

3 - Cartographie des génomes

E.P. CRIBIU, L. SCHIBLER

INRA, Laboratoire de Génétique Biochimique
et de Cytogénétique, 78352 Jouy-en-Josas cedex

e-mail : Edmond-Paul.Cribiu@biotec.jouy.inra.fr

Introduction à la cartographie des génomes complexes

Résumé. Les technologies mises en place dans le cadre des projets sur les génomes humain et murin ont stimulé le développement de la cartographie des espèces d'intérêt agronomique. Des cartes ont été construites pour les ruminants, le porc, le cheval et le poulet, permettant de localiser des gènes d'intérêt agronomique ou des QTL (quantitative trait loci) chez ces espèces. Ces travaux de cartographie génétique, cytogénétique et physique devraient se révéler être des outils indispensables pour le clonage positionnel de ces gènes ou QTL.

Les activités de cartographie sont classiquement subdivisées en cartographie génétique et cartographie physique. La cartographie génétique, basée sur l'analyse de ségrégation, consiste à positionner de façon relative des locus, avec comme objectif la mise en évidence de liaisons avec des caractères d'intérêt. La cartographie physique, quant à elle, cherche à déterminer la distance réelle en nucléotides entre locus et a pour objectif ultime le séquençage complet du génome. La distance physique entre deux marqueurs constitue donc une mesure absolue, contrairement à la mesure relative des cartes génétiques. Le lien entre ces deux approches complémentaires est assuré par la cartographie chromosomique, dont l'objectif est l'assignation des éléments de cartographie génétique et physique à des chromosomes ou à des régions de chromosomes. De plus, la cartographie chromosomique permet d'établir les cartes comparées entre les différentes espèces d'intérêt agronomique, l'Homme et la souris.

1 / La carte génétique

Les premières cartes génétiques ont été construites au début du siècle chez la drosophile par Morgan. Ces cartes se basaient essentiellement sur la ségrégation de caractères phénotypiques disposés linéairement sur un chromosome et de l'existence de crossing-over se produisant a priori aléatoirement le long de celui-ci. Ainsi, plus la distance séparant deux gènes portés par le même chromosome est importante, plus la probabilité de recombinaison entre ces deux gènes est élevée. La mesure du taux de recombinaison reflète donc la distance linéaire entre les gènes. Le centimorgan (cM), correspondant à 1 % de recombinaison, fut adopté comme unité de distance génétique. L'analyse de liaison, en suivant la transmission de caractères polymorphes, permet ainsi d'établir la position relative des gènes sur le chromosome et de mesurer la distance qui les sépare.

Chez l'Homme et les animaux d'élevage, la longueur de l'intervalle de génération et l'impossibilité de réaliser des croisements contrôlés entre individus ayant un phénotype donné (que ce soit pour d'évidentes considérations éthiques chez l'Homme ou des raisons d'ordre économique chez les animaux) ont longtemps limité le développement des cartes génétiques. Chez les animaux d'élevage, les premiers résultats obtenus concernaient les groupes sanguins et les protéines du sang ou du lait.

La cartographie complète du génome n'a pu être réalisée qu'avec l'apport des techniques de génétique moléculaire qui ont permis d'exploiter une partie de la variabilité des génomes en augmentant considérablement le nombre de locus potentiellement polymorphes. En effet, par suite de la dégénérescence du code génétique (plusieurs codons correspondent au même acide aminé), le polymorphisme protéique, observé sur gel d'électrophorèse, ne reflète environ que le quart des mutations présentes au niveau de la séquence d'ADN (Grosclaude 1988). De plus, l'étude du polymorphisme protéique réduit le champ d'investigation aux seules séquences codantes (représentant 5 à 10 % seulement du génome), alors que l'étude du polymorphisme nucléotidique permet de positionner des marqueurs anonymes sur l'ensemble du génome (Botstein *et al* 1980), sachant que ce réseau de marqueurs permettra ensuite de détecter des liaisons avec des gènes dont seul le phénotype est connu et donc de les cartographier.

Parmi les marqueurs les plus utilisés, on peut citer les RFLP et les microsatellites. D'autres techniques ont aussi été développées comme les techniques de RAPD et d'AFLP qui permettent, à des coûts moins élevés, d'identifier un polymorphisme.

2 / La carte chromosomique

Alors que la cartographie génétique produit une information de liaison, déconnectée du support

physique que constituent les chromosomes, la cartographie chromosomique a pour objectif la localisation des locus sur les chromosomes et recouvre deux approches différentes : les hybrides somatiques et la carte cytogénétique.

2.1 / Les cartes d'hybrides somatiques ou d'irradiation

Les hybrides de cellules somatiques ont été développés dans les années 1960, après la mise en évidence des phénomènes de fusion cellulaire spontanée et la mise au point de techniques d'induction (virus de Sendai, PEG). Ces hybrides résultent de la fusion de cellules primaires de l'espèce étudiée et d'une lignée établie de rongeur (souris ou hamster en général). Au cours de leur prolifération en culture, les clones de cellules hybrides perdent aléatoirement des chromosomes de la lignée primaire, alors que les chromosomes de rongeur sont conservés. Il est ainsi possible de créer une collection d'hybrides comportant chacun un jeu différent de chromosomes de la lignée primaire. Le nombre de chromosomes retenus est variable, en fonction de la durée de culture, et peut être réduit à un seul dans le cas d'hybrides monoporteurs, comme il en existe chez l'Homme ou le porc (Cuthbert *et al* 1995).

Les clones de la collection d'hybrides sont analysés par tests d'expression enzymatique, hybridation de sondes, amplification PCR, etc, afin de définir tout d'abord des synténies, c'est-à-dire des groupes de locus conservés ou perdus ensemble et donc localisés sur le même chromosome. L'utilisation des hybrides somatiques permet ainsi d'assigner des locus à des chromosomes et éventuellement d'obtenir une localisation régionale si les hybrides ne contiennent qu'un fragment d'un chromosome et/ou si leur analyse caryotypique est réalisable.

La construction d'hybrides d'irradiation suit le même protocole, mais, dans ce cas, les cellules de la lignée primaire sont soumises à une irradiation qui fragmente les chromosomes avant fusion. Les cartes d'hybrides d'irradiation sont basées sur l'apparition de cassures dues aux rayons X, tout comme les cartes génétiques sont basées sur l'apparition de cassures dues aux recombinaisons lors de la méiose et sur la recherche de liaisons entre locus. En revanche, les marqueurs n'ont pas besoin d'être polymorphes pour être informatifs et la résolution obtenue ne dépend que de la fréquence de cassure et de la fréquence de rétention. La fréquence de cassure est fonction de l'intensité d'irradiation et il est possible, en modifiant ce paramètre, de construire des hybrides dont la résolution est adaptée à la densité de marqueurs disponibles. Dans la mesure où les hybrides d'irradiation contiennent plusieurs fragments, l'ordre des locus ne peut pas être établi à partir d'un seul hybride, mais uniquement à partir d'une collection d'hybrides. Il a été montré qu'une centaine d'hybrides suffisent pour déterminer l'ordre des marqueurs et donc établir des cartes. Une analyse statistique des données sur l'ensemble de la collection est nécessaire pour déterminer cet ordre. Les distances calculées sont exprimées en centiRays (cR). Par définition, 1 cR_(N rad) correspond à une fréquence de cassure de 1 % entre deux locus, après exposition à une dose de N rad de rayons X. L'équivalence avec la distance physique réelle n'est pas stricte et dépend de l'intensité d'irradiation : 1 cR représente environ 50 à 100 kb. La

résolution est donc 10 à 20 fois supérieure à celle obtenue classiquement par cartographie génétique (1 cM, soit 1000 kb).

2.2 / Les cartes cytogénétiques

L'hybridation moléculaire *in situ* sur métaphases colorées (FISH) permet de localiser et d'ordonner directement des sondes correspondant à des gènes, des marqueurs génétiques, etc. Il est ainsi possible d'établir des cartes cytogénétiques dont la résolution maximale est la bande, soit environ 3 à 10 cM, ce qui implique, dans le meilleur des cas, une résolution d'environ une à trois Mb. En revanche, lorsque cette même technique est appliquée sur chromosomes interphasiques (Rappold *et al* 1984), la chromatine étant près de 20 fois moins condensée, la résolution peut atteindre 50 à 1000 kb (Trask *et al* 1989, Lawrence *et al* 1990). Une résolution de 10 à 200 kb peut être obtenue en hybridant les sondes directement sur la chromatine extraite des noyaux (l'ADN décondensé forme des boucles autour des restes de matrice cellulaire, donnant l'impression d'un halo). Enfin, une précision de l'ordre de la kb est à présent atteinte avec l'hybridation sur chromosomes étirés (Fiber FISH et peignage moléculaire). Les segments d'ADN, étirés au point d'atteindre un état double brin relâché, permettent l'hybridation et sont visibles sous microscope. L'utilisation de fluorochromes différents produit une image multicolore à partir de laquelle il est possible d'orienter, d'ordonner des contigs de YAC, BAC ou cosmides, de mesurer la taille des recouvrements et des trous et donc d'établir une carte fine de la région concernée.

3 / La cartographie comparée

La cartographie comparée consiste à déterminer les régions chromosomiques homologues conservées dans plusieurs espèces et à en préciser les bordures. Deux approches concourent à établir la carte comparée : d'une part, la localisation de gènes homologues dans plusieurs espèces, permettant de définir des segments conservés et, d'autre part, la cytogénétique comparée, permettant, par une approche plus globale, d'identifier de grandes régions d'homologie. L'identification de ces segments homologues permet de transposer les connaissances acquises dans des espèces très étudiées comme l'Homme ou la souris à des espèces moins bien connues comme les ruminants, facilitant ainsi le clonage positionnel. De plus, la cartographie comparée apporte un éclairage sur l'évolution des génomes.

Le développement des techniques de biologie moléculaire, la construction des collections d'hybrides somatiques et l'établissement des cartes génétiques à base de RFLP ont permis l'essor de la cartographie comparée. Hybrides somatiques et cartes génétiques ne fournissent cependant pas la même information.

En effet, l'assignation à un chromosome par la voie des hybrides somatiques permet de détecter la conservation de synténies, c'est-à-dire la localisation d'un ou plusieurs gènes homologues sur les chromosomes homologues de deux (ou plusieurs) espèces.

La carte génétique permet en plus d'étudier l'ordre des gènes et donc d'identifier des segments conservés, c'est-à-dire des associations synténiques de deux (ou plus) gènes non interrompues par d'autres fragments chromosomiques, et des segments d'ordre conservé, c'est-à-dire des associations synténiques de plusieurs gènes localisés dans le même ordre sur les chromosomes homologues de deux espèces. La localisation chromosomique par hybridation *in situ* ou par hybrides d'irradiation fournit une information équivalente à l'information de liaison, mais avec une résolution moindre pour l'hybridation *in situ* et supérieure pour les hybrides d'irradiation. Le terme de SCEUS (pour Smallest Conserved Evolutionary Unit Segment) désigne le plus petit segment conservé entre plusieurs espèces.

Les études comparatives, basées sur la morphologie et le dessin des bandes des chromosomes, s'appliquent difficilement aux espèces éloignées sur le plan phylogénétique et ayant subi de profonds remaniements, sauf cas particuliers comme ceux de l'Homme et du lapin (Dutrillaux *et al* 1980) et de l'Homme et du chat (Nash et O'Brien 1982) pour lesquels le dessin des bandes est conservé pour certains chromosomes. C'est pourquoi d'autres techniques comme l'hybridation de sondes marquées sur chromosomes métaphasiques ont été développées. L'hybridation *in situ* en fluorescence de sondes produites à partir de chromosomes isolés (triés, brûlés ou microdisséqués, amplifiés par DOP-PCR, PARM-PCR, etc) d'une espèce de référence (généralement l'Homme) sur les métaphases d'espèces appartenant à des ordres différents (porc, ruminants, cheval par exemple) permet de détecter les régions chromosomiques homologues entre ces espèces et ainsi d'étudier globalement la synténie et les remaniements chromosomiques. Les résultats obtenus sont cependant relativement imprécis puisque seuls les remaniements inter-chromosomiques peuvent être mis en évidence. En revanche, l'hybridation de sondes contenant des gènes sélectionnés à partir de leur position sur la carte humaine, permet d'observer les régions de rupture de synténie de façon précise, les régions sans information de cartographie comparée et, enfin, les éventuels remaniements intra-chromosomiques.

4 / La carte physique

Le but ultime des projets de cartographie est l'étude de gènes responsables de maladies chez l'Homme et l'étude de gènes d'intérêt économique pour les espèces d'élevage. Cette étude suppose de disposer de ces gènes sous forme clonée. Or, ni la carte génétique ni la carte chromosomique ne permettent d'accéder directement à l'ADN cloné, contrairement à l'approche physique pour laquelle les banques d'ADN génomique constituent l'outil fondamental. En effet, comme la taille maximale d'une séquence est de 600 à 1000 paires de bases, le séquençage s'effectue à partir de fragments d'ADN clonés en plasmides. L'organisation des séquences obtenues nécessite alors de positionner ces fragments, à la fois les uns par rapport aux autres et par rapport aux chromosomes dont ils sont issus, afin de reconstituer le génome dans sa globalité. C'est pourquoi la cartographie physique recouvre plusieurs technologies différentes et complémentaires permettant d'assurer la continuité de l'échelle du génome à l'échelle de la séquence : la cartographie en champ pulsé, les alignements de contigs.

De la même façon que l'on réalise une carte de restriction d'un plasmide avec différentes combinaisons d'enzymes de restriction et de sondes, il est possible d'établir une carte physique d'un génome complexe. La distance entre deux locus peut être déterminée directement par analyse de fragments de restriction : la taille du plus petit fragment contenant les deux locus permet d'estimer la distance qui les sépare. En règle générale, les locus utilisés correspondent à des marqueurs génétiques espacés de plusieurs cM, soit plusieurs mégabases. Il est donc nécessaire de réaliser des digestions partielles et/ou d'utiliser des enzymes de restriction à sites rares (NotI, MluI) afin de libérer des fragments d'ADN suffisamment grands. Compte tenu de leur taille importante, les fragments générés peuvent être séparés par électrophorèse en champ pulsé.

L'alignement de clones chevauchants (contigs) couvrant l'ensemble du génome correspond à la première étape de la cartographie physique. Ce travail d'alignement a été réalisé très tôt pour des organismes ayant un génome de taille modeste. Une carte physique d'*E. Coli*, par exemple, a été construite dès 1987 par alignement de phages sur 4,7 Mb. Mais il est évident que ces alignements ne sont pas envisageables dans le cas des mammifères dont la taille du génome est estimée à 3000 Mb. L'effort de construction des contigs étant inversement proportionnel à la taille des fragments étudiés, plusieurs types de vecteurs ont été développés afin de permettre le clonage de grands fragments, certains pouvant dépasser la mégabase :

- les cosmides sont des vecteurs hybrides, construits à partir de séquences plasmidiques et des séquences *cos* du phage lambda. L'ADN génomique est cloné classiquement dans le vecteur cosmétique. La taille des fragments à cloner doit impérativement se situer entre 35 et 45 kb afin de permettre l'encapsulation ;

- les YAC ont été développés à partir des trois éléments de séquences nécessaires et suffisants à la duplication et la ségrégation des chromosomes chez la levure (origine de répllication, centromère et télomère). Ils constituent ainsi de véritables chromosomes artificiels de levure, d'où leur nom « Yeast Artificial Chromosomes ». Leur capacité de clonage atteint deux mégabases (Méga YAC), avec une taille moyenne de 400 kb habituellement ;

- les P1 dérivent du phage P1 et permettent le clonage de fragments de 70 à 100 kb ;

- le vecteur PAC dérive du phage P1, en conserve les avantages, mais autorise une transformation plus simple par électroporation, ce qui permet de s'affranchir des limites d'encapsulation et de cloner des fragments plus grands, de 100 à 300 kb ;

- le système BAC utilise un vecteur basé sur le facteur F d'*E. Coli* dont la répllication est strictement contrôlée, améliorant ainsi la stabilité des inserts de même taille que les PAC.

La mise au point de ces nouveaux vecteurs a révolutionné le domaine de la cartographie physique. En effet, il devenait envisageable de constituer des banques ordonnées de clones individualisés et, en utilisant ces nouveaux outils, d'ordonner ces clones en contigs couvrant tout ou partie du génome. Pour ce faire, plusieurs techniques ont été développées : l'analyse de fragments de restriction (fingerprint), l'analyse du contenu en STS et l'hybridation croisée.

L'approche de fingerprinting, déjà utilisée pour la cartographie en phage d'*E. Coli*, de la levure ou du nématode *Caenorhabditis elegans*, a été appliquée pour l'établissement de la carte physique humaine. Le principe de cette technique est basé sur l'analyse comparée des fragments de restriction entre clones, afin de déterminer des fragments communs et donc le chevauchement.

Le développement de la cartographie génétique à l'aide de microsattellites a fourni un nombre important de marqueurs STS (Sequence-Tagged Site), faciles à utiliser par PCR. Le criblage des banques avec ces marqueurs permet d'isoler des clones chevauchants et l'analyse du contenu en STS de l'ensemble des clones permet la reconstitution de l'ordre chromosomique.

Les différentes stratégies exposées précédemment possèdent toutes de sérieuses limitations. C'est pourquoi, lors de la réalisation de la première carte physique complète du génome humain, l'équipe du Généthon (Cohen *et al* 1993) a cherché à tirer profit de toutes les techniques pour en combiner les avantages tout en minimisant les inconvénients. En effet, une partie des 98208 clones de la banque, soit 33000 YAC de 900 kb de moyenne, ont été analysés à la fois par fingerprinting (trois enzymes de restriction et deux sondes répétées), par leur contenu en STS (2100 marqueurs microsattellites incorporés dans la carte génétique) et par hybridation avec des sondes de PCR inter-Alu provenant de 25000 YAC. De plus, 500 YAC positifs pour des marqueurs génétiques ont été localisés cytogénétiquement par FISH. La poursuite de ce travail a conduit deux ans plus tard à la publication d'une carte de seconde génération (Chumakov *et al* 1995), couvrant près de 75 % du génome en 225 contigs de 10 Mb de moyenne et comportant près de 4500 STS.

Conclusion

Les travaux de cartographie génétique, cytogénétique et comparée ont permis de mettre en place les outils indispensables au clonage positionnel des gènes d'intérêt agronomique chez les animaux d'élevage. Le développement de la cartographie comparée avec l'Homme ou la souris, mais aussi entre toutes les espèces d'élevage (ruminants, porc et cheval), devrait permettre d'identifier des segments conservés entre toutes ces espèces, afin d'enrichir virtuellement la carte d'une espèce avec les gènes localisés dans les autres ou même d'identifier certains gènes ou QTL dans une espèce sur la base de l'existence d'un QTL dans une autre. On peut par exemple imaginer que certains QTL de production de viande ou de prolificité détectés chez le porc soient également présents chez les bovins ou le mouton. Une approche ciblée utilisant ces QTL «candidats» pourrait peut-être permettre de rattraper le retard accumulé dans ces domaines chez les ruminants.

Chez les animaux d'élevage, l'enjeu majeur des prochaines années résidera probablement dans notre capacité à exploiter efficacement en sélection les gènes ou QTL mis en évidence. A moyen terme, il est cependant probable qu'il soit nécessaire d'identifier et de cloner les gènes responsables des caractères d'intérêt, d'une part pour permettre une utilisation optimale de la sélection et d'autre part, pour permettre la création d'animaux transgéniques. La difficulté majeure pour les généticiens résidant alors dans l'obtention de clones contenant les gènes d'intérêt, le moment est donc venu de relever le défi du passage de la carte génétique à la carte physique.

Références

- Botstein D., White R.L., Skolnick M., Davis R.W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32, 314-331.
- Chumakov I.M., Rigault P., Le Gall I., Bellanne-Chantelot C., *et al*, 1995. A YAC contig map of the human genome. *Nature*, 377, 175-297.
- Cohen D., Chumakov I.M., Weissenbach J., 1993. A first-generation physical map of the human genome. *Nature*, 366, 698-701.
- Cuthbert A.P., Trott D.A., Ekong R.M., Jezzard S., *et al*, 1995. Construction and characterization of a highly stable human: rodent monochromosomal hybrid panel for genetic complementation and genome mapping studies. *Cytogenetics and Cell Genetics*, 71, 68-76.
- Dutrillaux B., Viegas-Pequignot E., Couturier J., 1980. [Great homology of chromosome banding of the rabbit (*Oryctolagus cuniculus*) and primates, including man]. *Annale de Génétique*, 23, 22-25.
- Grosclaude F., 1988. Le polymorphisme génétique des principales lactoprotéines bovines. *INRA Productions Animales*, 1, 5-17
- Lawrence J.B., Singer R.H., McNeil J.A., 1990. Interphase and metaphase resolution of different distances within the human dystrophin gene. *Science*, 249, 928-932.
- Nash W.G., O'Brien S.J., 1982. Conserved regions of homologous G-banded chromosomes between orders in mammalian evolution: carnivores and primates. *Proc Natl Acad Sci U S A*, 79, 6631-6635.
- Rappold G.A., Cremer T., Hager H.D., Davies K.E., Muller C.R., Yang T., 1984. Sex chromosome positions in human interphase nuclei as studied by in situ hybridization with chromosome specific DNA probes. *Human Genetics*, 67, 317-325.
- Trask B., Pinkel D., van den Engh G., 1989. The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics*, 5, 710-717.