

F. HATEY<sup>1</sup>, P. MARTIN<sup>2</sup>, M. DOUAIRE<sup>3</sup>, F. LE GAC<sup>4</sup>,  
G. DAMBRINE<sup>5</sup>, P. HERPIN<sup>6</sup>, P. MONGET<sup>7</sup>.

<sup>1</sup> INRA, Laboratoire de Génétique Cellulaire, BP 27,  
31326 Castanet-Tolosan cedex

<sup>2</sup> INRA, Laboratoire de Génétique Biochimique et de  
Cytogénétique, 78352 Jouy-en-Josas cedex

<sup>3</sup> INRA - ENSAR, UMR Génétique animale, 65 rue  
de Saint Brieuc, 35042 Rennes cedex

<sup>4</sup> INRA, Station commune de Recherches en  
Ichtyophysiologie, Biodiversité et Environnement :  
SCRIBE, Campus de Beaulieu, 35042 Rennes cedex

<sup>5</sup> INRA, Station de Pathologie Aviaire et de  
Parasitologie, 37380 Nouzilly

<sup>6</sup> INRA - ENSAR, UMR Veau et porc, 35590 Saint  
Gilles

<sup>7</sup> INRA - CNRS - Université de Tours, UMR  
Physiologie de la reproduction et des comporte-  
ments, 37380 Nouzilly

## 4 - Recherche de gènes associés à des fonctions

# Le programme Asteroger : vers un outil multifonctionnel pour les productions animales

## La génomique devient fonctionnelle

Les travaux conduits dans le domaine animal à l'INRA sont tournés essentiellement vers la compréhension et la maîtrise des caractères d'intérêt zooteknique en vue de leur amélioration : croissance, reproduction, lactation, comportements, résistance aux maladies, qualité des produits. Pour progresser et prendre également en compte les interactions entre caractères, l'étude la plus exhaustive possible de l'expression des gènes est nécessaire. Or, actuellement, pour étudier ces caractères gouvernés le plus souvent par plusieurs gènes, les techniques utilisées ne permettent d'analyser en même temps que l'expression d'un seul gène ou d'un petit nombre de gènes, parmi les 30 000 à 120 000 présents chez les animaux supérieurs.

Pour cibler les zones du génome où se trouvent les gènes d'intérêt pour un caractère donné, gènes majeurs et locus à caractère quantitatif (QTL : Quantitative Trait Locus), puis pour identifier ces gènes, des approches génétiques se sont développées chez les végétaux et les animaux d'élevage ainsi que chez l'Homme. Ces approches utilisent des cartes du génome, carte physique et carte génétique, établies à l'aide de balises ou marqueurs distribués le plus régulièrement possible sur les chromosomes. La carte physique la plus fine est constituée par la séquence des nucléotides, et le séquençage systématique de quelques génomes modèles et du génome humain a été réalisé ou est en cours. Mais le séquençage brut du génome fournit peu d'informations sur la fonction d'un gène ; ainsi, par exemple, le génome de la levure est intégralement séquencé, mais on ignore tout de la fonction de près de la moitié des gènes. Ce constat a conduit au développement de la génomique fonctionnelle, centrée, par le biais des ARN messagers (transcriptome), sur les gènes exprimés et dont l'objectif est de transformer les informations sur les séquences en données biologiques, en intégrant les données d'expression et de fonction des gènes.

Chez les animaux domestiques, le développement des cartes génétiques, réalisé dans le sillage des travaux de génétique humaine, a atteint et même dépassé aujourd'hui les objectifs que l'INRA s'était assignés au lancement de ce programme en 1991 (Gellin et Grosclaude 1991). Ce succès a confirmé, comme pour le génome humain, la justesse du choix

d'un investissement systématique et de la mise en place de collaborations ; celles que l'INRA a soutenues se sont développées dans le cadre des programmes financés par l'Europe. Pour réaliser ces travaux de cartographie, les outils essentiels pour l'analyse des génomes des principales espèces ont été développés et mis à la disposition des unités de recherche : génotypage à grande échelle (confié à un laboratoire d'analyses génétiques pour les espèces animales : GIE LABOGENA, création et caractérisation de banques de grands fragments d'ADN, panels d'hybrides cellulaires et d'hybrides d'irradiation.

Parallèlement, des travaux sont déjà engagés à l'INRA, grâce à une collaboration entre généticiens, physiologistes, nutritionnistes et pathologistes, pour identifier et répertorier les gènes impliqués dans différentes fonctions ou voies métaboliques : fonction mammaire, fonction ovarienne, métabolisme des lipides, résistance aux agents pathogènes. Par exemple, le projet Roger (pour Répertoire Ordonné de Gènes Exprimés et Régulés) a pour objectif de construire un premier répertoire de transcrits, sous forme de clones d'ADN complémentaire (ADNc) chez les bovins, d'établir les profils transcriptionnels des tissus mammaire, musculaire et embryonnaire et d'effectuer des criblages différentiels pour isoler des gènes dont l'expression est modulée selon le génotype des animaux ou leur état physiologique (ou pathologique). Ce projet est limité à une seule espèce et à un nombre restreint de tissus ou de fonctions.

Dans une perspective de génomique fonctionnelle, il faut maintenant élargir le champ d'analyse au plus grand nombre possible de gènes et à plusieurs espèces. Cela permettra de répondre aux besoins déjà exprimés et de faire progresser de nouveaux domaines de recherche en émergence comme la biologie du développement, le bien-être animal, la réponse immunitaire ou les encéphalopathies transmissibles.

C'est dans cette perspective qu'a été élaboré le programme Asteroger (pour Analyse Systématique des Transcrits en vue de l'Établissement d'un Répertoire Ordonné des Gènes pour l'Étude de leur Régulation). Ce programme associe plusieurs départements du secteur animal de l'INRA et devrait développer des synergies aussi bien dans

l'Institut qu'en dehors. Son objectif est de dresser un inventaire le plus large possible du transcriptome d'un nombre limité d'espèces choisies comme les plus pertinentes : le bovin, le porc, la poule et la truite. L'établissement de ce répertoire passe par la construction de banques d'ADNc pour différents tissus. Cette collection permettra d'identifier les gènes par séquençage et de construire des filtres haute densité ou des microréseaux qui seront exploités par les généticiens, les physiologistes, les pathologistes, les nutritionnistes et les technologues, en particulier pour analyser l'expression des gènes impliqués dans les fonctions qu'ils étudient.

## 1 / Des raisons d'investir

Les raisons d'investir dans le projet Asteroger sont d'ordre stratégique et scientifique.

### 1.1 / L'INRA est en position de force

En dépassant le seul effet de mode, l'INRA se doit de participer au formidable investissement technique, financier et intellectuel que représente la génomique fonctionnelle. Compte tenu du très faible nombre de gènes actuellement caractérisés dans les espèces d'élevage, la disponibilité de banques d'ADNc de ces espèces est de la première importance. Au-delà des seules informations de séquence, la mise en évidence de la (des) fonction(s) précise(s) d'un gène repéré dans un premier temps à partir d'une séquence d'ADN complémentaire constitue un enjeu décisif dans la compétition très vive autour des résultats de la génomique fonctionnelle, notamment en ce qui concerne la propriété intellectuelle des résultats. L'objectif est d'éviter que la totalité des informations de séquence, d'expression et de fonction des gènes des espèces sur lesquelles travaille l'INRA, en partenariat avec les organisations professionnelles, soient inaccessibles ou protégées par des brevets étrangers (anglo-saxons pour la plupart). De fait, les grandes firmes de l'agro-alimentaire et de la sélection animale investissent dans ce domaine. Des projets similaires à Asteroger ont démarré ou vont démarrer à l'étranger. Dans la perspective d'une éventuelle association avec des équipes de recherche étrangères, européennes en particulier, il faut au préalable un engagement fort du côté français.

En France, l'INRA est l'organisme de recherche le mieux placé pour être porteur du projet puisqu'il possède à la fois l'expertise et les ressources nécessaires : expertise en physiologie, en pathologie, en génétique et en biologie moléculaire, ressources avec les animaux nécessaires et les installations expérimentales correspondantes. De plus, l'INRA est déjà engagé dans cette voie à travers des actions institutionnelles (séminaire 'Cartographie du génome et fonction des gènes' 1996, comité 'Post-génomique' 1998), des programmes de recherche (AIP 'Génomiques et fonctions' 1998-99 : programmes Roger, déterminisme génétique de la fonction d'ovulation, ComparEST ; programmes européens Genet-Pig, Chicken Image) ainsi que des réflexions et des contacts individuels dans le cadre français ou européen.

### 1.2 / Identifier tous les gènes

Asteroger est un passage obligé pour que la génomique fonctionnelle soit intégrée dans la démarche quotidienne de nombreux chercheurs car il va les aider à identifier tous les gènes qui interviennent dans les grandes fonctions d'intérêt pour les productions animales. Actuellement, pour une fonction donnée (fonctions gonadique, intestinale, respiratoire, immunitaire, ...), une équipe de recherche de trois à quatre personnes ne travaille «que» sur quelques gènes. À l'échelle de la communauté scientifique, quelques centaines de gènes seulement sont étudiés pour chacune de ces fonctions. La difficulté majeure est de connaître l'importance relative de ces gènes dans la fonction étudiée, par rapport aux quelques dizaines de milliers présents dans le génome. Cette hiérarchisation de l'importance relative des gènes dans chacune des fonctions physiologiques est un enjeu primordial pour les années à venir. Elle requiert que l'ensemble de ces gènes soit connu, c'est-à-dire identifié, séquencé et répertorié et que les données d'expression correspondantes soient disponibles. Il y a, globalement, un fort déficit en ce domaine dans les espèces animales.

L'acquisition de données de séquence dans les quatre espèces est donc un premier objectif. Entreprendre le séquençage systématique du génome des espèces d'élevage n'est sans doute pas aujourd'hui le choix le plus pertinent, compte tenu de l'investissement important que cela représente et de l'absence de recul suffisant dans l'utilisation de ces données. En revanche, la séquence des ARN messagers, qui correspond à la fraction exprimée du génome et ne représente que 3 à 4 % du génome, est, elle, accessible et utilisable rapidement. La connaissance des séquences codantes est également nécessaire pour accompagner le formidable essor des études du protéome : pour identifier une protéine par spectrométrie de masse, il est absolument nécessaire d'en connaître la séquence dans l'espèce étudiée. Si pour les gènes d'intérêt la connaissance de la séquence complète des transcrits est utile, par exemple pour identifier des modules fonctionnels, la stratégie retenue dans le cadre d'un séquençage systématique de banques d'ADN complémentaire est de réaliser une seule lecture de séquençage des extrémités des inserts d'ADNc ; les séquences ainsi obtenues, de 500 à 1 000 bases, constituent des étiquettes (ou EST pour Expressed Sequence Tags) permettant d'identifier les gènes correspondants (Hatey *et al* 1998).

Le projet Asteroger permettra également d'aborder des thématiques jusque là hors de portée, parce que trop difficiles avec les méthodes actuelles, telles que les interactions entre grandes fonctions, par exemple entre nutrition et reproduction. Dans ce domaine, cela fait 30 ans que, faute de mieux, les chercheurs font de la phénoménologie à petite échelle, en dosant dans le sérum quelques substances connues telles que l'insuline, les IGF, les corticoïdes, l'hormone de croissance, les acides gras libres, le glucose, et, récemment, la leptine ; ils calculent ensuite des corrélations toujours difficiles à interpréter entre une baisse de fertilité, par exemple, et les variations des niveaux sériques des dits composés. Les outils développés dans le cadre d'Asteroger permettront de faire un «cliché» de tous les gènes dont l'expression est altérée en cas de désordre métabolique, et de chercher de nouvelles cibles pour agir sur les fonctions. Il faudra

ensuite choisir avec pertinence les gènes sur lesquels étudier prioritairement la fonction, ainsi que leur variabilité dans les espèces zootechniques. Un gène peut être important pour une fonction, mais ne pas être impliqué dans la variabilité du caractère. Les critères de choix devront tenir compte du niveau et surtout de la spécificité tissulaire d'expression, de la présence de polymorphisme dans les espèces modèles (dont l'Homme), et des phénotypes obtenus dans le cadre des grands programmes de mutagenèse systématique des génomes modèles (nématode, drosophile, souris). Cette approche expressionnelle est d'autant plus importante à développer que certaines études de la fonction des gènes sont bloquées par manque de moyens d'investigations tels que les cultures de cellules *in vitro* ou la transgénèse. C'est le cas des études sur l'hypothalamus, compartiment de quelques mm<sup>3</sup>, hétérogène et d'accès difficile, impliqué dans la régulation de l'ingestion, de la reproduction, de la croissance, du métabolisme énergétique ; c'est également le cas de l'épididyme. Dans ces cas, on peut envisager d'exploiter la variabilité phénotypique d'un troupeau à des fins fonctionnelles, en cherchant des corrélations entre le « profil expressionnel » complet (ARN messagers ou protéines) d'un tissu et un phénotype donné. Par exemple, identifier dans l'épididyme des protéines ou des ARN messagers systématiquement associés à une semence de haute qualité ou donnant de bons rendements après congélation permettrait d'une part d'utiliser ces molécules comme marqueurs de fertilité et, d'autre part, d'orienter les travaux sur l'étude de leur rôle fonctionnel.

Plus largement, comparer les « clichés expressionnels » des hypothalamus de truite, de poule, de porc et de bovin suite à une sous-alimentation donnera des éléments de physiologie comparée extrêmement intéressants, étant donné les très grandes différences d'interaction de ces espèces avec leur environnement nutritionnel.

Enfin, au-delà du choix pertinent des quelques gènes à étudier prioritairement, il faut aussi imaginer une approche globale permettant de tirer parti de l'ensemble de ces informations d'expression : cascade d'activation, interactions entre gènes, etc. Une démarche novatrice est ici nécessaire et doit être développée.

### 1.3 / La carte des transcrits

Dans une perspective de génétique, l'isolement et l'identification d'un très grand nombre de gènes attendus du programme Asteroger dans les quatre espèces de référence va permettre d'établir une carte des transcrits plus dense. Secondairement, la redondance résiduelle après normalisation va conduire à l'identification d'un certain nombre de polymorphismes de simples nucléotides (ou SNP pour Single Nucleotide Polymorphism).

Ces résultats vont permettre de progresser dans l'identification des gènes responsables de la variabilité phénotypique, identification qui passe par la construction de cartes génétiques basées sur des marqueurs polymorphes. L'étude de la co-ségrégation d'un allèle de ces marqueurs et de la valeur favorable d'un caractère d'intérêt permet de localiser dans le génome la position du gène responsable du caractère, qu'il s'agisse d'un gène majeur ou d'un gène à effet quantitatif (QTL). Cette démarche est

lourde et nécessite en général de constituer des familles *ad hoc* ('s'insérer' dans un programme QTL existant n'est malheureusement pas toujours possible). Pourtant, la recherche de gènes importants dans les fonctions d'intérêt à travers la mise en évidence de QTL devrait être plus développée ; chez les mammifères, la liste des gènes majeurs tels que RN, Booroola ou culard, plus facilement abordables, sera rapidement épuisée.

La possibilité de densifier la carte en marqueurs polymorphes de type SNP permettra, sans s'affranchir complètement des familles, au moins dans l'immediat, d'en alléger les protocoles. Les informations de séquence issues d'Asteroger devraient contribuer à l'identification de tels marqueurs.

La recherche systématique des régions du génome contrôlant la variabilité des caractères débouche sur la définition de zones plus ou moins larges qui peuvent comporter un grand nombre de gènes qu'il faudra identifier ; ce travail d'identification passe par une localisation de plus en plus précise avec l'utilisation de banques de grands fragments. Cette stratégie de clonage positionnel se trouve considérablement facilitée si les clones qui constituent les banques de grands fragments sont ordonnés les uns par rapport aux autres ; dans cette perspective, il est nécessaire de disposer de ces banques et de réaligner leur mise en ordre ou 'contigage'.

Les collections d'ADNc établies dans le cadre d'Asteroger, sous forme de banques d'ADNc normalisées et organisées ou de microréseaux, constituent des outils efficaces dans cette recherche :

- pour identifier les gènes présents dans les zones d'intérêt par hybridation avec les clones de grands fragments correspondants ;

- pour établir la carte des transcrits, c'est-à-dire positionner sur la carte physique le plus grand nombre possible de gènes. Cette carte va permettre d'identifier rapidement, grâce à leur localisation chromosomique, les gènes présents dans la région où un QTL ou un gène majeur aura été identifié. Cette localisation d'un gène dans une région identifiée par l'approche génétique le désigne comme candidat pour le caractère étudié. C'est la démarche de gène « candidat positionnel » (Ballabio 1993, Collins 1995), beaucoup plus efficace que la stratégie de clonage positionnel. Elle se poursuit par la recherche d'un polymorphisme de ce gène, puis d'une association entre l'un de ses allèles et l'expression du caractère étudié. Une telle association permet d'utiliser ce polymorphisme pour le typage des animaux, c'est-à-dire la détermination de leur génotype au locus considéré : une différence avec des marqueurs anonymes est de pouvoir s'affranchir des phénomènes de recombinaison. Il devient plus facile d'étudier ce qui différencie des animaux extrêmes ou des lignées divergentes. Une telle détermination est plus fiable et plus facile à réaliser que la mesure d'un caractère complexe tel que la tendreté de la viande ou la cinétique *post mortem* du pH dans un muscle. Elle peut, de plus, être réalisée dès la naissance (ou même avant : tri des embryons) sans attendre l'abattage de l'animal ; elle permet aussi de s'affranchir des contraintes liées aux caractères mesurables dans un seul sexe ou d'expression tardive (taux d'ovulation, production laitière). De plus, l'exemple des végétaux montre que l'effet d'un QTL est souvent fonction de la population et/ou du milieu dans lequel il s'exprime. La connaissance des gènes impliqués devrait permettre



de comprendre ce phénomène et de mieux le gérer ou de s'en affranchir. Enfin, sous réserve d'une cohérence entre la fonction (identifiée ou probable) du gène et le caractère, ce gène peut être considéré comme la cause du caractère (mutation causale) ; indépendamment de son intérêt pour la compréhension du caractère, cette mutation constitue le meilleur marqueur pour le génotypage. La démonstration de cette relation de cause à effet est obtenue par manipulation de l'expression du gène (remplacement allélique, inactivation ou surexpression), *in vitro* dans des cellules en culture ou *in vivo* chez l'animal (seulement la souris à l'heure actuelle).

La construction des cartes de transcrits constitue donc un objectif majeur auquel le programme Asteroger va contribuer de manière significative en produisant la séquence des gènes puis en identifiant la localisation de ces gènes. Cette activité se situe donc dans la continuité des efforts entrepris en cartographie. C'est une étape lourde et onéreuse qui passe actuellement par l'utilisation de panels d'hybrides ; elle a déjà permis de localiser près de 30 000 gènes chez l'Homme, dans le cadre d'un consortium international (Schuler *et al* 1996, Deloukas *et al* 1998). Des stratégies alternatives existent (recherche 'in silico', hybridation avec des grands fragments organisés en contigs) ou doivent être imaginées ; les unes et les autres doivent être mises en œuvre pour accélérer l'établissement de cette carte. En effet, il n'y a aujourd'hui qu'environ 600 gènes localisés chez la vache, 450 chez le porc, 200 chez la poule et beaucoup moins chez les poissons domestiques.

Les très nombreuses données recueillies dans le cadre du programme Asteroger vont enfin alimenter les études comparatives entre espèces. Sans oublier la possibilité d'analyse d'un même caractère dans différentes espèces, y compris l'Homme, et l'intérêt des comparaisons entre espèces dans une perspective de phylogénie, nous soulignerons ici les développements à partir des données de séquence ou de cartographie :

- le séquençage systématique d'étiquettes sur les clones ADNc permettra l'identification d'un certain nombre de gènes par comparaison avec les séquences d'autres espèces (gènes orthologues), même si cette approche risque d'être limitée du fait de la divergence entre génomes, surtout dans les régions non codantes (étiquettes 3') ;

- l'organisation des gènes sur les chromosomes (c'est-à-dire leur position les uns par rapport aux autres) présente une certaine conservation entre les espèces, de telle sorte qu'il est possible de déduire la position d'un gène à partir des données disponibles dans une autre espèce comme l'Homme ou la souris (recherche 'in silico'). Cette cartographie comparative est encore très imparfaite et nécessite un nombre plus important de gènes d'ancrage. Il y a, de plus, un intérêt particulier de l'espèce poule du fait de la structure particulière du génome liée à sa plus grande compacité (taille plus petite des introns et des régions intergéniques) et à la présence des microchromosomes dont l'origine et le rôle évolutif éventuel demeurent inconnus.

Pour illustrer ces différents intérêts, nous présentons ici deux exemples dont l'un concerne un tissu d'intérêt agronomique majeur, le muscle, et l'autre une filière de production, la truite.

## 1.4 / Une nouvelle approche de la fonction musculaire

Dans le domaine de la production de viande, la maîtrise et l'amélioration de la qualité passent par l'étude des mécanismes biologiques par lesquels les facteurs génétiques ou les conditions d'élevage affectent le développement des tissus musculaires et adipeux : dynamique du développement (qui conditionne la croissance des animaux et la qualité de leurs carcasses), caractéristiques physico-chimiques, cellulaires et métaboliques (qui conditionnent les qualités diététiques, technologiques et organoleptiques des viandes). La qualité de la viande dépend en effet des influences conjointes de plusieurs caractéristiques complexes (nombre, taille, caractéristiques métaboliques et contractiles des fibres, teneur en lipides intramusculaires, nature du collagène, intensité de la protéolyse, etc) qui peuvent être affectées par de nombreux facteurs génétiques, nutritionnels ou environnementaux.

Le programme Asteroger permettra l'accès simultané à l'ensemble des gènes impliqués dans ces processus et exprimés différemment dans différentes situations, en particulier :

- la mise en place, la croissance et la différenciation des fibres musculaires, qui conditionnent les qualités technologiques de la viande ;

- les différences de teneur en lipides intramusculaires, qui conditionnent en grande partie les qualités sensorielles de la viande ;

- la protéolyse, qui conditionne la maturation de la viande, ou la mise en place des différents types de collagène, qui conditionne la tendreté de la viande ;

- le développement des tissus adipeux organisés, sous-cutané, intermusculaire ou abdominal ;

- les interactions entre génotype et conditions d'élevage : par exemple, chez la poule, croissance rapide ou lente, développement accru des muscles pectoraux, animaux maigres ou gras, variations quantitatives ou qualitatives de l'apport nutritionnel.

Cela permettra de cibler de nouveaux gènes dont la fonction et la régulation par les facteurs d'élevage seront étudiées. De plus, les gènes identifiés sur les espèces modèles (bovin, porc, poule, truite) seront des candidats privilégiés pour l'étude de cette fonction dans les autres espèces (ovin, lapin, dinde, saumon...).

## 1.5 / Une nouvelle approche pour la filière «poissons »

Pour les poissons la situation est particulière car, dans ces espèces :

- les gènes clonés sont encore peu disponibles et les criblages de banque en hétérologue sont extrêmement laborieux et onéreux ;

- il existe des gènes spécifiques permettant l'adaptation des différentes fonctions à des environnements aquatiques variés et éminemment fluctuants et dont l'expression dépend de facteurs comme la température, le taux d'oxygène dans l'eau ou la salinité du milieu ;

- chez les salmonidés, la durée du cycle de production est relativement longue et la maîtrise du temps de croissance, de l'âge à la puberté et du cycle reproducteur constitue un objectif majeur ;

- le transfert des connaissances par le biais de la

génomique comparative est actuellement à peu près impossible avec des espèces trop « éloignées génétiquement » comme les mammifères. L'application des résultats obtenus chez la truite à d'autres poissons modèles ou de grand intérêt aquacole (méda-ka, tilapia, bar), favorisera la diversification des espèces et des systèmes de production (collaborations CNRS, INSERM, CIRAD, IFREMER).

Le programme Asteroger va permettre des progrès dans trois domaines :

1/ les connaissances fondamentales liées au modèle aquatique, par exemple :

- l'identification des gènes médiateurs de l'influence de la température sur le développement musculaire précoce chez le poisson ;

- la recherche de gènes spécifiques des différentes étapes de la spermatogenèse ;

2/ les avancées biotechnologiques, par exemple :

- l'identification de la cascade des gènes impliqués dans la différenciation sexuelle des gonades chez les poissons gonochoriques et les espèces hermaphrodites, différenciation génétiquement déterminée mais aussi dépendante de paramètres environnementaux comme la température, l'alimentation ou des paramètres sociaux. La maîtrise technologique du sexe permettra, chez de nombreuses espèces d'intérêt aquacole, une meilleure gestion de la croissance et de la fertilité des cohortes en élevage ;

- la recherche de gènes différentiellement régulés lors de la réponse précoce non spécifique à une infection virale ou bactérienne ;

3/ les retombées globales sociales, environnementales ou économiques, par exemple :

- l'identification des gènes impliqués dans l'utilisation des glucides comme ressource énergétique majeure, attendue par les nutritionnistes en vue du remplacement de la farine animale par des farines végétales dans les aliments d'élevage. Le programme Asteroger va permettre l'accès à une large panoplie des gènes impliqués dans le métabolisme des glucides et exprimés différemment chez la carpe ('bon' utilisateur des glucides) et la truite ('mauvais' utilisateur des glucides) ;

- la recherche d'indicateurs du bien-être animal : gènes impliqués dans la réponse et l'adaptation au stress chronique, indicateurs du stress chez les poissons d'élevage, marqueurs dans l'analyse des mécanismes généraux d'adaptation au milieu et de domestication des nouvelles espèces d'intérêt pour l'aquaculture.

## 2 / Un grand projet transversal : Asteroger

Dans la perspective de la description du transcriptome complet, c'est-à-dire de l'ensemble des transcrits correspondant à un génome, le projet Asteroger a pour but de concevoir, construire et exploiter un répertoire ordonné de gènes correspondant à une fraction significative (50 000 gènes) du transcriptome pour chacune des principales espèces animales d'élevage (bovin, porc, poule, truite).

Ce projet comprend trois phases :

1/ construction de banques normalisées d'ADNc, multi-tissus (une banque par espèce) ;

2/ séquençage systématique de l'extrémité des clones d'ADNc pour générer des étiquettes (EST) ;

3/ confection et développement des outils moléculaires d'analyse transcriptionnelle (filtres haute densité, microréseaux, biopuces ...).

### 2.1 / Phase 1, construction de banques d'ADNc multi-tissus (1999-2000)

Construire quatre répertoires ordonnés (un par espèce) de gènes transcrits suppose de collecter un grand nombre de séquences (50 000 au minimum, pour chaque espèce, soit 200 000 séquences). La stratégie envisagée est de construire des banques (généothèques) d'ADN complémentaires à partir d'ARNm extraits de différents tissus. Ces banques multi-tissus seront « normalisées » afin d'égaliser la représentation de chaque transcrit, en limitant la redondance et donc le séquençage répété de transcrits identiques. En complément des banques multi-tissus, des banques spécialisées, normalisées ou non, seront construites pour répondre aux besoins spécifiques des équipes concernées (par exemple embryon, glande mammaire, cerveau, foie ...) et pour compléter le répertoire partiel établi à partir des banques multi-tissus. Toutes ces banques seront organisées en plaques de microtitration (96 ou 384 puits). Cette première phase est en cours de réalisation.

### 2.2 / Phase 2, séquençage systématique (démarrage en 2000)

Le séquençage sera réalisé en plusieurs tranches par le Genoscope (Centre National de Séquençage) des laboratoires de l'INRA ou du CNRS ou des compagnies privées. L'objectif est d'obtenir, pour chacune des quatre espèces considérées, les étiquettes 3' et 5' de 50 000 clones différents, ce qui, compte tenu d'une redondance résiduelle après normalisation, nécessitera le séquençage d'un nombre de clones plus important, de l'ordre de 150 000. Le traitement des séquences obtenues ('nettoyage', contrôle et annotation des séquences) et la gestion des informations nécessiteront des moyens informatiques correctement dimensionnés et du personnel qualifié.

Dès ce stade, une première forme d'exploitation des données sera engagée, indépendamment des étapes ultérieures : elle aura pour objectif la localisation des gènes correspondants sur les chromosomes en utilisant les panels d'hybrides. Par contre, le séquençage systématique des clones ne constitue pas un préalable au regard de certaines applications. En particulier, la réalisation des filtres haute densité peut être menée sans attendre, la redondance résiduelle ne constituant pas à ce niveau un handicap majeur ; au contraire, par le biais des répétitions, elle introduit des contrôles internes.

### 2.3 / Phase 3, Filtres haute densité, microréseaux et puces

L'approche filtres haute densité (filtres HD ou macroarrays) et microréseaux (ou microarrays) est basée sur l'hybridation de collections d'ADNc avec des sondes complexes ; la mesure de l'intensité des signaux d'hybridation permet d'apprécier le niveau d'expression des gènes correspondants dans l'échantillon ayant servi à construire la sonde. Cette

méthode d'analyse en parallèle permet d'obtenir simultanément des milliers de données d'expression (Sчена *et al* 1996, DeRisi *et al* 1997, Iyer *et al* 1999).

Les collections d'ADNc à hybrider sont déposées sur un support solide, en Nylon pour les filtres HD ou en verre pour les microréseaux. Les densités varient de 10 à 100 dépôts par cm<sup>2</sup> pour les filtres HD à près de 10 000 pour les microréseaux. La confection de ces dispositifs nécessite le recours à des robots et les collections d'ADNc doivent au préalable être ordonnées.

En plus des robots, l'analyse au moyen des filtres HD et des microréseaux requiert des équipements spécifiques dont l'acquisition doit être envisagée : imageur, stations de travail. Au-delà de l'optimisation et de la standardisation des conditions d'hybridation, des collaborations seront nécessaires pour participer au développement de logiciels d'acquisition et de traitement des données et d'exploitation des résultats par des outils informatiques ; ces développements informatiques sont essentiels pour permettre au projet d'atteindre tous ses objectifs. Nous nous appuyerons pour cela sur l'expérience d'équipes françaises ayant déjà consenti un important investissement en la matière. Les puces à ADN sont basées sur le même principe d'hybridation en parallèle, mais les cibles sont constituées d'oligonucléotides. Produits à l'échelle industrielle, ces dispositifs conviennent parfaitement à des analyses en série, l'une des applications dans notre domaine concernant le génotypage. Des collaborations sont donc également envisagées.

En dehors des équipements, du matériel obtenu (banques) ou construit (filtres HD puis microréseaux) et des données de séquence, ce projet va aussi permettre l'acquisition de savoir-faire (normalisation et soustraction de banques d'ADNc) qui seront ensuite disponibles pour la communauté scientifique de l'INRA. Ces savoir-faire pourront être appliqués à des domaines ou à des projets qui

ne sont pas directement pris en compte par le projet Asteroger ; c'est le cas par exemple des recherches poursuivies dans d'autres espèces (lapin, mouton, canard) ou faisant appel à du matériel biologique peu abondant (embryon précoce, structures cérébrales).

## Conclusion

Le projet Asteroger a pour ambition d'établir, dans quatre espèces d'élevage, une large collection de transcrits, de les identifier par séquençage et d'en tirer parti en termes de localisation chromosomique et d'analyse de l'expression. Ces objectifs répondent à une demande scientifique exprimée par les chercheurs de l'INRA et, au-delà, par la communauté scientifique internationale. C'est un grand projet qui nécessite des moyens génériques disponibles dans le réseau des génopôles français et des moyens disponibles à l'INRA pour les génomes animaux : biologie moléculaire, informatique, unités expérimentales. Ce projet profitera directement à l'ensemble des chercheurs en production animale grâce aux outils et aux savoir-faire élaborés et acquis au cours de sa réalisation. Leur diffusion sera assurée par le caractère transversal du projet et leur exploitation ira au-delà des quelques utilisations évoquées.

Si nous savons déjà cloner, séquencer, localiser et mesurer l'expression des gènes, nous devons développer de nouvelles stratégies pour tirer efficacement parti du flux croissant des données de séquence, de localisation et d'expression ; ces approches 'à haut débit' vont sans aucun doute devenir de plus en plus accessibles et nous permettre d'aborder les questions biologiques sous un angle nouveau. La démarche mise en place dans le projet Asteroger est donc un passage obligé pour collecter et organiser toutes les données indispensables pour l'étude de la fonction des gènes qui sera le défi majeur des prochaines années.

## Références

- Ballabio A., 1993. The rise and fall of positional cloning. *Nature Genetics*, 3, 277-279.
- Collins F.S., 1995. Positional cloning moves from perditional to traditional. *Nature Genetics*, 9, 347-350.
- Deloukas P., Schuler G.D., Gyapay G., Beasley E.M., Soderlund C., Rodriguez-Tome P., Hui L., Matise T.C., McKusick K.B., Beckmann J.S., *et al*, 1998. A physical map of 30,000 human genes. *Science*, 282, 744-746.
- DeRisi J.L., Iyer V.R., Brown P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278, 680-686.
- Gellin J., Grosclaude F., 1991. Analyse du génome des espèces d'élevage : projet d'établissement de la carte génique du porc et des bovins. *INRA Productions Animales*, 4, 97-105.
- Hatey F., Tosser-Klopp G., Cloucard-Martinato C., Mulsant P., Gasser F., 1998. Expressed sequence tags for genes : a review. *Genetics Selection Evolution*, 30, 521-541.
- Iyer V.R., Eisen M.B., Ross D.T., Schuler G., Moore T., Lee J.C.F., Trent J.M., Staudt L.M., Hudson J.Jr., Boguski M.S., *et al*, 1999. The transcriptional program in the response of human fibroblasts to serum. *Science*, 283, 83-87.
- Sचना M., Shalon D., Heller R., Chai A., Brown P.O., Davis R.W., 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 10614-10619.
- Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D., Gyapay G., Rice K., White R.E., Rodriguez-Tome P., Aggarwal A., Bajorek E., *et al*, 1996. A gene map of the human genome. *Science*, 274, 540-546.