

Une autre innovation dont l'impact sur l'avancement de la carte génétique est moins évident est constitué par l'analyse de l'ADN de spermatozoïdes isolés (*sperm typing*). Il s'agit là encore d'une application de la PCR qui permet de déterminer, à partir de l'ADN d'un seul spermatozoïde isolé par micromanipulation, quel allèle d'un gène donné y est présent (à condition évidemment que la séquence de ce gène soit connue, que des amorces PCR aient été synthétisées et qu'il y ait, dans la zone définie par ces amorces, un polymorphisme informatif dans le cas étudié). La faisabilité d'une telle approche avait été démontrée dès 1988 [7] et elle a déjà fait l'objet d'applications à la mesure de distances génétiques [8]. Les points forts de cette méthode sont assez évidents : d'une part, le spermatozoïde est haploïde, ce qui simplifie largement l'analyse puisque l'on n'a pas à se demander si deux allèles particuliers des marqueurs étudiés sont situés sur le même chromosome (problème de « phase ») ; d'autre part et surtout, chaque spermatozoïde contient des chromosomes qui sont le produit de recombinaisons indépendantes de celles qui ont eu lieu pour les chromosomes d'un autre spermatozoïde : analyser cent spermatozoïdes revient donc à analyser cent enfants du

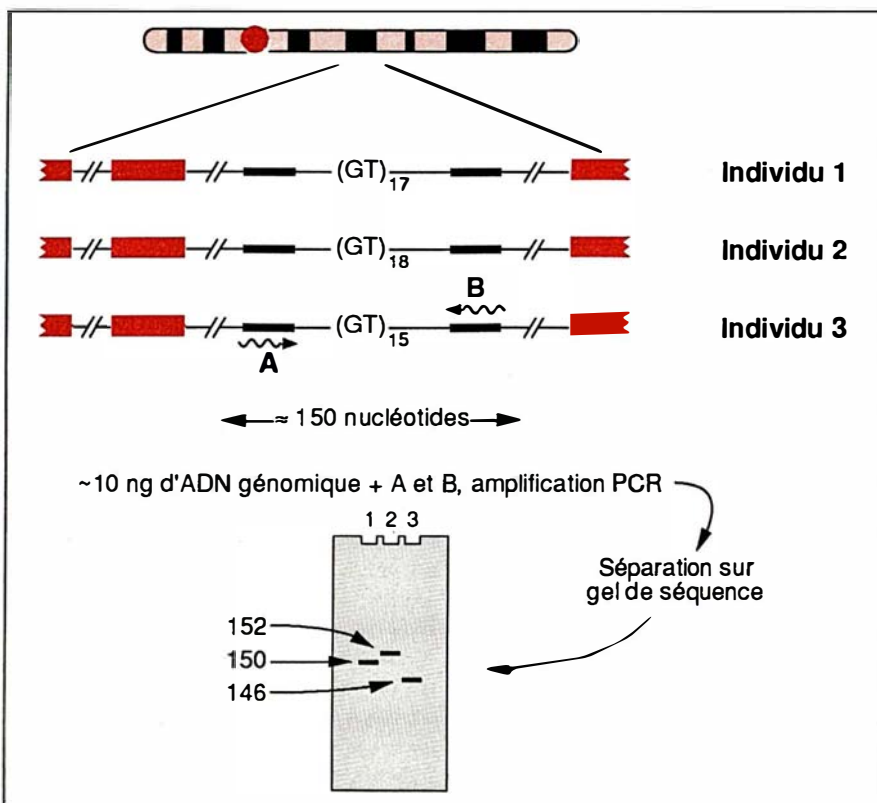


Figure 1. **Les microsatellites et leur analyse.** Les microsatellites, très fréquents dans le génome, sont constitués par des séquences simples souvent du type $(GT)_n$. La figure représente un tel microsatellite entre deux gènes ou entre deux exons d'un gène (rectangles rouges à gauche et à droite). D'un individu à l'autre, le nombre de répétitions est très variable (individus 1, 2 et 3). Si l'on a déterminé les séquences adjacentes (ADN copie unique), on peut définir deux sondes A et B permettant l'amplification sélective par PCR de la zone contenant le microsatellite ; l'analyse des produits d'amplification sur un gel de polyacrylamide très résolutif révèle le polymorphisme lié à la variation du nombre de répétitions et permet de l'utiliser pour une analyse génétique. Le résultat observé est en général plus complexe que dans ce schéma en raison de l'existence de deux loci (sur les deux chromosomes homologues) et de certains artefacts liés à la PCR.

même père — situation privilégiée (du point de vue du généticien !) qui est rarement rencontrée dans la réalité... Toute plaisanterie mise à part, cette technologie donne un avantage considérable pour l'analyse de petites distances génétiques, au prix d'une certaine difficulté technique (micromanipulation, amplification à partir d'une très petite quantité d'ADN). Il est possible de typer simultanément deux *loci* [8], peut-être plus ; en revanche, il est peu probable que, par manque de matériel, l'on puisse ainsi typer simultanément un grand nombre de *loci* sur le même spermatozoïde (par nature unique et différent de chaque autre spermatozoïde), ce qui fait que la contribution de cette méthode à la carte génétique générale restera sans doute limitée. C'est néanmoins une technique à ne pas négliger...

Un troisième espoir pour la carte génétique réside tout simplement dans l'automatisation des procédures très répétitives employées, par exem-

ple le *Southern blot*. Cette automatisation ne va pas sans problèmes (voir la *Chronique génomique du mois d'avril*) mais elle progresse : on devrait ainsi pouvoir effectuer un plus grand nombre d'analyses dans de bonnes conditions de fiabilité, même si les problèmes de coût restent sérieux, comme pour le séquençage (nous en parlerons une autre fois).

Après une certaine vague de pessimisme exprimé à de nombreuses occasions [1, 2], un optimisme mesuré semble donc prévaloir, et une carte génétique à 5, 2 ou peut-être même 1 centimorgan paraît de nouveau être un objectif réalisable dans un avenir relativement proche ■

Bertrand Jordan

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

RÉFÉRENCES

1. Roberts L. Whatever happened to the genetic map? *Science* 1990 ; 247 : 281-2.
2. Anderson GC, Creation of linkage map falters, posing delay for genome project. *The Scientist* 1990 ; 4 : 1-13.
3. Donis-Keller H, et al. A genetic linkage map of the human genome. *Cell* 1987 ; 51 : 319-37.
4. Jordan BR. Megabase methods, a quantum jump in recombinant DNA techniques. *Bioessays* 1988 ; 8 : 140-5.
5. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucl Acids Res* 1989 ; 17 : 6463-71.
6. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 1989 ; 44 : 388-96.
7. Li H, Gyllensten UB, et al. Amplification and analysis of DNA sequences in single human sperm and diploid cells. *Nature* 1988 ; 335 : 442-4.
8. Cui X, Li H, et al. Single sperm typing : determination of genetic distance between the G-globin and parathyroid hormone *loci* by using the polymerase chain reaction and allele-specific oligomers. *Proc Natl Acad Sci USA* 1989 ; 86 : 9389-93.