

J. GELLIN

INRA, Laboratoire de Génétique Cellulaire,
BP 27, 31326 Castanet Tolosan cedex

e-mail : Joel.Gellin@toulouse.inra.fr

3 - Cartographie des génomes

La cartographie comparée des génomes des vertébrés

Résumé. La cartographie comparée des génomes consiste à étudier le plus précisément possible les homologies existantes dans la structure des génomes des organismes vivants. Ces informations sont très utiles notamment pour l'étude des animaux domestiques. Elles permettent d'accélérer l'acquisition et l'exploitation des données de cartographie génétique dans chaque espèce. Les méthodes et les techniques utilisées sont présentées dans un cadre historique et les résultats les plus marquants sont résumés. Cette présentation permet de comprendre pourquoi cette activité évolue rapidement aujourd'hui dans le cadre du développement des cartes génétiques dites à haute densité. Dans ce contexte sont précisées les limites de ces études et les méthodes d'analyse à privilégier ou à développer.

Les travaux actuels sur le génome humain, à notre connaissance, sont : le programme Hugo pour obtenir la séquence complète du génome humain, les différents programmes pour réaliser des cartes à haute densité avec des marqueurs de type SNP (Single Nucleotide Polymorphism) disposés tous les 50/10 kb sur le génome et le programme Unigene d'obtention de la collection complète des gènes humains.

Les projets de séquençage complet du génome d'espèces comme la poule, les bovins, le porc, sont une suite logique aux efforts consentis sur le génome humain. Le séquençage intégral des génomes est, en effet, le seul moyen sûr d'accéder à la totalité des gènes. Cette connaissance est importante, par exemple dans le cas du porc pour lequel des xéno-transplantations sont envisagées chez l'Homme. Plus généralement, la possession d'un réseau dense de marqueurs permet des analyses globales et systématiques grâce aux technologies émergentes des filtres à haute densité et des puces à ADN. Ces données pourront être corrélées aux variations des performances zootechniques des animaux d'élevage ; c'est une solution pour étudier les QTL dans l'ensemble des populations animales (Kruglyak 1999). Nous pourrions également quantifier globalement la 'réactivité' du génome aux variations de l'environnement et aux agressions des pathogènes, apprécier la diversité génétique des espèces et assurer une 'traçabilité' des produits de l'agro-alimentaire.

Les différentes stratégies de cartographie comparée présentées dans cet article ont permis d'obtenir des premières cartes comparées génétique et physique des génomes des animaux en utilisant principalement des données disponibles essentiellement chez l'Homme. Cette activité - comme les autres concernant le génome - devient globale et systématique et ceci en accord avec l'accumulation des données dans un plus grand nombre d'espèces. Dans ce cadre, les initiatives de

séquençage deviennent prépondérantes. La comparaison des séquences d'ADN est un travail de bioinformatique, activité dite 'in silico', qui consiste à intégrer au mieux les données de séquences - devenant la base de l'analyse dans les comparaisons des génomes - avec des données supplémentaires aussi différentes que les localisations chromosomiques, les fonctions et les régulations des gènes orthologues⁽¹⁾ décrites dans chaque espèce.

1 / Le concept de la cartographie comparée

La cartographie comparée consiste à déterminer les régions chromosomiques homologues conservées dans plusieurs espèces. Il convient de préciser ces homologies de plus en plus finement.

Les génomes des vertébrés, et plus précisément des mammifères, sont bâtis avec un plan d'ensemble assez commun. Celui-ci a été par la suite plus ou moins 'brouillé' par la spéciation, mais l'organisation spatiale des gènes est souvent restée semblable pour les différentes espèces. La comparaison de toutes les espèces entre elles doit, à terme, permettre de comprendre quelques grands mécanismes de l'évolution et relier la localisation des gènes sur les chromosomes à leurs fonctions dans la cellule. On utilise actuellement, de façon pragmatique, les données partielles obtenues dans une espèce pour faire des hypothèses sur la structure génomique locale dans une autre et progresser ainsi plus vite dans la recherche des QTL et des gènes majeurs.

⁽¹⁾ Un gène orthologue est un gène homologue dans différentes espèces ayant le même gène ancêtre provenant du plus proche ancêtre commun.

L'assignation à un chromosome 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. L'étude, dans les différentes espèces, de l'ordre des gènes permet de définir d'une part des segments conservés, c'est-à-dire des associations synténiques de deux gènes (ou plus) non interrompues par d'autres fragments chromosomiques, et, d'autre part, 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.

Ces données permettent d'augmenter le nombre de marqueurs disponibles dans une zone chromosomique où un gène majeur ou un QTL ont été repérés. Prenons un exemple : sur le chromosome p du porc, nous avons repéré une région qui contient un gène, pour le moment inconnu, intervenant de façon majeure dans la vitesse de croissance de l'animal. Trois marqueurs génétiques permettent de délimiter une zone de quelques cM à l'intérieur de laquelle se trouve ce gène. En étudiant les données disponibles sur la carte humaine, on constate que ces trois gènes marqueurs (les versions supposées orthologues de ces gènes) sont également synténiques chez l'Homme. Ils sont localisés dans une zone restreinte à quelques cM sur le chromosome humain h dans une région bien connue. Des recherches plus nombreuses chez l'Homme ont permis d'isoler aux alentours un grand nombre de marqueurs supplémentaires. Ces marqueurs sont potentiellement utilisables et donnent de nouvelles informations sur la zone du chromosome p du porc. Bien entendu, il sera parfois difficile ensuite de trouver un phénotype équivalent chez l'Homme et de pister directement un gène candidat.

Plusieurs résultats montrent l'intérêt zootechnique de ces méthodes pour caractériser un gène et la mutation responsable d'un phénotype remarquable. Les deux exemples les plus typiques sont d'une part la recherche du gène halothane chez le porc (McCarthy *et al* 1990) et, d'autre part, la recherche du gène culard chez les bovins (Grobet *et al* 1997). Des exemples plus actuels concernent le gène RN chez le porc, le gène Booroola chez le mouton et le gène PIS chez la chèvre.

2 / Les grandes étapes des progrès méthodologiques

On peut discerner quatre types d'approche :

- créer le plus possible de points d'ancrage, c'est-à-dire de repères, communs aux différents génomes. On utilise alors des gènes faciles à comparer et à étudier d'une espèce à l'autre ;
- localiser physiquement des gènes sur les chromosomes. On utilise des panels d'hybrides, cellulaires ou d'irradiation, ou des hybridations *in situ* sur des chromosomes en métaphase ;
- hybrider *in situ* avec des sondes complexes sur les chromosomes pour évaluer les grandes régions d'homologie sur l'ensemble des chromosomes. Il s'agit d'expériences de 'painting' ;
- réaliser la collection systématique des gènes de chaque espèce et comparer les structures des génomes 'in silico'. Il s'agit, dans ce dernier cas, de l'approche bioinformatique.

Une brève présentation historique est nécessaire. Pour comparer les génomes, on a utilisé essentiellement des méthodes de cartographie physique. Dans un premier temps (à partir de 1980), les recherches ont porté sur l'analyse de l'expression des protéines dans des hybrides cellulaires inter-spécifiques. Les premières comparaisons ont concerné la conservation éventuelle de synténies d'une espèce à l'autre. Rapidement, la biologie moléculaire a permis l'analyse de l'ADN des gènes, qu'ils soient ou non exprimés sous forme de protéines dans l'hybride cellulaire. Cette analyse devenait donc plus globale et plus fiable. Néanmoins, elle restait complexe car réalisée par une technique assez laborieuse (RFLP et Southern Blotting) et avec des sondes d'ADN souvent, à cette époque, d'origine humaine qui ne fonctionnaient pas toujours très bien sur l'ADN des animaux étudiés.

Les premières cartes physiques composées de groupes de synténie sont apparues, permettant l'étude du maintien ou non de ces synténies d'une espèce à l'autre. L'assignation des gènes sur un chromosome était rarement réalisée vu la complexité du caryotype des hybrides cellulaires. Néanmoins, quelques assignations sur un chromosome ou fragment de chromosome ont été effectuées de cette façon. Ce type d'analyse, fort complexe à l'époque et peu sûr, a été rapidement laissé de côté. En 1990, alors que chez l'Homme on disposait déjà d'une carte génétique, certains chromosomes étaient encore dépourvus de tout marqueur dans les espèces d'élevage. Un grand nombre de groupes de liaison et certaines synténies n'étaient pas assignées à un chromosome (moins de 40 gènes étudiés chez le porc par exemple).

L'hybridation *in situ* s'est développée dès que l'usage des sondes d'ADN s'est répandu (à partir de 1985). Elle permet une localisation sur un chromosome et pas seulement une assignation. Il a fallu, pour plus d'efficacité, essayer de récupérer des sondes d'ADN spécifiques de l'espèce. Les premières localisations *in situ* ont été obtenues par une méthode laborieuse d'hybridation avec des sondes marquées au tritium. Rapidement est apparue l'hybridation *in situ* en fluorescence (FISH : Fluorescence *In Situ* Hybridisation) que l'on pratique beaucoup actuellement. Sa sensibilité est excellente avec les grands vecteurs de clonage (cosmides, BAC).

D'importants programmes de cartographie, auxquels participe l'INRA (Gellin et Grosclaude 1991), ont été initiés en Europe : pour le porc en 1990 (PiGMap, A. Archibald, Roslin Institute, UK), pour les bovins en 1992 (BovMap, H. Levéziel, INRA), puis pour la poule en 1996 (ChickMap, D. Burt, Roslin Institute, UK). Simultanément, des protocoles expérimentaux de grande ampleur étaient mis en place à l'INRA pour recueillir des informations sur les performances des animaux et chercher des QTL. C'est dans ce contexte général et pour accélérer ces travaux que la cartographie comparée s'est développée.

Avec l'utilisation systématique de la technique PCR, les données s'accroissent maintenant très vite. Les panels d'hybrides cellulaires permettent une analyse plus fine. Le contenu en chromosomes ou fragments de chromosomes de chaque hybride est étudié par des méthodes de painting (Yerle 2000, cet ouvrage) et l'analyse de synténie est ainsi asso-

ciée à une localisation plus ou moins précise du gène sur un chromosome. L'analyse des panels d'hybrides d'irradiation représente la méthode de localisation la plus récente et la plus fine (Yerle 2000, cet ouvrage). Ce niveau d'analyse est devenu nécessaire avec l'obtention des cartes de plus en plus denses qui apparaissent à l'heure actuelle.

3 / Localisation systématique de gènes dans différentes espèces

3.1 / La méthode ZooPCR

L'idée pour comparer les génomes a été de localiser dans d'autres espèces des gènes bien connus localisés chez l'Homme et la souris. Ce sont des gènes essentiels au métabolisme de base des cellules, qui sont exprimés sous forme de protéines dans tous les tissus. Ils sont utilisés ici comme marqueurs physiques (des balises en quelque sorte) placés sur un chromosome permettant d'accumuler des données de synténie. Après une analyse génétique adéquate, ils peuvent devenir des marqueurs génétiques s'ils sont polymorphes. Appelés marqueurs de type I ou «d'ancrage», ils sont à opposer aux marqueurs de type II, comme les microsatellites, très polymorphes mais qui ne permettent pas, en général, la transposition de l'information d'une espèce à l'autre.

Une coordination de ces recherches a été proposée par O'Brien *et al* (1993) qui ont publié une liste de 321 locus d'ancrage. L'idée essentielle était de coordonner et focaliser les efforts afin d'optimiser les possibilités de comparaison de l'ensemble des vertébrés. Ces locus, localisés dans les deux espèces (Homme et souris), ont également été choisis pour couvrir au mieux les chromosomes humains en espérant observer également une bonne distribution dans les autres espèces. Cependant, cette première liste désignait des gènes ne couvrant pas vraiment tous les chromosomes humains. A cette époque ces marqueurs étaient étudiés avec la technique RFLP et le Southern Blotting (Riquet et Pitel 2000, cet ouvrage) sur un panel d'hybrides cellulaires. Cette technique n'était pas simple à mettre en œuvre. Les hybridations *in situ* pouvaient également être utilisées, mais en marquant les sondes au tritium, ce qui n'était pas aisé non plus. Cette initiative a néanmoins orienté le choix de gènes à localiser et a facilité la synthèse des données recueillies dans le cadre de publications régulières par O'Brien et ses collaborateurs.

Le concept de O'Brien *et al* (1993) a été repris dans le cadre des nouvelles possibilités techniques de la PCR. Les séquences d'ADN de nombreux gènes sont maintenant disponibles. Ces données 'in silico' sont souvent recueillies pour différentes versions orthologues du même gène après des études dans diverses espèces. La comparaison de ces séquences permet de définir des amorces 'consensus' qui, en PCR, amplifient le gène correspondant dans différents génomes suffisamment apparentés. Afin de vérifier que les amorces ont bien fonctionné et bien amplifié le gène ciblé, la spécificité du pro-

duit obtenu est contrôlée par séquençage. Ces méthodes (ZooPCR) ont permis d'ajouter beaucoup de nouveaux marqueurs de type I sur les cartes des animaux domestiques. Des approches différentes, mais très similaires, ont été présentées dans la littérature (CATs, TOASTs, ComparEST, COMPASS...). Dans tous ces cas, les localisations sont obtenues avec des panels d'hybrides cellulaires ou les panels d'hybrides d'irradiation disponibles chez les bovins, le porc et bientôt chez la poule.

Lyons *et al* (1997) ont prolongé l'initiative de O'Brien en sélectionnant des amorces pour les gènes sélectionnés : 537 gènes ont été répertoriés et des amorces consensus ont été caractérisées pour une utilisation potentielle chez les vertébrés. Des amorces ont été finalement définies pour 410 gènes qui ont été appelés des CATs (Comparative Anchor Tagged Sequences).

Dans le cas des CATs, les amorces sont en général définies dans des exons consécutifs, l'idée étant d'optimiser les différences dans la taille des fragments amplifiés entre les différentes espèces et de faciliter la localisation des gènes. La localisation effectuée avec des hybrides cellulaires ou d'irradiation (hamster ou souris / animaux domestiques, porc, bovins) aboutit souvent à l'amplification du gène dans les deux espèces composant l'hybride. Si la différence de taille des deux produits d'amplification est suffisamment grande, un simple gel d'agarose suffit pour la visualisation et l'analyse des clones positifs et négatifs. Dans le cas contraire, il est nécessaire d'utiliser des gels SSCP plus résolutoires mais plus compliqués à mettre en œuvre et d'optimiser la probabilité de trouver un polymorphisme du gène dans chaque espèce grâce à la partie non codante amplifiée. Dans le cas des CATs, le choix des séquences consensus est en fait orienté pour amplifier correctement l'ADN du chat (espèce favorite des auteurs de la méthode). Ces amorces fonctionnent en effet avec un bon pourcentage pour cette espèce. L'usage des CATs pour les autres espèces comme le porc et les bovins donne des résultats variables et pas toujours faciles à interpréter. Les choix des auteurs ont des inconvénients importants. En effet, les amorces choisies de part et d'autre d'un intron donnent comme on l'a vu des produits de taille variable d'une espèce à l'autre. Cet effet positif est contrebalancé par le fait qu'il est souvent difficile d'affirmer qu'il s'agit bien d'une amplification spécifique, c'est-à-dire de l'amplification du gène orthologue. En fait, il est recommandé dans tous les cas de séquencer le produit obtenu pour une vérification finale, ce qui alourdit beaucoup les expériences.

Les initiatives sur les marqueurs d'ancrage ont permis de faire progresser les cartes comparées et ont aussi popularisé le concept d'amorces consensus malgré les résultats en demi-teinte de l'approche CATs. Les choix, pour définir les amorces consensus, n'ont pas toujours été des plus pertinents. Des résultats ont néanmoins été publiés (Lyons *et al* 1999).

De nouveaux marqueurs d'ancrage ont également été proposés avec des critères différents de choix de séquences d'amorces consensus. Les TOASTs (Traced Orthologous Amplified Sequence Tags) ont été récemment développés par Jiang *et al* (1998). Comme pour les CATs, des couples d'amorces spécifiques de 146 gènes orthologues ont été choisis après comparaison

de séquences de différentes espèces pour les utiliser en cartographie comparée. Cependant, ces marqueurs TOASTs sont plus prometteurs pour la cartographie comparée, notamment chez le porc, puisque les amorces correspondantes ont essentiellement été choisies sur des séquences porcines et en excluant la plupart du temps des séquences introniques. Aujourd'hui, sur ces 146 TOASTs, 50 gènes sur 66 testés ont été localisés et séquencés. Beaucoup des localisations obtenues sont en accord avec les données de cartographie comparée déjà recueillies. Actuellement les TOASTs sont testés chez le chien (Vignaux *et al* 1999) et aussi utilisés pour isoler les BAC correspondants chez le cheval (G. Guérin, communication personnelle).

Depuis 1996, notre groupe s'est engagé avec d'autres dans des programmes d'utilisation d'étiquettes (ou EST : expressed sequence tags) humaines comme outil de cartographie des génomes des vertébrés. Le but est également de multiplier les points d'ancrage sur les différents génomes étudiés. Le Généthon nous a fourni les amorces humaines correspondant à des EST finement cartographiées chez l'Homme avec un panel d'hybrides d'irradiation. Dans le cadre du projet de cartographie humaine, les chercheurs du Généthon ont testé en ZooPCR environ 10 000 EST sur l'ADN de porc, de bœuf, de chien et de souris. Les amorces qui amplifient l'ADN de ces différentes espèces ont été fournies aux laboratoires impliqués dans la cartographie des espèces en question : 10 % des amorces amplifient l'ADN d'au moins une espèce autre que l'Homme. 344 EST ont été testées chez le porc et 600 chez les bovins. Aujourd'hui, après analyse sur les panels d'hybrides cellulaires de chaque espèce, 65 EST sont localisées chez le porc et 120 chez les bovins.

La synthèse des résultats reste à faire. Un certain nombre d'EST ont été testées chez la souris ou le chien. Il semble difficile d'obtenir des EST consensus, c'est-à-dire fonctionnant chez de nombreuses espèces. En effet, ces EST sont très peu nombreuses contrairement à ce que l'on espérait au début du projet et sont les plus difficiles à utiliser. D'une part, ils amplifient, comme attendu, l'ADN du hamster présent dans les hybrides cellulaires aussi bien que l'ADN de l'espèce et, d'autre part, donnent souvent la même taille d'amplification pour toutes les espèces. Ceci oblige, pour l'analyse sur des panels d'hybrides cellulaires, à utiliser des méthodes de séparation d'ADN sur gel plus sophistiquées qu'un classique gel d'agarose.

H. Lewin et collaborateurs ont voulu rendre encore plus systématique cette idée de comparaison. Ces auteurs définissent une nouvelle notion pour la cartographie comparée qu'ils intitulent COMPASS (CComparative Mapping by Annotation and Sequence Similarity ; Ma *et al* 1998, Ozawa *et al* 2000). Les données de séquences recueillies dans les bases internationales sont analysées pour prédire la localisation des EST bovines en se basant sur l'ensemble des informations plus ou moins complètes de cartographie comparée. Ils ont localisé ainsi 'in silico' plus de 1000 EST. De l'ordre de 100 EST ayant un gène humain orthologue ont été réellement localisées sur un panel d'hybrides d'irradiation et beaucoup se localisent dans la zone attendue. Cette approche COMPASS préfigure la cartographie comparée de demain qui sera essentiellement fondée sur la gestion bioinformatique des données de localisations chromosomiques des séquences spécifiques obtenues dans chaque espèce.

3.2 / Les méthodes FISH, de painting et ZooFISH

Parallèlement aux approches de biologie moléculaire, la cartographie essentiellement cytogénétique a permis le développement d'une connaissance moins précise mais plus systématique et globale des similitudes entre les génomes des mammifères.

Le 'painting hétérologue' consiste à utiliser une sonde complexe spécifique d'un chromosome d'une espèce donnée et de l'hybrider (par FISH) sur les chromosomes en métaphase d'une autre espèce, afin de déterminer les équivalences au niveau chromosomique. Cette technique permet d'obtenir une vue générale comparée de la structure des génomes des mammifères et de définir les grands fragments de synténie conservés. Les détails comme l'ordre des gènes et les limites exactes des fragments conservés ne peuvent ici être analysés ; on obtient uniquement un schéma général de comparaison. Cette approche, d'abord mise au point pour déterminer les homologies entre les chromosomes humains et ceux de divers primates, a été ensuite appliquée à la comparaison des chromosomes humains avec ceux de mammifères plus éloignés sur le plan phylogénétique, tels que le porc et le bovin.

Des hybridations de l'ADN de chaque chromosome porcine ont été effectuées sur des chromosomes humains en métaphase, et l'expérience a également été réalisée dans le sens inverse, c'est-à-dire une hybridation des sondes spécifiques de chaque chromosome humain sur des chromosomes de porc en métaphase. Ce processus bidirectionnel de painting a permis de déterminer les homologies entre les chromosomes des deux espèces segment à segment (tableau 1 et figure 1, Goureau *et al* 1996).

Des études similaires ont été réalisées entre l'Homme et les bovins et entre le porc et les bovins. La synthèse de l'ensemble de ces données permet d'établir les homologies entre les chromosomes humains, porcins et bovins. Elles sont d'une grande aide dans des stratégies de recherche de gènes candidats. Par exemple, pour l'étude du gène porcine RN, localisé sur Sscr 15 (Sscr 15 pour *Sus scrofa* chromosome 15). Cette région du Sscr 15 étant équivalente à la région Hsap 2q12-q37 (Hsap 2 pour *Homo sapiens* chromosome 2), des recherches de nouveaux marqueurs et de gènes candidats ont été entreprises dans cette zone du chromosome humain (Milan *et al* 2000, cet ouvrage).

Les résultats de painting chez le porc sont détaillés sur le site internet du Laboratoire de Génétique Cellulaire - INRA Toulouse (<http://www.toulouse.inra.fr/lgc/pig/compare/compare.htm>) et, pour les animaux d'élevage en général, dans une revue de Chowdhary *et al* (1998). Aujourd'hui, ce travail a été réalisé dans les principales espèces zootechniques comme les bovins, le porc, le lapin, le mouton et le cheval. Des efforts complémentaires doivent être faits chez la poule, espèce pour laquelle un painting hétérologue n'a pu être obtenu jusque là. Des essais sont en cours actuellement sur les génomes de la poule et du canard (A. Vignal, communication personnelle).

Tableau 1. Schéma général des homologies entre les chromosomes humain et porc.

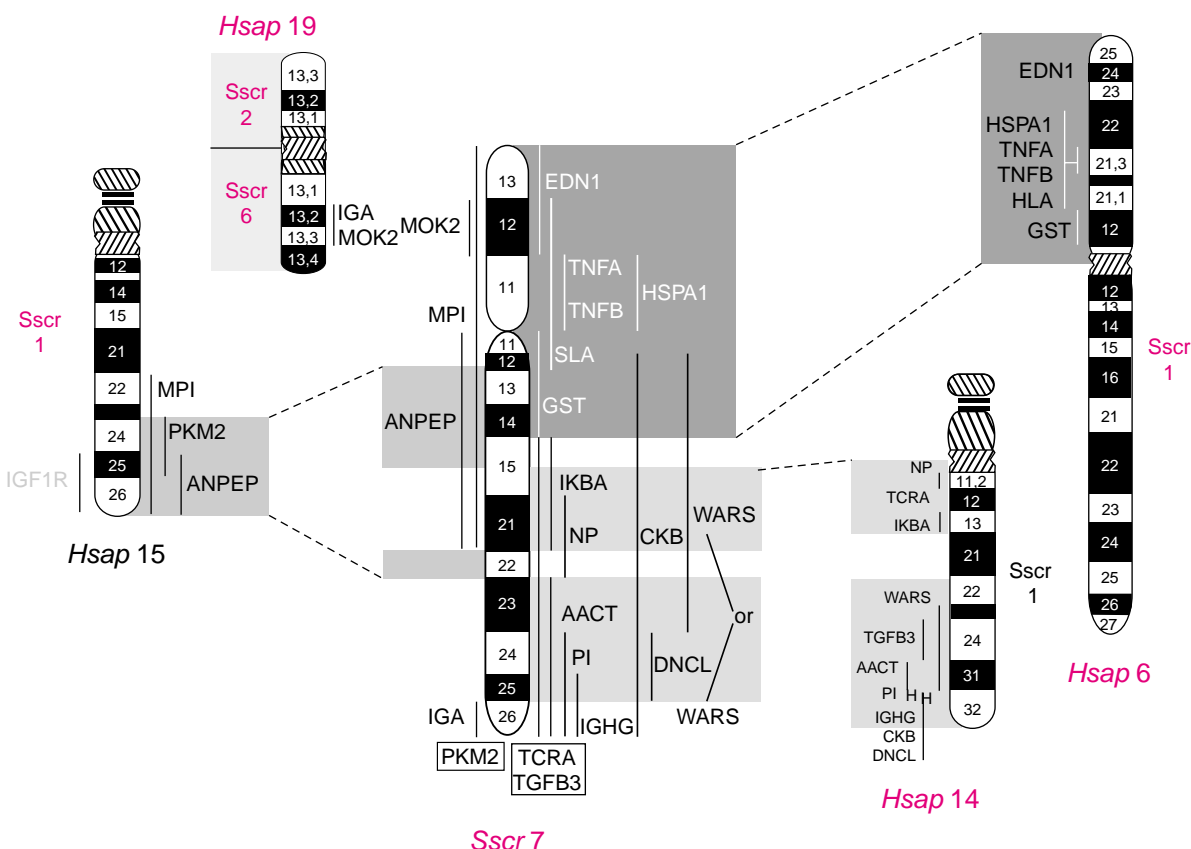
La technique de painting hétérologue (voir le texte) permet d'obtenir une vue générale comparée de la structure des génomes des mammifères et de définir les grands fragments de synténie conservés. Les détails, comme l'ordre des gènes et les limites exactes des fragments conservés, ne peuvent ici être analysés ; on obtient uniquement un schéma général de comparaison.

Hsap pour Homo sapiens et Sscr pour Sus scrofa. Les résultats de painting chez le porc sont détaillés sur le site internet du Laboratoire de Génétique Cellulaire - INRA Toulouse (<http://www.toulouse.inra.fr/lgc/pig/comparer/comparer.htm>)

Chromosome humain	Correspondances avec les chromosomes porcins
Hsap 1	Sscr 04 , Sscr 06 , Sscr 09 , Sscr 10 , Sscr 14
Hsap 2	Sscr 03 , Sscr 15
Hsap 3	Sscr 13
Hsap 4	Sscr 06 , Sscr 08 , Sscr 15 , Sscr 17
Hsap 5	Sscr 02 , Sscr 16
Hsap 6	Sscr 01 , Sscr 07
Hsap 7	Sscr 03 , Sscr 09 , Sscr 18
Hsap 8	Sscr 04 , Sscr 14 , Sscr 15
Hsap 9	Sscr 01 , Sscr 10 , Sscr 14
Hsap 10	Sscr 10 , Sscr 14
Hsap 11	Sscr 02 , Sscr 09
Hsap 12	Sscr 05 , Sscr 14
Hsap 13	Sscr 11
Hsap 14	Sscr 01 , Sscr 07
Hsap 15	Sscr 01 , Sscr 07
Hsap 16	Sscr 03 , Sscr 06
Hsap 17	Sscr 02 , Sscr 12
Hsap 18	Sscr 01 , Sscr 06
Hsap 19	Sscr 02 , Sscr 05 , Sscr 06 , Sscr 07
Hsap 20	Sscr 17
Hsap 21	Sscr 06 , Sscr 09 , Sscr 13
Hsap 22	Sscr 05 , Sscr 14

Figure 1. Comparaisons entre le chromosome 7 porc (Sscr 7) et les fragments équivalents sur des chromosomes humains (Hsap 6, 14, 15, 19).

Des hybridations de l'ADN de chaque chromosome porc ont été effectuées sur des chromosomes humains en métaphase, et l'expérience a également été réalisée dans le sens inverse, c'est-à-dire hybridation de l'ADN de chaque chromosome humain sur des chromosomes de porc en métaphase. Ce processus bidirectionnel de painting a permis de déterminer les homologies entre les chromosomes des deux espèces segment à segment (Goureau et al 1996). Des marqueurs de type I (voir le texte) sont également indiqués et complètent la comparaison.



La localisation cytogénétique (par FISH notamment) de gènes identifiés et de clones à grandes insertions (BAC) contenant des gènes ou des marqueurs génétiques est devenue l'approche fondamentale pour l'établissement des homologies entre génomes. Elle est maintenant potentiellement très utile, car il est démontré que l'on peut utiliser systématiquement des BAC comme marqueurs d'ancrage.

Schibler *et al* (1998) augmentent le nombre de marqueurs de type I en fabriquant des amorces à partir des informations 'in silico' sur la séquence de gènes localisés sur le génome humain et isolent, par hybridation hétérologue, des BAC de chèvres. Ces BAC sont localisés *in situ* et constituent autant de points de comparaison avec l'Homme. Ces BACs ont également été localisés par FISH sur le génome du porc (166 BACs de chèvre ont été testés sur le génome du porc, 114 ont été localisés sur le génome du cheval).

Chez la poule, les nombreux microchromosomes, impossibles à différencier par des méthodes de «banding» classique, ont été «étiquetés» en utilisant des BAC en FISH et par analyse du marquage des sondes en deux couleurs (Fillon *et al* 1998). Il faut également noter que ces BAC de poule semblent fonctionner à 100 % sur le génome du canard (V. Fillon, communication personnelle).

Les BAC présentent donc aujourd'hui un intérêt fondamental pour multiplier les points de comparaison entre plusieurs espèces. Toutes ces études en ZooFISH utilisant des BAC contenant un gène sont remarquablement efficaces. Ils représentent le meilleur compromis entre des vecteurs de clonage permettant des insertions plus grandes mais structurellement souvent chimériques et plus difficiles à obtenir comme les YAC (extraction, préparation de l'ADN). Les cosmides, quant à eux, ne clonent que des petits fragments d'ADN. De plus, nous avons constaté que les YAC sont plus difficiles que les BAC à utiliser dans les conditions d'hybridation hétérologue. Ils donnent souvent des résultats en FISH faibles ou non exploitables quelles que soient les techniques utilisées pour préparer une sonde fluorescente à partir de l'ADN de l'insertion. A cette remarque il faut ajouter le fait que les grands projets de «contigage» qui prendront corps rapidement chez le porc, les bovins et la poule seront effectués comme chez l'Homme avec des BAC. La possibilité de croiser les données 'in silico' ou de FISH pour comparer «BAC à BAC» l'ordre des gènes dans des régions entières de génomes de différents mammifères sera probablement très utile pour une aide à la réalisation des cartes à haute densité. Les banques de BACs sont maintenant disponibles pour toutes les espèces majeures (Rogel-Gaillard 2000, cet ouvrage). Ces banques doivent être bien représentatives, suffisantes en terme d'équivalent génome et être constituées de clones avec de grandes insertions non chimériques.

4 / Vers une carte complète du génome pour chaque espèce

Malgré les résultats obtenus par ZooFISH et ZooPCR, cette activité de comparaison au sens strict a ses limites. Il est maintenant impératif d'isoler les gènes spécifiques de chaque espèce et d'assurer ainsi une analyse complète du génome. Les banques d'ADNc obtenues chez l'Homme et la souris ne sont pas suffisantes, ni pour l'étude des QTL, ni pour isoler les gènes que nous recherchons chez les animaux domestiques. Il faut en effet éviter une analyse trop complexe sur des gènes orthologues avec souvent des familles de gènes différemment structurées d'une espèce à l'autre, avec des différences fonctionnelles importantes entre gènes apparentés (schémas d'expression, «enhancers», isoformes et dosage de gènes). L'analyse de la séquence des gènes ne permet pas toujours de discerner sans ambiguïté les orthologies des paralogies⁽²⁾, par exemple pour les gènes appartenant à des familles multigéniques complexes.

Le Comité de suivi sur la génomique animale entre les USA et l'Europe⁽³⁾, qui a été créé en 1998 à l'occasion d'une réunion à Bruxelles (activités de la commission européenne DG XII) a proposé de retenir comme actions prioritaires les travaux sur l'isolement des EST et sur la bioinformatique (isolement de banques d'ADNc à partir de différents organes des espèces majeures - bovin, porc, poule, saumon - séquençage, localisation et intégration des données dans des bases de données). Cette priorité a été récemment rappelée dans un workshop Hugo sur la cartographie comparée (Gellin *et al* 1999).

De nombreuses initiatives sont déjà engagées dans ce sens. En Europe, le programme Genetpig (Genes controlling economic traits in pig, 1999-2000) consiste à localiser 800 nouveaux gènes sur le génome du porc en utilisant des banques d'ADNc isolés de différents organes. CompRad (Comparative Radiation hybrid mapping) est un programme équivalent chez les bovins. Ces deux projets doivent être considérés comme des programmes pilotes dans ce domaine.

Il s'agit de créer un répertoire très large des gènes exprimés (50 000 à 100 000 séquences). A terme, l'objectif est un répertoire exhaustif de tous les gènes exprimés de l'espèce étudiée. Cette information, qui sera alors complémentaire de la connaissance totale de génomes des vertébrés (Homme, souris, tétraodon ou fugu), permettra d'accéder rapidement à l'identification de gènes impliqués dans la variabilité de fonctions (analyse fonctionnelle), ainsi qu'à l'analyse génétique dans les populations (identification de

⁽²⁾ Des gènes paralogues sont des gènes qui se ressemblent chez un individu et qui appartiennent à une même famille multigénique. Ils peuvent être synténiques ou pas. Ils sont issus d'une duplication ancestrale.

⁽³⁾ Steering group on Farm Animal Genomics. Dave Burt et Jim Womack sont les coordinateurs Europe/USA de ce groupe composé de 6 membres (3 pour les USA et 3 pour l'Europe, dont deux membres INRA).

QTL) par la possibilité de définir de nombreux polymorphismes SNP (marqueurs de type III). Ces données serviront bien entendu aux futurs travaux de cartographie génétique comparée.

5 / Synténie et ordre des gènes : des idées nouvelles

En l'état actuel des connaissances, il est intéressant de s'interroger sur l'existence de mécanismes spécifiques pour l'évolution des génomes des mammifères. Les méthodes de localisation des gènes sont nombreuses et deviennent plus précises (localisation génétique en cM ou localisation physique en cR). Nous pouvons maintenant ordonner les gènes avec une grande fiabilité et poser la question de la conservation de cet ordre d'une espèce à l'autre dans les zones de synténies conservées. Des travaux sur la souris et sur l'Homme ont mis en évidence un nombre insoupçonné de réarrangements internes (Carver et Stubbs 1997). Mais la comparaison des cartes de ces deux espèces a montré des particularités que l'on ne peut pas généraliser aux autres mammifères.

Des panels d'hybrides d'irradiation permettent de faire ce travail dans d'autres espèces. Les réarrangements sont fréquents, mais leur nombre et leur nature dépendent de la zone étudiée. Des comparaisons ont été réalisées entre les bovins et la chèvre (Piumi *et al* 1998). Sun *et al* (1999) ont présenté une comparaison détaillée de l'ordre des gènes chez l'Homme (Hsap 3) et le porc (Sscr 13) dans une zone de synténie considérée comme conservée après des études de painting hétérologue et montré l'existence de remaniements. Yang et Womack (1998) utilisent, quant à eux, la cartographie par hybrides d'irradiation en parallèle dans les deux espèces qu'ils comparent, réalisent un «paralleled RH mapping» et montrent qu'il y a beaucoup de remaniements dans les zones clairement définies jusque là comme des zones de synténie conservées.

Nous devons faire la synthèse de tous ces résultats et de ceux qui vont s'accumuler maintenant très vite suite à la grande disponibilité prévisible des EST, des panels d'hybrides d'irradiation dans la majorité des espèces et du 'contigage' systématique. L'existence fréquente de recombinaisons dans les régions réputées conservées entre espèces pose le problème des limites de l'approche comparée. Celle-ci peut devenir complexe à l'échelle plus précise de quelques kb quand on recherche un gène particulier. On peut espérer que, dans de nombreux cas, il y aura conservation de l'ordre. Des exemples de grande conservation ont permis d'isoler rapidement des gènes comme ceux du syndrome halotha-

ne chez le porc et de la myostatine dans le phénotype chez les bovins. Néanmoins, ces difficultés probables justifient, l'acquisition d'une connaissance la plus complète possible sur chaque espèce.

6 / L'aide de l'informatique

Nous assistons à une accumulation très importante de données de séquençage. L'analyse de ces séquences deviendra à terme l'outil fondamental de comparaison des génomes, la référence pour les comparaisons étant, pour l'ensemble des mammifères, la carte humaine et, secondairement, la carte de la souris. Des informations complémentaires (parentés évolutives, géographiques, fonctionnelles, structurales) seront collectées et combinées d'une façon systématique en continuant les comparaisons entre mammifères mais également chez des vertébrés particuliers comme le fugu, le *Tetraodon nigroviridis*, qui possède des génomes compactés, et la poule, au génome en partie compacté. Ainsi, nous pourrions lever les ambiguïtés de localisation dans le cas de gènes paralogues ou appartenant à des familles de gènes. Ces données sont globalement difficiles à utiliser. Très disparates, elles sont obtenues avec des méthodes plus ou moins précises. Les données utilisables sont, d'une part les résultats de liaison génétique (analyse génétique des familles informatives), exprimés en cM, et d'autre part les résultats de cartographie physique (analyse des panels d'hybrides d'irradiation), exprimés en cR. Dans les deux cas, on a une bonne idée de l'ordre des gènes dans différentes espèces, ce qui constitue un cadre pour la localisation la plus probable d'un nouveau gène. Les données complémentaires sont issues de cartographie physique, par exemple de painting et bien entendu de localisation chromosomique d'EST, de BAC et après 'contigage' de ces éléments.

Conclusion

Des avancées très importantes ont déjà été obtenues par comparaison des cartes des différents espèces animales. Aujourd'hui le développement rapide des cartes génétiques à haute densité entraîne une nécessaire évolution des concepts dans ce domaine. Les banques de BACs et les banques d'ADNc spécifiques d'organes et d'espèces représentent désormais les deux outils de base pour construire des cartes détaillées. Les grands projets de 'contigage' seront effectués avec des BAC, et les collections de gènes seront réalisées avec des banques d'ADNc de référence. L'analyse des séquences d'ADN devient de plus en plus le point central de l'activité de comparaison des génomes et cette nouvelle orientation suppose des développements informatiques conséquents.

Références

Carver E.A., Stubbs L., 1997. Zooming in on the human-mouse comparative map: genome conservation re-examined on a high-resolution scale. *Genome Research*, 7, 1123-1137.

Chowdhary B.P., Raudsepp T., Fronicke L., Scherthan H., 1998. Emerging patterns of comparative genome organization in some

mammalian species as revealed by Zoo-FISH. *Genome Research*, 8, 577-589.

Fillon V., Morisson M., Zoorob R., Auffray C., Douaire M., Gellin J., Vignal A., 1998. Identification of 16 chicken microchromosomes by molecular markers using two-colour fluorescence in

- situ hybridization (FISH). *Chromosome Research*, 6, 307-313.
- Gellin J., Grosclaude F., 1991. Analyse du génome des espèces d'élevage : projet d'établissement de la carte génétique du porc et des bovins. *INRA Productions Animales*, 4, 97-105.
- Gellin J., Brown S., Graves J., Rothschild M., Shook L., Womack J., Yerle M., 2000. 2nd Comparative mapping workshop Toulouse 3rd and 4th May 1999, HUGO report. *Mammalian Genome*, 11, 140-144.
- Goureau A., Yerle M., Schmitz A., Riquet J., Milan D., Pinton P., Frelat G., Gellin J., 1996. Human and porcine correspondence of chromosome segments using bidirectional chromosome painting. *Genomics*, 36, 252-262.
- Grobet L., Martin L.J., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., Dunner S., Ménéssier F., Massabanda J. *et al*, 1997. A deletion in the bovine myostatin gene causes the double-musled phenotype in cattle. *Nature Genetics*, 17, 71-74.
- Jiang Z., Priat C., Galibert F., 1998. Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. *Mammalian Genome*, 9, 577-587.
- Kruglyak L., 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, 22, 139-144.
- Lyons L.A., Laughlin T.F., Copeland N.G., Jenkins N.A., Womack J.E., O'Brien S.J., 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics*, 15, 47-56.
- Lyons L.A., Kehler J.S., O'Brien S.J., 1999. Development of comparative anchor tagged sequences (CATS) for canine genome mapping. *Journal of Heredity*, 90, 15-26.
- Ma R.Z., van Eijk M.J., Beever J.E., Guérin G., Mummery C.L., Lewin H.A., 1998. Comparative analysis of 82 expressed sequence tags from a cattle ovary cDNA library. *Mammalian Genome*, 9, 545-549.
- McCarthy T.V., Healy J.M., Heffron J.J., Lehane M., Deufel T., Lehmann-Horn F., Farrall M., Johnson K., 1990. Localization of the malignant hyperthermia susceptibility locus to human chromosome 19q12-13.2. *Nature*, 343, 562-564.
- Milan D., Robic A., Chardon P., Iannuccelli N., Caritez J.C., Yerle M., Gellin J., Looft C., Andersson L., Elsen J.M., Le Roy P., 2000. Exemple de cartographie fine : le cas du gène RN chez le porc. *INRA Productions Animales*, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 137-139.
- O'Brien S.J., Womack J.E., Lyons L.A., Moore K.J., Jenkins N.A., Copeland N.G., 1993. Anchored reference loci for comparative genome mapping in mammals. *Nature Genetics*, 3, 103-112.
- Ozawa A., Band M.R., Larson J.H., Donovan J., Green C.A., Womack J.E., Lewin H.A., 2000. Comparative organization of cattle chromosome 5 revealed by comparative mapping by annotation and sequence similarity and radiation hybrid mapping. *Proc. Natl Acad. Sci. USA*, 97, 4150-4155.
- Piumi F., Schibler L., Vaiman D., Oustry A., Cribiu E.P., 1998. Comparative cytogenetic mapping reveals chromosome rearrangements between the X chromosomes of two closely related mammalian species (cattle and goats). *Cytogenetics and Cell Genetics*, 81, 36-41.
- Riquet J., Pitel F., 2000. Les techniques de base de la génétique moléculaire. *INRA Productions Animales*, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 29-35.
- Rogel-Gaillard C., 2000. Les banques de grands fragments d'ADN. *INRA Productions Animales*, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 79-85.
- Schibler L., Vaiman D., Oustry A., Giraud-Delville C., Cribiu E.P., 1998. Comparative gene mapping: a fine-scale survey of chromosome rearrangements between ruminants and humans. *Genome Research*, 8, 901-915.
- Sun H. F., Ernst C.W., Yerle M., Pinton P., Rothschild M.F., Chardon P., Rogel-Gaillard C., Tuggle C.K., 1999. Human chromosome 3 and pig chromosome 13 show complete synteny conservation but extensive gene-order differences. *Cytogenet. Cell Genet.*, 85, 273-8.
- Vignaux F., Priat C., Jouquand S., Hitte C., Jiang Z., Cheron A., Renier C., Andre C., Galibert F., 1999. Toward a dog radiation hybrid map. *Journal of Heredity*, 90, 62-67.
- Yang Y.P., Womack J.E., 1998. Parallel radiation hybrid mapping: a powerful tool for high-resolution genomic comparison. *Genome Research*, 8, 731-736.
- Yerle M., 2000. Etablissement des cartes cytogénétiques et physiques. *INRA Productions Animales*, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 87-93.