

La Base de données Mapgena

A. NEAU¹, C. CHEVALET², B. BIBÉ³

¹ INRA, Département de Génétique Animale,
78352 Jouy-en Josas cedex

² INRA, Laboratoire de Génétique Cellulaire,
BP 27, 31326 Castanet-Tolosan cedex

³ INRA, Département de Génétique Animale,
BP 27, 31326 Castanet-Tolosan cedex

e-mail : André.Neau@jouy.inra.fr

Résumé. La base de données Mapgena a un double objectif : regrouper et gérer les informations utiles aux programmes d'étude du génome (cartographie, détection de QTL, analyse de populations) et ses applications ; créer des interfaces entre ces données et les outils d'analyse (crimap, linkage, qtlmap, phylip, génepop, fstat...). Elle est multi-espèce et contient des typages variés (protéines, hémotypes, gènes, microsatellites, RFLP...) ainsi que des caractères divers pour les performances. Cette base de données sous Oracle et son application développée avec l'AGL Uniface sont accessibles sur le serveur dga5 au Centre de Traitement de l'Information Génétique. A ce jour, les programmes de recherches gérés dans Mapgena sont : les programmes de détection de QTL dans plusieurs espèces (bovins laitiers, porcs, poule), le programme international de cartographie génétique chez le cheval ainsi que les données de typage dans le programme de sélection des reproducteurs chez les caprins.

Le début des années 1990 a vu les progrès en génétique moléculaire s'accélérer avec l'arrivée de nouvelles technologies permettant de révéler de nouveaux marqueurs génétiques polymorphes, bien répartis sur l'ensemble du génome et autorisant l'automatisation des génotypages réalisés à grande échelle. Ces nouveaux outils disponibles sont à l'origine de l'essor des programmes de cartographie génétique, puis, un peu plus tard, des programmes de détection des QTL. Ces marqueurs sont aussi utilisés dans le cadre des programmes d'étude de la diversité génétique et pour certaines applications comme le contrôle de filiation et les génotypes pris en compte pour le choix des reproducteurs. Pour l'ensemble de ces programmes, le volume et la complexité des informations à gérer et à traiter nécessitent l'utilisation d'outils informatiques adaptés. Très tôt, une réflexion s'est engagée au sein du Département INRA de Génétique animale sur les besoins en bases de données nécessaires à la mise en oeuvre de ces projets. A côté des bases de données internationales publiques de cartographie génétique comme Bovmap (Eggen *et al* 1993) et de bases locales dans les laboratoires comme Gemma (Iannucelli *et al* 1996), le développement d'une base de données ouverte à la communauté des chercheurs de notre Département devait être décidé. Pour mener à bien cette entreprise, un groupe de projet⁽¹⁾ animé par C. Chevalet fut constitué au début de l'année 1994. Ce texte retrace les diffé-

rentes étapes jalonnant la construction de cet outil avec la définition des objectifs, la description des principaux éléments du dictionnaire des données, le chemin emprunté dans la conception puis le développement de l'application et les programmes de recherches actuellement présents dans Mapgena.

1 / Les objectifs

Le premier objectif de cette base de données est de regrouper, sous forme structurée et normalisée, l'ensemble des informations utiles aux programmes d'étude du génome et de ses applications. Le deuxième objectif est de créer des interfaces entre une sélection de ces données et les outils d'analyse de données sous forme de fichiers conformes aux logiciels d'analyse utilisés. Ceux-ci sont relatifs à l'analyse de liaison des marqueurs, à la détection de QTL, à l'analyse de populations. Un autre besoin exprimé est l'accessibilité de l'application pour l'ensemble des utilisateurs de la base, quel que soit le site où il travaille et quel que soit le matériel informatique dont il dispose (PC, MAC ou terminal X). De surcroît, l'application doit avoir une interface homme/machine graphique conviviale et simple à utiliser. L'ensemble de ces spécifications a été pris en compte dans le choix de l'architecture matérielle et logicielle à mettre en place. Concernant le matériel, un serveur dédié « base de données » a été acheté et mis en place au cours du dernier semestre 1994 au Centre

(1) Le groupe de projet conduit et animé par C. Chevalet a vu sa composition évoluer au cours du temps et des thèmes abordés. La liste indique les participants répartis par unité. Elle se veut la plus exhaustive possible. DGA : B. Bibé, A. Neau et J. Arhainx ; CTIG : G. Steier ; LGBC : H. Levéziel, G. Guérin et E. Cribiu ; LGC : C. Chevalet, D. Milan et M. San Cristobal ; GIE LABOGENA : M.Y. Boscher, Y. Amigues et J.C. Mériaux ; SGQA : D. Boichard, P. Le Roy et J.P. Bidanel ; SAGA : J.M. Elsen, A. Piacère, D. François et J.M. Astruc ; Direction Informatique : C. Christophe.

de Traitement de l'Information Génétique (CTIG). Ce serveur dga3 a été remplacé par un autre serveur dga5 (RS6000 - H50) avec 512 Mo de RAM et 8 disques de 4,5 Go en septembre 1998. Peu après, le système de gestion de bases de données relationnelles (SGBDR) Oracle, cœur du dispositif, a été installé sur cette machine. Toute l'infrastructure informatique était disponible pour héberger le dictionnaire des données.

2 / Le dictionnaire des données

Classiquement dans le cadre d'une base de données relationnelles, ce dictionnaire décrit le modèle conceptuel des données, chaque entité (table), les propriétés (champs ou variables) attachées à chaque entité ainsi que les règles de gestion attribuées à chaque propriété ou entre les entités (relations). Il n'est pas possible de décrire ici de façon exhaustive le contenu de ce dictionnaire, mais on peut le résumer ainsi : il est multi-espèce, accepte des typages de nature variée (protéines, hémotypes, gènes, RFLP, microsatellites, etc) et des critères divers de performances (performances brutes, index ou moyennes). Bien que le modèle de données soit relationnel, il est possible de hiérarchiser la présentation du contenu de ce dictionnaire. Les informations étant relatives à une espèce donnée, l'espèce est d'un point de vue hiérarchique l'entité de plus haut niveau. Sous ce niveau, se trouvent trois grandes catégories de données : les généalogies, les typages et les performances.

2.1 / Les généalogies

Sous cette appellation, plusieurs entités regroupent les informations sur l'animal. L'entité animal résume l'état-civil, le pedigree de l'animal, son origine, sa race. L'entité synonyme permet d'attribuer des codifications connues de l'animal en dehors de celle utilisée (en général, son numéro officiel) pour l'identifier dans la base de données. Ainsi, le numéro d'échantillon d'ADN dans le laboratoire chargé du typage ou le numéro de dossier pour LABOGENA ou encore le numéro IE (Institut de l'Élevage) pour les taureaux d'insémination artificielle, peuvent être des numéros utiles à connaître. Pour des besoins de traitement, un numéro pourra aussi être créé (par exemple en analyse de liaison, pour laquelle le logiciel Crimap ne reconnaît que le numérique pour l'identification des animaux). Tous les synonymes utilisés sont décrits dans une table de référence qui peut être actualisée en fonction des besoins. L'entité race comprend l'information sur la race ou le type génétique de l'animal. L'entité origine comprend l'information sur le lieu d'élevage de l'animal.

2.2 / Les typages

Dans cette catégorie de données figurent l'information sur le locus, son polymorphisme et l'entité typage proprement dite. L'entité locus est identifiée par un code locus et une technique. Cette précision permet de distinguer, pour un même locus, deux échantillons de nature différente (par exemple lait et ADN pour la caséine- α s1). Cette entité contient notamment le numéro de chromosome, son groupe de synténie, son groupe de liaison, sa position physique sur le chromosome et sa position génétique

(distance par rapport à un marqueur de référence sur le chromosome), son type (dominant, codominant), etc. L'entité allèle permet la description du polymorphisme attaché à chaque locus, un numéro permet d'identifier chaque allèle, un code (par exemple, l'allèle numéro 1 peut correspondre au code A, cas des caséines- α s1 ou α s2 pour le typage des laits de chèvres) et une longueur d'allèle sont aussi présents dans cette entité. L'entité typage comprend, pour un animal et un locus donnés, le génotype, l'origine de chaque allèle, le statut du typage après analyse (laboratoire), le statut après traitement des données, le laboratoire qui a effectué le typage (une table de référence décrit les différents laboratoires). Il faut ajouter l'entité géno-phénotypes, qui permet, dans le cas où le marqueur est dominant, d'avoir une correspondance entre le génotype et le phénotype observé.

2.3 / Les performances

Dans cette catégorie de données se trouvent l'information sur le caractère et l'entité performance proprement dite. L'entité type de la performance est un descripteur du caractère attaché à la performance. Cette entité contient le type de la valeur du caractère (continu, discret, binaire), sa nature (index, moyenne, performance brute), l'héritabilité, la répétabilité. L'entité performance comprend, pour un animal et un caractère donnés, la valeur de la performance, la date de la performance, le coefficient de détermination pour les index, le nombre de descendants (index et moyenne).

2.4 / Le sous-ensemble des données

Un des objectifs de l'outil est de réaliser des interfaces entre une sélection des données du référentiel et les différents outils d'analyse des données. Pour répondre à cet objectif, l'entité dossier a été définie. Celle-ci contient un code pour l'identifier, un libellé, le nom du créateur, la date de création ainsi qu'une zone libre de commentaire. On peut associer à chaque dossier un ensemble d'animaux structurés par famille ou population, un ensemble de loci et un ensemble de caractères. Ce dispositif intégré dans le dictionnaire permet ainsi de délimiter et de mémoriser le sous-ensemble des données objet de l'étude pour l'extraction des fichiers d'analyse.

Le dictionnaire étant défini, le modèle conceptuel a été traduit en modèle logique, puis en modèle physique. Ce modèle physique correspond à une suite d'instructions SQL (langage du SGBDR) permettant l'implémentation du dictionnaire de données avec ses règles de gestion dans Oracle. En janvier 1995, la base de données était mise en place.

3 / L'application

3.1 / Présentation de l'AGL et de la configuration retenue

En mai 1995, l'AGL (Atelier de Génie Logiciel) Uniface était choisi. Cet outil est un produit OOO (Outil Orienté Objet) et il permet de faire de la programmation événementielle grâce aux triggers⁽²⁾

(2) Un objet est constitué de propriétés (taille, couleur, etc) mais aussi de codes répartis dans des structures appelées triggers. Un événement « utilisateur », comme le click de la souris sur un bouton ou sur une zone donnée, peut déclencher l'exécution du code contenu dans un ou plusieurs de ces triggers.

associés aux différents objets. Cet AGL est multi-plateforme. En outre, Uniface peut s'interfacer avec un nombre important de SGBDR du marché (Oracle, DB2, Sybase, etc). Ce produit est conçu pour fonctionner dans le cadre d'une architecture centralisée ainsi que dans un mode client/serveur aujourd'hui très répandu. En fait, ce produit est caractérisé par une architecture dite « 3 Tiers » qui distingue trois niveaux : les données prises en charge par le serveur de données (Oracle), le niveau logique de l'application avec ses différents objets (propriétés, règles de gestion et code associé à un traitement) et enfin le niveau utilisateur avec la représentation graphique de l'application. Dans le cadre de notre application, les trois niveaux se trouvent centralisés sur le serveur dga5.

3.2 / Les composants de l'application

L'application comprend six composants principaux dénommés « tables de référence », « généalogies », « typages », « performances », « dossiers » et « outils ». Les cinq premiers correspondent à l'organisation du modèle des données vue sous un angle hiérarchisé. Les « tables de référence » comprennent la gestion des espèces, des races, des types de synonyme, des élevages ou origines des animaux, des laboratoires d'analyse et la gestion des utilisateurs avec leurs droits sur les données. Cette dernière partie n'est accessible qu'à l'administrateur de la base de données. Les « généalogies » contiennent la gestion des animaux et des synonymes. Les « typages » prennent en compte la gestion des loci, des allèles (polymorphisme), des géno-phénotypes et des typages proprement dits. Les 'performances' assurent la gestion des caractères et des performances. Les 'dossiers' sont dédiés à la gestion des dossiers et des listes associées (animaux, loci et caractères). Le dernier composant dénommé 'outils' comprend les fonctions spécifiques de l'application.

3.3 / Les fonctions générales de l'application

Il existe d'abord une fonction classique de consultation et de mise à jour (si l'utilisateur a les droits requis) sous forme de fiche ou de liste pour les éléments constituant les cinq premiers composants. Pour chaque fiche, il y a aussi possibilité de visualiser les objets qui lui sont rattachés. Une autre fonction générale est la possibilité de sélectionner des ensembles de données à partir de la définition d'un profil de recherche établi par l'utilisateur. Pour la définition des profils de recherche, on peut utiliser les opérateurs relationnels ou logiques sur chaque champ, on peut aussi utiliser des jokers comme le « * » ou le « ? » sur les champs correspondant à une chaîne de caractères pour la discrimination des données à sélectionner. Le principe de fonctionnement de la gestion des requêtes est d'abord une analyse de la demande faite par l'utilisateur. L'interprétation de la demande étant réalisée, l'application analyse le contexte de la base de données en établissant des requêtes intermédiaires vers celle-ci. En fonction des réponses, qui peuvent différer d'un jeu de données à l'autre pour une même requête initiale, il y a la construction d'une requête finale la plus optimisée possible vers le serveur de données, qui retourne alors les occurrences dans l'ensemble. Lorsque l'ensemble est constitué, il est possible de le modifier par ajout, suppression ou en le combinant avec d'autres ensembles préalable-

ment mémorisés. On peut aussi faire l'extraction des occurrences de celui-ci dans un fichier en sélectionnant les rubriques (ou champs) et le format des données. Pour les ensembles d'animaux, de loci ou de caractères, il est possible de constituer les listes associées aux dossiers. Avec les trois ensembles précités, il est aussi possible de réaliser des sous-sélections d'ensemble à partir des typages (loci), des performances (caractères) ou des deux (animaux). Cette sous-sélection s'effectue à partir de la définition d'un autre profil de recherche. Le développement de cet outil de requêtes non prédéfinies pour la constitution de ces ensembles de données est donc un outil puissant pour consulter, extraire et manipuler les données de Mapgena.

3.4 / Les fonctions spécifiques de l'application

Les fonctions spécifiques de l'application sont incorporées dans le sixième composant dénommé 'outils'. Il existe une fonction de vérification mendélienne des génotypes pour un dossier. C'est aussi à ce niveau que se trouve les interfaces entre un sous-ensemble de données (dossier) et les outils d'analyse. Ces outils recouvrent l'analyse de liaison, la détection de QTL et l'analyse de populations. En analyse de liaison, nous trouvons l'interface avec Crimap (Green *et al* 1990) et Linkage (Lathrop *et al* 1993). En détection de QTL se trouve l'interface avec QTLmap (Elsen *et al* 1999) pour trois espèces (le bovin, le porc et le poulet) avec une différence dans l'interface utilisateur due au fait que le caryotype représenté (pour le choix des marqueurs) diffère entre ces trois espèces. En analyse de population, nous trouvons l'interface avec Phylip (Felsenstein 1993), Génepop (Raymond et Rousset 1995) et Fstat (Goudet 1995). Les fichiers générés par ces interfaces avec les outils d'analyse peuvent être transférés sur dga2 (serveur dédié au traitement au CTIG). Cette fonction de transfert apporte une cohérence à l'ensemble. Chaque machine, située sur le même site, bénéficie d'un service de haut niveau (sauvegarde, accès sécurisés, compétences regroupées) et de moyens puissants. Il existe aussi des fonctions attachées à des programmes particuliers (cartographie du génome du cheval, génotypages pris en compte dans le choix des reproducteurs chez les caprins).

3.5 / L'étape de développement

Le développement a commencé en début d'année 1996, sur les bases exposées précédemment, avec une version client/serveur d'Uniface (la version AIX-RS6000 n'était pas disponible à l'époque). Les données du gène RN du porc avaient été introduites dans la base de données en cours d'année 1995, avec les généalogies et les performances (6 caractères) fournies par P. Le Roy (Le Roy *et al* 1990) et les typages (marqueurs du chromosome 15) fournis par D. Milan (Milan *et al* 1995). Ces données ont servi de jeu d'essai pour tester les développements. En mars 1996, M.H. Raymond (Direction Informatique, INRA Le Magneraud) nous rejoignait pour assurer une partie du développement. La première version de l'application était disponible en fin d'année 1996. Aujourd'hui, l'application est constituée de 180 fenêtres et le fichier d'export du référentiel de l'application présente une taille de 8 Mo. Ces chiffres donnent un petit aperçu des développements effectués.

4 / Les programmes de recherches

A ce jour, cinq programmes sont gérés dans Mapgena. Il y a trois programmes de détection de QTL (dispositif « petites filles » en bovins laitiers, « Pigmap2 » chez le porc, lignées R+/R- chez le poulet), un programme d'analyse de liaison avec le projet international de cartographie du génome du cheval ainsi que les données de typage dans le cadre des programmes de sélection des boucs d'insémination artificielle et des mères à boucs.

4.1 / Détection de QTL : dispositif « petites filles » en bovins laitiers (Boichard *et al* 2000, cet ouvrage)

Ce programme implique 1568 taureaux (14 pères et 1554 fils). Il y a donc 14 familles (9 Prim'Holstein, 3 Normandes et 2 Monbéliardes) constituées de demi-frères de père. Les marqueurs utilisés sont au nombre de 169 (157 microsattellites, 11 groupes sanguins et le gène BLAD). Au total, le typage de l'ensemble de ces marqueurs sur l'ensemble des animaux concernés, représente environ 259 000 typages réalisés pour l'essentiel au GIE LABOGENA. Les données de performances comprennent des index et des moyennes relatifs aux performances des petites-filles des 14 pères. Elles concernent 11 caractères en rapport avec la quantité et la qualité du lait et représentent environ 17 000 enregistrements de la base.

4.2 / Détection de QTL : programme « Pigmap2 » chez le porc (Bidanel et Milan 2000, cet ouvrage)

Les animaux impliqués dans ce programme sont 924 descendants F2 (LW/MS)/(LW/MS) issus de 6 mâles Large-White et 6 femelles Meishan en F0, ce qui correspond à 6 familles de demi-germains de père ou 24 familles de germains. Ces familles ont été créées et élevées au domaine INRA du Magneraud. Les marqueurs utilisés sont au nombre de 108 et représentent environ 94 000 typages réalisés pour une grande part au GIE LABOGENA. Les données de performances correspondent à 93 caractères : mesures zootechniques, résultats d'analyse biologique, données d'abattage et mesures de comportement. L'ensemble de ces mesures représentent environ 42 000 performances brutes avec leurs facteurs de variation.

4.3 / Détection de QTL : lignées R+ et R- chez le poulet Rhode Island Red (Tixier-Boichard *et al* 1998)

Les animaux impliqués dans ce programme sont 550 descendants F2 (R+/R-)/(R+/R-) issus de 4 mâles et de 39 femelles (R+/R-) en F1, avec 13 coqs et 21 poules R+ ou R- en F0, soit 4 familles de demi-germains de père ou 39 familles de germains. Ces familles d'animaux ont été créées et élevées au Laboratoire de Génétique Factorielle (INRA Jouy-en-Josas), à l'époque, situé à la Minière. Pour les F2, les marqueurs utilisés sont au nombre de 66 (199 marqueurs ont été testés sur les F1) et représentent environ 5 400 typages réalisés au Laboratoire de Génétique Cellulaire (INRA Toulouse). La base contient des données de performances brutes (10 caractères) ainsi que des performances pré-correctées (17 variables) par des effets de milieu en vue de l'analyse QTL. Les mesures mémorisées (crois-

sance, ponte et consommation d'aliment) représentent environ 6 800 enregistrements.

4.4 / Cartographie internationale du génome du cheval (Guérin *et al* 1998)

Après que la décision ait été prise en 1995 de construire la carte « mâle » du cheval, G. Guérin avait proposé que les données attachées à ce programme soient regroupées et traitées à l'INRA. Lors du congrès « International Plant & Animal Genome » de janvier 1997 à San Diego, cette proposition fut acceptée. Vingt-six laboratoires se sont inscrits pour contribuer à la mise en oeuvre de ce programme. A l'origine, 12 familles de référence furent constituées (familles de demi-germains de père) provenant de 10 sites (laboratoires) différents. Une 13^{ème} famille australienne a été ajoutée depuis à ce programme. Ces 13 familles comportent au total 923 chevaux (13 étalons, 407 juments et 503 produits). A partir de fichiers du tableur Excel, outil utilisé par les différents laboratoires participants, nous avons établi un fichier préformaté (proche du format Crimap) pour la saisie des génotypes. Une fonction spécifique d'import des typages à partir de ces fichiers en format « csv » (format variable dans lequel les rubriques sont délimitées par un « ; »), fut développée pour la circonstance. Ces fichiers externes non structurés (au sens classique du terme) sont traités par l'application comme une chaîne de caractères unique qui sera découpée séquentiellement au fur et à mesure du traitement. Le traitement associé à cet import est la vérification mendélienne des typages, la mise à jour des tables marqueurs, allèles et typages ainsi que la création de trois descripteurs en sortie : un fichier avec le journal des opérations effectuées pour assurer la traçabilité, un fichier contenant les génotypes incompatibles et la nature de l'incompatibilité (envoyé au laboratoire) et, enfin, un fichier contenant les génotypes absents et leurs nombres. Cette fonction d'import fut utilisée dès l'arrivée des premiers typages à l'automne 1997. Depuis lors, environ 90 imports ont été effectués, à partir de données venant de plus de 20 laboratoires. La base possède 304 marqueurs qui représentent environ 90000 typages. En sortie, deux fonctions spécifiques ont aussi été développées, l'une extrait un bilan sur l'informativité des marqueurs (méioses informatives par famille et totales, nombre de pères hétérozygotes), l'autre extrait un bilan sur la transmission des allèles (nombre et pourcentage par allèle, χ^2).

4.5 / Une application avec les génotypages en sélection caprine (Manfredi *et al* 1995, Piacère *et al* 1997)

Les données nationales caprines sont hébergées dans la base de données raciale de cette espèce sur le serveur dga1 du CTIG. Cette base de données contient les généalogies et les performances zootechniques de plusieurs centaines de milliers d'animaux. C'est à partir de cette information que l'indexation des reproducteurs s'effectue. Par ailleurs le rôle d'un certain nombre de gènes, comme ceux codant pour les caséines $\alpha s1$ et $\alpha s2$, est maintenant mieux connu pour leurs effets sur la composition du lait des caprins. Certains allèles sont plus ou moins favorables pour la qualité, voire la quantité, du lait produit lors d'une lactation. Depuis 1995, dans le cadre du programme national de sélection pour les races Saanen et Alpine, le typage sur le gène codant

pour la caséine- α 1 est ainsi réalisé régulièrement et pris en compte comme critère dans le choix des boucs d'insémination artificielle et celui des mères à boucs. Il faut ajouter que le GIE LABOGENA effectue également le contrôle de filiation de ces animaux à partir d'un panel sélectionné de 6 microsattellites et retourne à la base caprine les résultats de compatibilité. Concernant la caséine- α 1, le typage se fait sur l'ADN pour les boucs et peut se faire sur le lait et l'ADN pour les mères. Les typages ADN sont pris en charge par Y. Amigues (LABOGENA), les typages lait étant pris en charge par M.F. Mahé (Laboratoire de Génétique Biochimique et de Cytogénétique, INRA Jouy-en-Josas). Même si le polymorphisme diffère entre les deux techniques, des liens étroits existent entre le typage PCR de l'ADN et le typage électrophorétique des protéines du lait. La centralisation de ces génotypes dans Mapgena permet aux différents acteurs d'avoir connaissance de l'information complète. En retour, Mapgena fournira à la base nationale caprine un fichier de génotypes simplifiés ou synthétiques établi à partir de règles régissant la correspondance allélique (ou les incohérences si elles existent) entre les deux techniques d'analyse. Les typages de la caséine- α 2 sur le lait des mères et du microsatellite β pour les boucs et leurs mères sont également insérés dans la base. Environ 5 200 typages sur les protéines du lait et 4 500 typages ADN sont aujourd'hui présents dans Mapgena.

5 / Les perspectives

Les programmes décrits vont se poursuivre et d'autres devraient être mis en place rapidement. Ainsi, les données pour l'analyse de liaison dans le cadre des programmes de recherche de gènes tels que ceux responsables de l'achondroplasie et de l'amyotrophie spinale ou le gène 'sans corne' chez

les bovins, doivent être introduites. En détection de QTL, les informations sur les lignées grasses et maigres chez le poulet, celles sur la caille avec les caractères attachés aux comportements ainsi que celles du dispositif "petites filles" en ovin lait de races Lacaune et Manech tête rousse sont en cours de préparation. La mise en place des informations sur les typages des codons du gène Prn-p (résistance/sensibilité à la tremblante) chez les ovins, est prévue. Dans ce cadre, des développements spécifiques sur la consultation et l'export des haplotypes sont déjà réalisés. L'introduction des données de typage du gène mh (hypertrophie musculaire) ou caractère culard, pour différentes races de bovins viande, est aussi envisagée. Les programmes de recherches sur la biodiversité connaissent des développements importants. Concernant l'application, les interfaces Phylip, Génepop et Fstat sont développés pour générer les fichiers utiles à ces programmes de génétique des populations. On doit noter que les différents programmes de recherches nécessitent de plus en plus une collaboration externe avec des partenaires à l'échelle nationale, européenne ou mondiale. Cette ouverture de la base de données à l'extérieur pose à l'évidence quelques problèmes du point de vue de la sécurité liée à l'accès au réseau du CTIG et de la confidentialité des données de Mapgena. Ces questions ont déjà été débattues et plusieurs solutions ont été proposées. Une réponse, sur le court terme, pourrait consister à mettre en place une base de données miroir sur un serveur placé à l'extérieur du réseau CTIG. Cette base de données externe ne contiendrait que les informations autorisées. Pour conclure, rappelons que cet outil a été conçu à partir d'une réflexion globale, il a montré son efficacité en intégrant des programmes non prévus initialement. C'est grâce à une construction très modulaire que cet outil peut s'adapter rapidement aux nouveaux besoins, qui évoluent très vite dans la discipline de la génétique moléculaire.

Références

- Bidanel J.P., Milan D., 2000. La recherche de QTL à l'aide de marqueurs : résultats chez le porc. INRA Productions Animales, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 223-228.
- Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Milan D., Rupp R., Amigues Y., Boscher M.Y., Levéziel H., 2000. La recherche de QTL à l'aide de marqueurs : résultats chez les bovins laitiers. INRA Productions Animales, numéro hors série « Génétique moléculaire : principes et application aux populations animales », 217-222.
- Eggen A., Christophe C., Papelier S., Piry S., Steinbach D., Neau A., Levéziel H., 1993. Bovmap a database of the bovine genome. CEC Framework 3 Biotechnology, Second European Bovine Genome Mapping Programme (BovMap) meeting, Munich, 24-26 March 1993.
- Elsen J.M., Mangin B., Goffinet B., Boichard D., Le Roy P., 1999. Alternative models for QTL detection in livestock. I. General introduction. Genetics Selection Evolution., 31, 213-224.
- Felsenstein J., 1993. PHYLIP - Phylogenie Inference Package, Version 3.5 Edition, Département of Genetics, Washington University.
- Goudet J., 1995. FSTAT (version 1.2) : A computer program to calculate F-Statistics, Journal of Heredity, 86 : 485-486.
- Green P., Falls K.A., Crooks S., 1990. Documentation for CRI-MAP, version 2.4. Washington University School of Medicine, St. Louis, MO.
- Guérin G., Bailey E., Anderson I., *et al*, 1998. Report of the International Equine Gene Mapping Workshop : First linkage map. XXVI International Conference on Animal Genetics, Auckland, New Zealand, August 9-14, 1998. Animal Genetics, 29, suppl. 1, 42 (abstract).
- Iannuccelli E., Woloszyn N., Arhainx J., Gellin J., Milan D., 1996. GEMMA : a database to manage and automate microsatellite genotyping. XXVth Congress of International Society of Animal Genetics, Tours, 21-25 Juillet 1996. Animal Genetics, 27, suppl. 2, 55.
- Lathrop M., Ott J. *et al*, 1993. User's guide for LINKAGE program (version 5.2, May 31, 1993).
- Le Roy P., Naveau J., Elsen J.M., Sellier P., 1990. Evidence for a new major gene influencing meat quality in pigs. Genetical Research., 55, 33-40.
- Manfredi E., Piacère A., Ricordeau G., Elsen J.M., Bibé B., Grosclaude F., 1995. Polymorphisme de la caséine alpha-s1 et sélection caprine. Rencontres Recherches Ruminants, 2, 167-170.
- Milan D., Le Roy P., Woloszyn N., Caritez J.C., Elsen J.M., Sellier P., Gellin J., 1995. The RN locus for meat quality maps to pig chromosome 15. Genetics Selection Evolution., 27, 195-199.
- Piacère A., Boulloc Duval N., Sigwald J.P., Larzul C., Manfredi E., 1997. Utilisation de l'index combiné caprin et du polymorphisme de la caséine α 1 dans le schéma de sélection caprin. Rencontres Recherches Ruminants, 4, 187-190.
- Raymond M., Rousset F., 1995. GENEPOP (version 1.2) : population genetics software for exact tests and ecumenicism. Journal of Heredity, 86, 248-249.
- Tixier-Boichard M., Pitel F., Bergé R., Gellin J., Bordas A., Vignal A., 1998. Mapping a major gene affecting skeletal growth in chicken with microsatellite markers and DNA pools. XXVI International Conference on Animal Genetics, Auckland, New Zealand, August 9-14, 1998. Animal Genetics, 29, suppl. 1, 73 (abstract).