

Une puce à ADN ciblée sur la région du Complexe Majeur d'Histocompatibilité chez le porc : construction et premières exploitations

Laurence FLORI (1), Christine RENARD (1), Céline URIEN (1), Xiao-Xiang HU (2), Bao-Liang FAN (2), Jérôme LECARDONNEL (1), Céline DUCROIX-CREPY (1), François PIUMI (1), Karine HUGOT (1), Jean-Pierre BIDANEL (3), François LEFEVRE (4), Claire ROGEL-GAILLARD (1), Patrick CHARDON (1)

(1) INRA-CEA UMR 314, Laboratoire de Radiobiologie et d'étude du Génome, 78352 Jouy en Josas

(2) State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing

(3) INRA, Station de Génétique Quantitative et Appliquée, 78352 Jouy-en-Josas

(4) INRA, Unité de Virologie et Immunologie Moléculaires, 78352 Jouy-en-Josas

laurence.flori@jouy.inra.fr

Une puce à ADN ciblée sur la région du Complexe Majeur d'Histocompatibilité chez le porc : construction et premières exploitations

Chez le porc, plusieurs QTL (Quantitative Trait Locus) contrôlant des caractères économiques ont été localisés sur le chromosome 7 en p12-q12, définissant une région candidate à proximité du complexe SLA (Swine Leucocyte Antigen). De nombreux paramètres immunologiques ont aussi été associés à des haplotypes SLA. Le complexe SLA contient des gènes codant pour des molécules impliquées dans la réponse immunitaire innée et adaptative, en particulier les molécules permettant la présentation des antigènes aux lymphocytes T, des gènes codant pour des protéines non immunologiques et des gènes de fonction inconnue. En se basant sur la séquence porcine et la cartographie comparée homme-porc, nous avons développé une puce à ADN ciblée sur un segment chromosomique de 27Mb correspondant à la région candidate. La puce SLA de première génération contient 1856 sondes dont 530 sont situées dans le SLA. Elle couvre 80 % de la région du complexe majeur d'histocompatibilité chez l'homme et 65 % de la région candidate dans son ensemble. Cette puce est utilisée dans deux projets. Le premier projet a pour objectif de rechercher des gènes candidats contrôlant la teneur en lipides du muscle chez des porcs Large White et Meishan. Neuf gènes différemment exprimés ont été mis en évidence. Le second projet concerne l'étude de l'impact de vingt ans de sélection sur la réponse immunitaire des porcs Large White.

A DNA chip targeting the Major Histocompatibility Complex region in the pig: design and first examples of use

In the pig, several QTLs (Quantitative Trait Locus) controlling economical traits have been mapped on chromosome 7 at position p12-q12, which delimits a candidate region close to the SLA (Swine Leucocyte Antigen) complex. Several immunological parameters have been associated with SLA haplotypes. The SLA complex contains genes encoding molecules involved in the innate and adaptative immune response, such as proteins presenting antigens to T cells, genes encoding non immunological proteins and genes of unknown function. Based on the pig sequence and the comparative map between man and pig, we are developing a microarray DNA chip targeting a 27 Mb chromosomal segment corresponding to the candidate region. The SLA chip contains 1856 probes, among which 530 map to the SLA complex. About 80% of the human major histocompatibility complex and 65% of the whole candidate region are covered. This first-generation DNA chip is currently used in two projects. The first project aims at identifying candidate genes involved in the control of intramuscular fat content in Large White versus Meishan pigs. Until now, 9 differentially expressed genes have been identified. In the second project, our objective is to study the impact of the selection during the last 20 years on the immune response capacity of Large White pigs.

INTRODUCTION

Chez le porc, plusieurs caractères d'intérêt économique ont été localisés sur le chromosome 7 en p12-q12, à proximité du complexe majeur d'histocompatibilité (CMH) ou SLA (pour Swine Leucocyte Antigen). Cette région contient, d'une part, des QTL (Quantitative Trait Locus) influençant des caractères de production comme la vitesse de croissance des animaux, l'épaisseur du lard dorsal (Bidanel et al., 2001), la quantité de lipides intramusculaires (Renard et Mourot, 2000), la composition de la carcasse (Milan et al., 2002) ou la qualité de la viande et, d'autre part, des QTL impliqués dans la physiologie de la reproduction comme le nombre de corps jaunes, le poids de l'utérus (Wilkie et al., 1999) ou l'âge de la puberté (Cassady et al., 2001). Pour certains caractères de production, la région impliquée a été récemment réduite à un intervalle de 6cM qui exclut la région SLA au sens strict (Demeure et al., 2005). Des paramètres immunologiques ont également été associés à des haplotypes SLA : taux d'anticorps (IgG), hypersensibilité retardée, phagocytose et activité bactéricide des PBMC (Peripheral Blood Mononuclear Cell) dirigées contre *Salmonella typhimurium* et *Staphylococcus aureus*, activité hémolytique du complément (Vaiman et al., 1998).

Chez les mammifères, la région du CMH est l'une des régions du génome les plus riches en gènes (plusieurs centaines). La moitié de ces gènes codent pour des molécules impliquées dans la réponse immunitaire innée et adaptative avec notamment les molécules permettant la présentation des antigènes aux lymphocytes T. Des gènes codant pour des protéines non immunologiques et des gènes de fonction inconnue sont également localisés dans cette région. Le séquençage et l'annotation du génome humain montrent que les régions flanquant le CMH sont également riches en gènes et la cartographie complète du CMH étendu est maintenant bien connue chez l'homme (Horton et al., 2004). Chez le porc, la séquence complète du complexe SLA (2,7Mb) est terminée et annotée et les cartes RH (hybrides irradiés) et physique de la région flanquante sur le bras q sont bien établies (Barbosa et al., 2004 ; Demeure et al., 2003).

Pour compléter les approches génétiques et aider au clonage positionnel des gènes d'intérêt, se mettent en place des études du transcriptome. Le transcriptome correspond à l'ensemble des ARN messagers (ARNm) présents dans une cellule à un instant donné. Les études du transcriptome permettent d'évaluer la quantité de transcrits présents dans des cellules ou des tissus et, par extrapolation, rendent compte du niveau d'expression des gènes. Elles sont basées sur l'utilisation de réseaux de plusieurs centaines ou milliers de gènes, dont les transcrits sont quantifiés simultanément. Il existe plusieurs types de réseaux : des réseaux pan-génomiques chez les espèces dont les génomes sont entièrement séquencés et des réseaux partiels spécifiques de fonctions physiologiques ou de régions chromosomiques.

Nous développons une approche transcriptome ciblée sur le complexe SLA et les régions flanquantes, définissant une région candidate de 27Mb. Nous présentons ici la construction de cette puce SLA et son utilisation actuelle dans deux

projets de recherche. Le premier projet a pour but de proposer des gènes candidats contrôlant la quantité de lipides intramusculaires chez des porcs Large White et Meishan. Le second projet s'inscrit dans un vaste programme d'évaluation du progrès génétique réalisé en France entre 1977 et 1998 sur des porcs Large White (Tribout et al., 2004) et consiste à étudier l'impact de ces vingt ans de sélection sur la réponse immunitaire des animaux.

1. MATÉRIELS ET MÉTHODES

1.1. Conception et production de la puce SLA

Les puces à ADN sont des supports solides (lames de verre ou membranes de nylon) sur lesquels sont fixés de manière ordonnée plusieurs centaines à plusieurs milliers de séquences d'ADN appelées sondes (produits PCR ou oligonucléotides) correspondant à des gènes dont on souhaite étudier le niveau d'expression. Les ARN (cibles) des cellules ou des tissus sont rétrotranscrits en ADNc (ADN complémentaire), marqués et hybridés sur la puce. Sur lame de verre, deux échantillons marqués par un fluorochrome différent cyanine 3 (Cy3) ou cyanine 5 (Cy5) sont comparés.

1.1.1. Choix des sondes dans la région candidate

La liste des gènes présents dans la région orthologue chez l'homme en 6p21-p23 a été établie en exploitant l'assemblage de la séquence disponible en mai 2004 sur le site UCSC (<http://genome.ucsc.edu/>). Les séquences porcines correspondantes ont été recherchées à l'aide du logiciel de cartographie comparée ICCARE (<http://genopole.toulouse.inra.fr/lccare/>, Muller et al., 2004). Des clones porcins d'ADNc ont été sélectionnés lorsqu'ils étaient disponibles. Ils proviennent de la banque produite par le programme AGE-NAE et de deux banques américaines MARC 1PIG et MARC 2PIG construites aux Etats-Unis par l'United States Department of Agriculture (USDA). Dans le cas contraire, des exons ont été identifiés par comparaison entre les ESTs (Expressed Sequence Tag) et l'ADN génomique et des amorces ont été dessinées à l'aide du logiciel Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Ces exons ont été amplifiés par PCR puis sous-clonés (plasmide PGEMT Easy, Promega).

1.1.2. Production de la puce

La puce SLA a été produite au Centre de Ressources Biologiques GADIE (Génomique des Animaux Domestiques et d'Intérêt Economique, LREG, INRA, Jouy-en-Josas). Les ADNc et les exons sous-clonés ont été amplifiés par PCR (Taq PCR Master Mix Kit, Qiagen) avec des amorces universelles présentes dans le vecteur (35 cycles avec 30 sec à 94°C, 30 sec à 60°C et 2 min à 70°C). Après purification (Multiscreen-PCR plates, Millipore), les produits de PCR ont été contrôlés sur gel d'agarose 1%, quantifiés (Fluoroscanner), évaporés puis repris dans 13µL de SSC 3X. Les sondes ont été déposées sur des lames de verre 25x75 mm (UltraGAPS Coated Slides, Corning) à l'aide d'un robot à 16 aiguilles (Chipwriter Pro, Virtek). Des ADNc du commerce (Lucidea Microarray Scorecard, Amersham Biosciences) ont été dépo-

sés à deux endroits du réseau comme contrôles. Les lames ont été traitées à la vapeur pour homogénéiser l'hydratation des dépôts puis l'ADN a été dénaturé pendant 5 sec à 100°C et fixé aux UV (300mJ). Les lames sont conservées en atmosphère sèche.

1.2. Echantillons et animaux

Dans le premier projet, des porcs Large White sélectionnés pour une faible épaisseur du lard dorsal et des porcs Meishan, caractérisés par une forte adiposité ont été comparés. Des échantillons de muscle longissimus dorsi provenant de 7 porcs Large White de 80 kg et de 5 porcs Meishan de 60 kg ont été utilisés. Pour ces échantillons, le taux d'enzyme malique et la quantité de lipides intramusculaires avaient été mesurés au cours d'une étude antérieure (Renard et Mourot, 2000).

Dans le second projet, qui a pour objet l'étude de l'effet de la sélection au cours des 20 dernières années sur la réponse immunitaire des porcs, des animaux issus de reproducteurs de 1977 n'ayant pas subi le processus de sélection sont comparés à des animaux issus de reproducteurs de 1998. Cette étude s'appuie sur un dispositif expérimental mis en place pour étudier le progrès génétique réalisé entre 1977 et 1998 (Tribout et al., 2004). Des femelles Large White de 1998 ont été, soit inséminées avec de la semence congelée de verrats de 1977 (17 verrats en activité dans les centres d'insémination artificielle en 1978), soit croisées avec des verrats de 1998 (23 verrats d'insémination artificielle). Certains animaux nés dans ces portées (L77 et L98) ont été conservés et accouplés pour produire une seconde génération (G77 et G98). Cinq générations successives (77 et 98) ont ainsi été obtenues. A partir de la deuxième génération, les animaux ont été accouplés en fonction de leur sérotype SLA, dans le but de produire des animaux homozygotes pour l'haplotype H01. Les animaux ont été sérotypés pour les gènes SLA de classe I à l'aide d'un test de cytotoxicité, et génotypés pour les gènes SLA de classe II à l'aide d'un lot de microsatellites.

1.3. Extraction des ARN, hybridation des lames et analyse du transcriptome

Les ARN ont été extraits par la méthode Trizol-chloroforme ou à l'aide du Kit RNeasy (Qiagen), contrôlés et quantifiés à l'aide du Bioanalyser Agilent (Agilent Technologies France) puis stockés à -80°C à une concentration de 1µg/µl.

Pour l'hybridation des lames, les cibles ont été produites par rétrotranscription et marquage direct de 10 µg d'ARN avec incorporation de dCTP couplés à Cy3 ou à Cy5 (Amersham Biosciences). Les cibles marquées ont été purifiées (Kit Qiaquick, Qiagen), quantifiées (Nanodrop, Nyxor Biotech), évaporées puis reprises dans du tampon d'hybridation (Kit Pronto, Corning) à une concentration de 2pmoles/µl. Des ARN contrôles du commerce (Lucidea Universal Scorecard, Amersham Biosciences), s'hybridant avec les témoins déposés sur la lame, ont été marqués en même temps que les cibles. Les lames ont été préhybridées, hybridées avec 20pmoles de chaque cible marquée, puis rincées selon le protocole du fournisseur (Kit Pronto, Corning).

Les lames ont été scannées à l'aide du scanner Chipreader (Virtek) et du logiciel fourni avec le scanner. Deux lasers (un pour chaque couleur) excitent les fluorochromes et les photons émis sont récupérés par un tube photomultiplicateur et convertis en électrons, qui seront à leur tour convertis en signal numérique. Sur les images obtenues, les sondes sur lesquelles s'hybrident les cibles marquées par Cy3 et Cy5 apparaissent en vert et en rouge, respectivement. La superposition de ces deux images permet de visualiser l'expression différentielle des gènes : dans le cas de cibles marquées par Cy3, tous les gènes surexprimés apparaissent en vert et les sous-exprimés en rouge. Les gènes exprimés en quantité équivalente dans les deux conditions étudiées apparaissent en jaune. Les paramètres du scanner ont été définis en utilisant des lames hybridées avec le même ARN marqué avec Cy3 et Cy5 de manière à obtenir des signaux combinés jaunes (Cy3 : puissance 40, sensibilité 600 et gain 10 ; Cy5 : puissance 10, sensibilité 600 et gain 10). Le paramétrage du scanner établi avec cette lame jaune a été conservé pour toutes les autres lames. Les signaux obtenus ont été quantifiés à l'aide du logiciel ImageJ (Biodiscovery).

Les données brutes obtenues après quantification ont été rendues comparables par une étape de normalisation à l'aide du logiciel MIDAS (TIGR, <http://www.tigr.org/software/microarray.shtml>). Les données normalisées ont été analysées pour mettre en évidence les gènes différentiellement exprimés : une surexpression du signal Cy5 correspond à un rapport (signal Cy5/signal Cy3) supérieur à 2 et une sous-expression à un rapport inférieur à 0,5.

2. RÉSULTATS ET DISCUSSION

2.1. Le réseau SLA

Nous avons sélectionné les gènes présents sur la puce à l'intérieur d'une région allant du gène PRL au gène NFYA et correspondant aux pics de liaison obtenus avec les caractères de production (entre les microsatellites SW1369 et S0102). La puce contient 530 dépôts (401 clones d'ADNc et 129 exons sous-clonés), qui couvrent 80 % de la région du CMH chez l'homme et 65 % de la région candidate dans son ensemble. A cette série de gènes, ont été ajoutés 80 gènes de la réponse immunitaire (I. Oswald, communication personnelle) et 80 exons correspondant aux gènes du virus de la maladie d'Aujeszky. Afin d'aider à la normalisation des données, 1166 clones d'ADNc provenant des banques porcines construites et séquencées dans le cadre du programme AGENAE par le Laboratoire de Génétique Cellulaire (INRA, Toulouse) ont été choisis de manière aléatoire et incorporés au réseau. Cette puce SLA de première génération comprend donc un total de 1856 sondes, dont 530 sont spécifiques de la région candidate. Plusieurs lots ont été produits et leur qualité a été validée par des hybridations avec le même ARN marqué par Cy3 et Cy5. La position et l'allure des dépôts témoins ont été vérifiées. Une hétérogénéité de la taille des dépôts est parfois observée mais la qualité des lames semble satisfaisante.

2.2. Recherche de gènes candidats influençant la teneur en lipides du muscle

Après extraction d'ARN du muscle *longissimus dorsi* et vérification de leur qualité, trois animaux de chaque race (Large White et Meishan) ont été comparés. L'analyse des données normalisées a permis d'identifier 9 gènes différemment exprimés entre les deux races : dans la race Meishan, un gène est sous-exprimé et 8 sont surexprimés. Parmi ces gènes, quatre sont localisés dans la région SLA (C6orf25 ; TULP1 ; CRSP3 ; LOC221382). C6orf25 est une protéine transmembranaire de la famille des immunoglobulines qui contient un motif ITIM (immunoreceptor tyrosine-based inhibitory motif) et un domaine SH2. CRSP3 est un cofacteur requis pour l'activation transcriptionnelle de SP1, une protéine qui se fixe sur l'ADN. TULP1 est principalement impliquée chez l'homme dans la dégénérescence de la rétine. Ces gènes sont en cours de validation par PCR quantitative. De nouvelles hybridations sur d'autres animaux de poids différents permettront de confirmer ces premiers résultats.

2.3. Etude de l'impact de la sélection sur la réponse immunitaire

Quinze animaux issus des verrats de 1977 et 15 animaux issus des verrats de 1998 seront comparés. La capacité de réponse immunitaire de ces animaux sera étudiée à partir d'un modèle d'infection *in vitro* des cellules dendritiques avec le virus de la maladie d'Aujeszky, un pathogène majeur bien connu du porc. Chez les porcs âgés de 3 mois, les leucocytes seront extraits du sang. Après isolement, les monocytes seront différenciés *in vitro* en cellules dendritiques immatures à l'aide de cytokines (IL4 et GM-CSF) et infectés par le virus de la maladie d'Aujeszky. Les hybridations permettront d'identifier les gènes cellulaires viro-induits. Les cellules dendritiques sont les premières cellules du système immunitaire à interagir avec le virus et jouent un rôle clef dans l'immunité innée. Ce modèle permettra d'étudier la réponse individuelle à l'infection par le virus. La modélisation expérimentale visant à mettre en évidence les gènes différemment exprimés dans les cellules dendritiques infectées par le virus est en cours d'optimisation.

Des études préliminaires d'infection de la lignée épithéliale rénale PK15 ont mis en évidence des gènes cellulaires différemment exprimés et permis de détecter les sondes virales sur la puce, ce qui montre que les gènes viraux sont induits par l'infection, validant ainsi le modèle d'infection.

CONCLUSION ET PERSPECTIVES

La puce SLA participera à la cartographie des QTL contrôlant les caractères de production identifiés sur le chromosome 7. Elle a déjà permis d'identifier quelques gènes dont l'expression différentielle chez des porcs Meishan et Large White est à confirmer. Cette puce est également en cours d'utilisation dans des études visant à évaluer l'impact de la sélection sur la capacité de réponse immunitaire, avec un modèle d'infection de cellules dendritiques par le virus d'Aujeszky. Ces études transcriptomiques centrées sur une petite région du génome seront complétées par l'utilisation des réseaux porcins pan-génomiques en cours d'élaboration chez le porc, notamment dans le cadre du programme AGE-NAE.

Une approche transcriptome ciblée sur une région candidate permet d'exploiter des informations préalables de cartographie et de saturer la région concernée avec des gènes qui ne sont pas toujours présents dans les réseaux pan-génomiques disponibles. En effet, pour l'espèce porcine, ces réseaux n'offrent actuellement qu'une couverture partielle avec 10000 à 15000 gènes. La puce SLA de première génération comprend des ADNc et des exons amplifiés. L'étape suivante consistera à produire un réseau avec des sondes oligonucléotidiques pouvant distinguer les gènes appartenant à une même famille et ayant de fortes homologies de séquences.

La puce régionale SLA est un outil précieux pour l'étude de l'expression des gènes de la région SLA et de la région SLA étendue. Elle permettra, outre l'identification de gènes d'intérêt, d'appréhender l'expression coordonnée des gènes de la région. L'analyse simultanée du profil transcriptionnel de tous les gènes pourra mettre en évidence des gènes corégulés. La séquence pourra ensuite permettre d'étudier les éléments cis-régulateurs de ces gènes et d'identifier des séquences régulatrices communes.

REMERCIEMENTS

Nous tenons à remercier le personnel de l'élevage de expérimental porcin de Bourges et en particulier Jean Gogué pour la gestion des animaux. Nous remercions également Sophie Pollet et Emmanuelle Zalachas (plateforme PICT, INRA, Jouy en Josas) pour leur aide lors du scan et de la quantification des lames. Le programme portant sur la réponse immunitaire bénéficie d'un financement Genanimal.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Barbosa A., Demeure O., Urien C., Milan D., Chardon P., Renard C., 2004. A physical map of large segments of pig chromosome 7q11-q14: comparative analysis with human chromosome 6p21. *Mamm. Genome*, 15, 982-995.
- Bidanel J.P., Milan D., Iannuccelli N., Amigues Y., Boscher M.Y., Bourgeois F., Caritez J.C., Gruand J., Le Roy P., Lagant H., Quintanilla R., Renard C., Gellin J., Ollivier L., Chevalet C., 2001. Detection of quantitative trait loci for growth and fatness in pigs. *Genet. Sel. Evol.*, 33, 289-309.
- Cassady J.P., Johnson R.K., Pomp D., Rohrer G.A., Van Vleck L.D., Spiegel E.K., Gilson K.M., 2001. Identification of quantitative trait loci affecting reproduction in pigs. *J. Anim. Sci.*, 79, 623-633.
- Demeure O., Renard C., Yerle M., Faraut T., Riquet J., Robic A., Schiex T., Rink A., Milan D., 2003. Rearranged gene order between pig and human in a QTL region on SSC 7. *Mamm. Genome*, 14, 71-80.
- Demeure O., Sanchez M.P., Riquet J., Iannuccelli N., Demars J., Fève K., Kernaleguen L., Gogué J., Billon Y., Caritez J.C., Milan D., Bidanel J.P., 2005. Exclusion of the swine leukocyte antigens as candidate region and reduction of the position interval for the *Sus scrofa* chromosome 7 QTL affecting growth and fatness. *J. Anim. Sci.*, 83, 1979-1987.
- Horton R., Wilming L., Rand V., Lovering R.C., Bruford E.A., Khodiyar V.K., Lush M.J., Povey S., Talbot C.C. Jr, Wright M.W., Wain H.M., Trowsdale J., Ziegler A., Beck S., 2004. Gene map of the extended human MHC. *Nat. Rev. Genet.*, 5, 889-899.
- Milan D., Bidanel J.P., Iannuccelli N., Riquet J., Amigues Y., Gruand J., Le Roy P., Renard C., Chevalet C., 2002. Detection of quantitative trait loci for carcass composition traits in pigs. *Genet. Sel. Evol.*, 34, 705-728.
- Muller C., Denis M., Gentzbittel L., Faraut T., 2004. The Iccare web server: an attempt to merge sequence and mapping information for plant and animal species. *Nucleic Acids Res.*, 32 (Web Server issue):W429-34.
- Renard C., Mourot J., 2000. Exemple d'approche fonctionnelle : le gras intramusculaire chez le porc. *Productions Animales*, numéro hors série « Génétique Moléculaire », 161-163.
- Tribout T., Caritez J.C., Gogué J., Gruand J., Bouffaud M., Billon Y., Péry C., Griffon H., Brenot S., Le Tiran M.H., Bussièrès F., Le Roy P., Bidanel J.P. Estimation, par utilisation de semence congelée du progrès génétique réalisé en France entre 1977 et 1998 dans la race porcine Large White : résultats pour quelques caractères de production et de qualité de tissus gras et maigres. *Journées Rech. Porcine*, 36, 275-282.
- Vaiman M., Chardon P., Rotschild M.F., 1998. Porcine major histocompatibility complex. *Rev. sci. tech. Off. Int. Epiz.*, 17, 95-107
- Wilkie P.J., Paszek A.A., Beattie C.W., Alexander L.J., Wheeler M.B., Schook L.B., 1999. A genomic scan of porcine reproductive traits reveals possible quantitative trait loci (QTLs) for number of corpora lutea. *Mamm. Genome*, 10, 573-578.