

P. MULSANT

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

e-mail : Philippe.Mulsant@toulouse.inra.fr

## 4 - Recherche de gènes associés à des fonctions

# Cartographie fine d'un gène : bilan INRA et autres résultats marquants

**Résumé.** Plusieurs programmes de cartographie fine et d'identification de gènes majeurs sont actuellement en cours à l'INRA, et certains ont abouti (caséine, RN) ou devraient déboucher rapidement sur l'identification de la mutation causale (PIS, Polled, Booroola). Les caractéristiques des principaux programmes sont présentées, ainsi que les raisons de leur progression. Les possibilités d'amélioration à court terme sont discutées.

On peut estimer, avec Georges et Andersson (1996), que la génétique moléculaire s'est réellement développée à partir du moment où la mise en évidence des microsatellites a fourni les marqueurs génétiques polymorphes, nombreux et d'usage relativement simple, qui étaient nécessaires au typage à grande échelle. Au cours de la dernière décennie, l'INRA a développé plusieurs programmes de localisation et d'identification de gènes majeurs, dont l'existence avait été préalablement démontrée par une étude statistique de la distribution des performances dans des pedigrees. Ces gènes avaient été mis en évidence dans la plupart des espèces domestiques (bovin, porc, poule, mouton, chèvre, lapin) et concernent une grande variété de caractères (qualité de la viande ou du lait, résistance aux maladies, efficacité alimentaire, prolificité, déterminisme du sexe, etc). Dans plusieurs cas (PIS, RN et Booroola notamment), le gène pertinent agit comme un facteur clé dans un système biologique encore mal connu (déterminisme du sexe, développement musculaire et qualité de la viande, ovulation), son identification présente donc un intérêt scientifique évident. Trois exemples, ceux des gènes EAC chez les bovins, PIS chez la chèvre et RN chez le porc, sont présentés dans d'autres articles de cet ouvrage. Les stratégies utilisées ne seront pas détaillées ici, mais les caractéristiques principales des programmes en cours seront brièvement présentées, en insistant sur les raisons majeures de succès ou parfois d'échec de ces programmes.

## 1 / Méthodologie

Jusqu'à maintenant, les différents programmes d'identification de gènes menés à l'INRA l'ont été de façon relativement similaire. Ils ont été basés sur des analyses classiques de ségrégation dans des familles informatives, avec leurs trois étapes successives : une première mise en évidence d'un marqueur lié, la cartographie fine de la région chromosomique portant le gène étudié et, enfin, l'inventaire

et l'analyse des séquences codantes de la région (Eggen 2000, cet ouvrage). Pour que ce type d'analyses familiales soit efficace, il est nécessaire de disposer non seulement d'un nombre suffisant de familles et de marqueurs, mais aussi d'une classification satisfaisante du génotype des animaux au locus du gène majeur supposé. Ces conditions sont généralement remplies chez les espèces d'intérêt zootechnique, ce qui explique l'homogénéité des stratégies utilisées. Les principales différences entre les programmes tiennent à l'importance plus ou moins grande qui a été donnée aux différentes méthodes possibles de cartographie fine : suivant les cas, l'accent a été mis sur les hybrides d'irradiation (RN), l'étude détaillée d'un contig de BAC (PIS) ou la cartographie comparée (Booroola).

### 1.1 / Animaux

Le nombre d'animaux informatifs des programmes INRA était en général de quelques centaines (de 300 pour PIS à un millier pour RN), ce qui est largement suffisant pour une première localisation et donne ensuite une précision potentielle de l'analyse de liaison de moins d'un centimorgan. Ces tailles de population peuvent, lors de l'étape de cartographie fine, se révéler un peu faibles pour des gènes situés dans des régions chromosomiques à faible taux de recombinaison (où les distances physiques entre marqueurs sont élevées par rapport à leurs distances génétiques). Même si le nombre d'individus informatifs est généralement moins limitant dans les espèces domestiques que chez l'Homme, le petit nombre d'animaux disponibles a été un des éléments majeurs expliquant l'impossibilité de localiser, jusqu'à présent, le gène présumé d'intersexualité du porc.

### 1.2 / Marqueurs

Il est clair que la situation dans ce domaine a radicalement évolué au cours des années 1990. Si les

premiers programmes (Booroola, par exemple) ont souffert au début d'une certaine rareté des marqueurs disponibles, on dispose maintenant pour la plupart des espèces zootechniques de cartes génétiques relativement détaillées, permettant au minimum une première localisation du gène étudié (de l'ordre de quelques centimorgans). De plus, la réalisation de banques de BAC (Rogel-Gaillard 2000, cet ouvrage) a fourni l'outil « universel » permettant ensuite de densifier la couverture par des marqueurs de la région définie autour du gène d'intérêt (marche sur le chromosome via le « contigage » et recherche de microsatellites dans les BAC isolés). On peut de fait constater que la convergence vers le gène étudié a été relativement rapide pour les principaux programmes au cours des deux dernières années.

### 1.3 / Classification

Il s'agit là d'un paramètre important, parfois sous-estimé, des programmes de recherche de gènes. Les causes d'erreurs de classification sont multiples : erreurs de manipulation, « adoption » d'un animal par une autre mère, pénétrance incomplète du gène étudié, mais aussi hétérogénéité génétique du caractère étudié ou empreinte génétique. Les effets peuvent aller jusqu'à l'impossibilité d'obtenir une localisation du gène (gène Mu affectant la composition corporelle chez le porc). Ils se limitent le plus souvent à une imprécision dans la localisation fine. A titre indicatif, dans la population Booroola, sur 400 méioses informatives, dix correspondent à des doubles recombinaisons n'affectant que le gène *FecB* lui-même, soit 2,5 %. Sans importance réelle si les animaux ne sont pas recombinants, ce type d'erreurs de classification dues à une pénétrance incomplète de la mutation étudiée aboutit à un déplacement de la zone de localisation probable du gène, quand elles affectent un animal porteur d'une recombinaison proche du gène étudié.

## 2 / Situation actuelle

Seuls deux des gènes étudiés à l'INRA ont été identifiés pour le moment : le gène de la caséine alpha-s1 (revue : Grosclaude *et al* 1994), responsable de la qualité fromagère du lait de chèvre, et le gène RN du porc, qui affecte la qualité de la viande (Milan *et al* 2000, cet ouvrage). Dans trois autres cas (Booroola, Polled et PIS), on peut considérer que l'étape de cartographie fine est terminée : isolement de marqueurs ne recombinant pas avec le locus étudié, obtention d'un contig de BAC couvrant la région d'intérêt et, enfin, mise en évidence d'un déséquilibre de liaison. Néanmoins, dans ces trois cas, la comparaison avec les données génétiques humaines ne fournit pas pour le moment de gène « candidat positionnel », ce qui impliquera la recherche puis le test systématiques des séquences codantes présentes dans le ou les BAC définissant la région d'intérêt. Ce travail sera rendu plus délicat par notre connaissance souvent incomplète du site et de la période d'expression du gène en cause. On notera à ce sujet que le clonage positionnel *sensu stricto*, sans gène candidat, n'a abouti, chez les animaux de ferme, que dans le cas du gène RN. Les autres mutations identifiées dans ces espèces modifient un gène « candidat fonctionnel » fort (caséine par exemple) ou un gène « candidat positionnel » (récepteur de la ryanodine pour la sensibilité à l'ha-

lothane chez le porc : Fujii *et al* 1991 ; myostatine pour le phénomène d'hypertrophie musculaire chez les bovins : Grobet *et al* 1997).

## 3 / Perspectives

Quelles sont les perspectives d'amélioration à court terme de ces programmes ou de ceux qui suivront ? Des gains non négligeables sont possibles aux trois étapes de localisation, de cartographie fine et d'identification indiquées plus haut. En ce qui concerne la première localisation, dite grossière, d'un gène, il est envisageable de diminuer fortement le travail de typage en remplaçant dans un premier temps les typages individuels par des typages des « mélanges » ou pools des animaux de phénotype extrême de chaque famille (Tixier-Boichard *et al* 1998). Toutefois, une telle approche paraît difficilement applicable aux espèces peu prolifiques pour lesquelles elle implique probablement des transferts d'embryons. Une autre possibilité est l'analyse directe d'un gène majeur par des méthodes d'identité par descendance (IBD), qui a permis à Charlier *et al* (1996) de localiser génétiquement le gène responsable de la syndactylie chez les bovins. Une telle procédure n'a été que très peu utilisée et son efficacité n'est sans doute pas systématique, en particulier pour des mutations récentes. Par ailleurs, le développement du typage semi-automatique de marqueurs multiples comme les AFLP permet maintenant d'envisager l'analyse de gènes trouvés dans des espèces mal « cartographiées » (caille ou lapin, par exemple).

La cartographie fine devrait quant à elle bénéficier de la réalisation de contigs régionaux de BAC, envisagée dans quelques espèces de référence. L'ordonnement préalable des BAC d'une région permettra en effet de remplacer la marche sur le chromosome, à quoi correspond actuellement le contigage, par des sauts sur le chromosome en choisissant BAC et marqueurs potentiels.

Le dernier point, probablement le plus important, concerne l'identification finale des mutations causales. Comme indiqué ci-dessus, cette identification est actuellement rendue délicate par la rareté relative des gènes « candidats positionnels ». Le séquençage systématique du génome humain devrait bientôt fournir une masse d'informations sur les gènes présents dans une région chromosomique donnée, même si la recherche informatique des séquences codantes est encore très imparfaite (à titre d'exemple, trois des meilleurs logiciels actuels ne proposent aucun gène potentiel commun à partir de la séquence d'un BAC). Combinée avec les résultats du programme Asteroger (identification de très nombreux gènes exprimés dans les principales espèces d'élevage), la disponibilité de cette information de séquence devrait permettre de fortement accélérer la recherche des gènes candidats présents dans une petite région chromosomique.

## Conclusion

Le bilan des différents programmes INRA d'identification de gènes majeurs paraît positif : deux gènes ont été identifiés et trois autres devraient l'être prochainement. A l'avenir, ce type de programme aura, semble-t-il, un concurrent qu'il engendre lui-même : la sélection assistée par mar-

queurs (SAM) et ses variantes (introgression assistée par exemple). Tout programme d'identification a en effet comme résultat intermédiaire l'isolement de marqueurs génétiques proches permettant une sélection assistée par marqueurs efficace. La SAM est d'ailleurs actuellement utilisée pour l'obtention de taureaux reproducteurs homozygotes pour le gène «sans cornes» et, à titre encore expérimental,

pour la création d'un troupeau homozygote pour le gène Booroola en race Romanov. Le choix entre les deux démarches, identification de la mutation causale ou caractérisation indirecte à l'aide de marqueurs liés, devrait dépendre des progrès des méthodes d'identification d'un gène mais aussi de l'intérêt scientifique et/ou agronomique que peut représenter sa connaissance.

## Références

Charlier C., Farnir F., Berzi P., Vanmanshoven P., Brouwers B., Georges M., 1996. IBD mapping of recessive traits in livestock: application to map the bovine syndactyly locus to chromosome 15. *Genome Research*, 6, 580-589.

Eggen A., 2000. Cartographie fine d'un gène et clonage positionnel. INRA Productions Animales, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 133-136.

Fujii J., Otsu K., Zorzato F., de Leon S., Khanna V.K., Weiler J.F., O'Brien P.J., MacLennan D.H., 1991. Identification of a mutation in porcine ryanodine receptor gene associated with malignant hyperthermia. *Science*, 253, 448-451.

Georges M., Andersson L., 1996. Livestock genomics comes of age. *Genome Research*, 6, 907-921.

Grobet L., Martin L.J., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., Dunner S., Ménéssier F., Massabanda J., Fries R., Hanset R., Georges M., 1997. A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nature Genetics*, 17, 71-74.

Grosclaude F., Ricordeau G., Martin P., Remeuf F., Vassal L., Bouillon J., 1994. Du gène au fromage : le polymorphisme de la caséine  $\alpha_{s1}$  caprine, ses effets, son évolution. *INRA Productions Animales*, 7, 3-19.

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.

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.

Tixier-Boichard M., Pitel F., Berge R., Gellin J., Bordas A., Vignal A., 1998. Mapping a major gene affecting skeletal growth in chicken with microsatellite markers and DNA pools. *Animal Genetics*, 29, s73 (abstract).