

Développement d'un système de traçabilité génétique chez le porc basé sur le séquençage de régions riches en SNP

Bernard CHINA, Antoine CLINQUART, Georges DAUBE

*Université de Liège, Faculté de Médecine Vétérinaire, Département des Sciences des denrées alimentaires,
Sart Tilman B43b, B-4000 Liège, Belgique.*

Avec la collaboration technique de L. DAMS et M.-A. VAESEN

Développement d'un système de traçabilité génétique chez le porc basé sur le séquençage de régions riches en SNP

La traçabilité est un sujet de préoccupation majeur dans l'industrie agro-alimentaire. Les marqueurs génétiques comme les polymorphismes simple nucléotide (SNP) peuvent être utilisés en traçabilité génétique. Deux locus riches en SNP ont été séquencés à partir de 96 individus différents. Trente-deux SNP significatifs ont été observés. Tous les individus testés présentaient un génotype différent. Pour simplifier la procédure, un amplicon hybride réunissant les deux régions d'intérêt a été généré. Une expérience de traçabilité utilisant cette stratégie a été menée avec succès en permettant d'identifier le bon couple « oreille-viande ».

Development of a genetic traceability system in pig based on the sequencing of SNP-rich regions

Traceability is a major concern in food industry. Genetic markers such as single nucleotide polymorphisms (SNPs) can be used in the genetic traceability context. Two loci rich in SNPs were sequenced in 96 different individuals. Thirty-two significant SNPs were observed. All tested individuals presented a different genotype. In order to simplify the procedure, a hybrid amplicon joining the two regions of interest was generated. A traceability experiment based on this strategy was successfully performed allowing to identify the right couple « ear-meat ».

INTRODUCTION

La traçabilité de la viande est une priorité du secteur agro-alimentaire, surtout depuis les crises alimentaires qui ont mis en évidence la nécessité d'une traçabilité sans faille de «l'étable à la table» ou de la «fourche à la fourchette» selon les expressions consacrées. Dans ce contexte, les technologies liées à l'ADN peuvent être utiles. En effet, d'une part, toutes les cellules d'un animal possèdent la même information génétique et, d'autre part, celle-ci est propre à l'individu.

L'ADN possède des régions polymorphes qui servent de marqueurs génétiques (CHINA et al., 2001). Parmi ceux-ci, deux sont particulièrement utilisés en matière de traçabilité génétique : il s'agit des microsatellites et des mutations ponctuelles (ou SNP pour « Single Nucleotide Polymorphism »).

Les microsatellites sont des séquences courtes répétées en tandem et caractérisées par un grand nombre d'allèles (une dizaine). Les allèles diffèrent par le nombre des répétitions. Cependant, la multiplexation et l'interprétation des signaux générés sont difficiles en raison de la présence de bandes ou de pics parasites (« stutters »). Par opposition, les SNP sont le plus souvent bialléliques donnant un signal de type oui ou non. Ceci est plus favorable à une automatisation de la lecture des résultats. Les inconvénients des SNP sont leur faible degré de polymorphisme par rapport aux microsatellites, ce qui implique de tester plus de SNP que de microsatellites pour avoir la même information, mais aussi le caractère population-spécifique des SNP.

Certaines régions d'ADN sont particulièrement riches en SNP (GONG, 2002). C'est le cas des gènes codant pour les protéines du système majeur d'histocompatibilité (ANDO et al., 2001).

Dans cette étude, nous avons génotypé l'ADN de porcs en séquençant deux régions d'ADN riches en SNP : AF355774 (genbank accession number AF355774) et SLADQA (genbank accession number M29939).

1. MATÉRIEL ET MÉTHODES

1.1. Échantillons

Les porcs ont été sélectionnés au hasard dans quatre abattoirs belges. Au niveau génétique, ils ne sont pas de race pure mais résultent de croisements entre les races Landrace, Piétrain et Large-White.

1.2. Extraction d'ADN

Des morceaux de muscles (*Longissimus dorsi*) et d'oreilles furent prélevés sur les carcasses à l'abattoir. L'ADN a été extrait en utilisant une trousse commerciale (wizard SV genomic DNA purification kit, Promega).

1.3. PCR

Le locus AF355774 a été amplifié par les amorces suivantes (GONG, 2002) :

AF355774L 5'-AAGCAAGTGACTCCTCCACCAC-3' et AF355774R 5' TCTCTTCTCCAGCAGACTCCTG. Le locus SLADQA a été amplifié en utilisant les amorces suivantes : SLADQA152U18: 5'-CGACCATGTTGCCTCCTA-3' et SLADQA377L15: 5'-GCCGCGGTGTTGTTG-3'. La réaction d'amplification (PCR) a été réalisée dans un volume de 20 µl contenant 10-50 ng d'ADN génomique, 200 µM dNTPs, 400 pM d'amorces et 1 unité de Taq polymérase (Amersham Biosciences) et son tampon concentré une fois (50 mM KCl, 1,5 mM MgCl₂ et 10 mM Tris-HCl, pH 9). Les cycles suivants (FRED60) furent réalisés : 94°C 5 min, 40 X (94°C 30 sec., 60°C 30 sec., 72°C 1 min), 72°C 4 min dans un thermocycleur (Master cycler gradient, Eppendorf).

Pour l'amplification hybride, le locus AF355774 a été amplifié en utilisant les amorces suivantes : AF355774L et AF-SLADQdwn :

5'-TAGGAGGCAACATGGTCGTCTCTTYCCAGCAGACTCCTG-3' et le locus SLADQA a été amplifié en utilisant les amorces suivantes :

AF-SLAup 5'-CAGGAGTCTGCTGGARAAGAGACGACCATGTTGCCTCCTA et SLADQA377L15 en utilisant les mêmes cycles d'amplification que ci-dessus. Les produits de PCR ont été purifiés en utilisant une trousse commerciale (Fast'n' Easy PCR products clean-up kit, Anansa). Huit microlitres de chaque produit de PCR purifié ont été mélangés avec 200 µM dNTPs, 1 unité de Taq polymérase (Amersham Biosciences) dans son tampon (50 mM KCl, 1,5 mM MgCl₂ and 10 mM Tris-HCl, pH 9) et de l'eau ultra-pure pour un volume total de 20 µl. Les cycles suivants furent appliqués : 94°C 5 min., 40 X (94°C 30 sec., 65°C 30 sec., 72°C 1 min), 72°C 4 min dans un thermocycleur (Master cycler gradient, Eppendorf). Deux microlitres du mélange d'amplification furent utilisés comme ADN cible dans une amplification itérative (PCR) avec les amorces oligonucléotidiques : AF355774L et SLADQA377L15 en utilisant les cycles du programme FRED60. La présence d'un fragment de 680 paires de bases a été vérifiée par une électrophorèse en gel d'agarose 2 %.

1.4. Séquençage

Avant le séquençage, les produits de PCR ont été purifiés en utilisant un système enzymatique (ExoSap-IT, Amersham Biosciences). La réaction de séquençage a été réalisée en utilisant une trousse commerciale (DYEnamic ET terminator cycler sequencing kit, Amersham Biosciences). Chaque brin de l'ADN a été séquençé. Pour les produits d'amplification des locus AF355774 et SLADQA, les amorces de PCR ont été utilisées. Pour l'amplification hybride AF-SLADQA, les amorces oligonucléotidiques

AF355774seq 5'-CTCCTCCACCACCCTGTC-3' et SLADQA373L17 5' -CGCGGTGTTGTTGGAAC-3' ont été utilisées. Le mélange réactionnel était constitué de 7 µl de produit de PCR purifié, 8 µl de « premix » de séquençage, 2,5 µl d'amorce de séquençage (2 pmole µl⁻¹) et 2,5 µl d'eau. Les cycles suivants ont été appliqués : 25 cycles (95°C 20 s, 50°C 15 s, 60°C 1 min) dans un thermocycleur (Master cycler gradient, Eppendorf). Les produits du séquençage ont été purifiés par filtration sur gel (Autoseq96, Amersham Biosciences) et analysés par électrophorèse capillaire (MegaBace 500, Amersham Biosciences) en utilisant les

paramètres suivants : voltage d'injection : 2 kV, temps d'injection des échantillons : 75 s, voltage de migration : 9 kV, temps de migration : 100 min ; PMT1 voltage : 750 V, PMT2 voltage : 750 V, température de migration : 44°C, base caller : Cimarron 3.12. La séquence a été déduite en utilisant le logiciel approprié (Sequence analyzer, Amersham Biosciences). Les séquences ont été alignées en utilisant le logiciel Clustalw software (<http://www.ebi.ac.uk/clustalw>).

2. RÉSULTATS

2.1. Identification de SNP

L'objectif était d'évaluer la possibilité de différencier les individus sur la base de leur génotype. Ce génotype était la combinaison des allèles présents pour chaque SNP situé au niveau des locus AF355774 et SLADQA. En pratique, l'ADN de 96 individus a été isolé. L'ADN des locus choisis a été amplifié par PCR et les produits de PCR ont été séquencés (figure 1). Les séquences ont été alignées et les SNP identifiés. Au niveau du locus SLADQA, 18 SNP ont été identifiés et 14 SNP au niveau du locus AF355774 (figure 1).

2.2. Calcul du polymorphisme

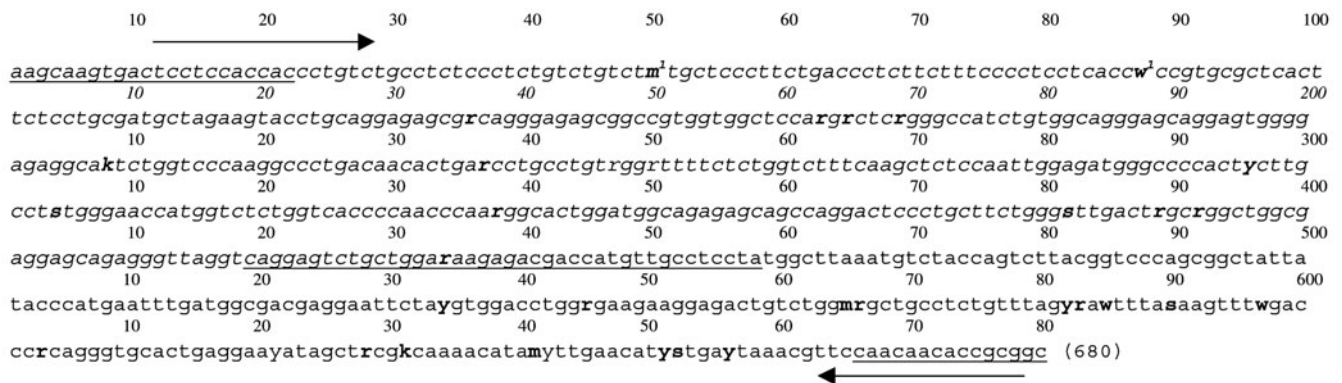
Pour chaque SNP, les fréquences des allèles et l'hétérozygotie ont été calculées (tableau 1).

L'hétérozygotie moyenne était de 0,29. La probabilité qu'un individu A pris au hasard dans une population soit identique à un individu B pris au hasard dans la même population pour r SNP (pour un marqueur i, il y a n allèles possibles avec des fréquences égales à q_{ij}, j=1,2...ni) vaut :

$$P(A = B) = \prod_{i=1}^r (\sum_{j=1}^{n_i} q_{ij}^4 + 4 \sum_{j=1}^{n_i-1} \sum_{k=j+1}^{n_i} q_{ij}^2 q_{ik}^2)$$

(PEELMAN et al., 1998). Cette probabilité est de 1,22 x 10⁻¹⁰ pour les 32 SNP détectés. De plus, tous les génotypes individuels sont différents pour les 96 individus testés.

Puisque les SNP d'un même locus sont proches les uns des autres, la liaison entre les allèles présents pourrait être importante. Nous avons calculé par un test d'indépendance le fait qu'un génotype présent pour un SNP n'avait pas d'influence sur le génotype présent pour un autre SNP situé au niveau du même locus. Les valeurs statistiques (non mon-



La séquence du locus AF355774 est en italique alors que la séquence du locus SLADQA est en format normal. Les SNP sont indiqués en gras. La position des amorces de PCR est soulignée. Les flèches indiquent la position des amorces de séquençage. SNP décrits dans Genbank (accession AF355774) mais peu clair dans notre étude.

Figure 1 - Séquence de l'amplicon hybride.

Tableau 1 - Caractéristiques des SNP

Locus	AF355774																	
	position*	136	163	165	169	208	237	247	250	296	304	338	382	389	392	434		
Mutation	R	R	R	R	K	R	R	R	Y	S	R	S	R	R	R			
p	0,6	0,8	0,71	0,78	0,9	0,81	0,64	0,78	0,63	0,78	0,55	0,8	0,83	0,84	0,75			
q	0,4	0,2	0,29	0,22	0,1	0,19	0,36	0,22	0,37	0,22	0,45	0,2	0,17	0,16	0,25			
H	0,48	0,31	0,42	0,35	0,18	0,31	0,47	0,33	0,46	0,33	0,5	0,32	0,27	0,27	0,38			
Locus	SLADQA																	
	position*	534	545	565	566	582	583	585	590	597	603	621	628	631	641	642	651	652
Mutation	Y	R	M	R	Y	R	W	S	W	R	Y	R	K	M	Y	Y	S	Y
p	0,75	0,94	0,93	0,5	0,64	0,38	0,98	0,98	0,96	0,96	0,62	0,9	0,78	0,89	0,92	0,88	0,81	0,87
q	0,25	0,06	0,07	0,5	0,36	0,62	0,02	0,02	0,04	0,04	0,38	0,1	0,22	0,11	0,08	0,12	0,19	0,13
H	0,38	0,11	0,13	0,5	0,46	0,47	0,03	0,03	0,07	0,07	0,48	0,1	0,34	0,19	0,15	0,22	0,32	0,22

*: dans la molécule d'ADN hybride p = fréquence de l'allèle le plus fréquent q = fréquence de l'allèle le plus rare H = hétérozygotie R = G ou A, K = G ou T, Y = C ou T, S = G ou C, M = A ou C, W = A ou T.

trées) indiquaient que seulement 32 % des combinaisons étaient significativement associées pour le locus SLADQA et 26 % pour le locus AF355774.

2.3. Multiplexation

Dans le but de réduire les coûts d'analyse, une méthode a été développée pour séquencer un seul produit de PCR. Dans un premier temps, les locus AF355774 et SLADQA ont été amplifiés séparément en utilisant des amorces hybrides de manière à créer une zone d'homologie (figure 2). Les produits de PCR ont été mélangés et réamplifiés avec une amorce supérieure correspondant à la séquence 5' de l'amplicon AF355774 et une amorce inférieure correspondant à la séquence 3' de l'amplicon SLADQA. Le fragment de 680 paires de bases ainsi généré a été séquencé. Cette expérience a été réalisée sur 24 ADN précédemment génotypés par séquençage individuel des 2 locus. Les génotypes déduits de l'amplicon hybride étaient les mêmes que ceux déduits à partir des amplicons séparés.

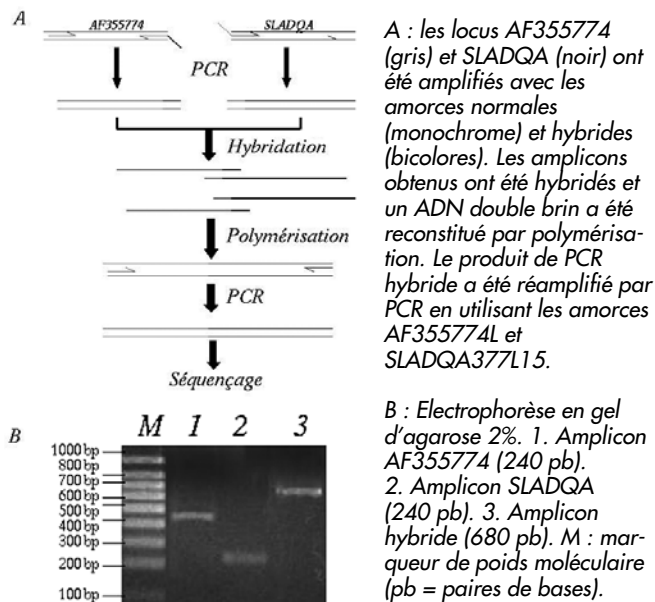


Figure 2 - Construction de l'amplicon hybride PCR

La répétabilité de la méthode a été testée sur 16 ADN en triple. Les génotypes observés étaient les mêmes, indiquant la bonne répétabilité de la méthode basée sur le séquençage de l'amplicon hybride.

2.4. Traçabilité

Enfin, une expérience de traçabilité génétique a été menée. Un morceau de viande et un morceau d'oreille ont été prélevés sur le même animal. En parallèle, un morceau d'oreille a

été prélevé à partir de 6 autres porcs. L'ADN génomique a été extrait à partir des 8 échantillons de tissus. Les génotypes établis par séquençage de l'amplicon hybride ont été comparés. Les résultats ont indiqué que le bon couple oreille-viande a pu être identifié.

3. DISCUSSION - CONCLUSION

La capacité à tracer la viande de la ferme à l'étable est une préoccupation majeure du secteur agro-alimentaire. La traçabilité classique est basée sur des données administratives comprenant des documents ainsi que des boucles d'oreille (EVRARD et al., 2001). Ce type de traçabilité est efficace jusqu'à l'abattoir mais, ensuite, la trace est plus difficile à suivre en raison de la constitution de lots. Puisque la séquence d'ADN présente dans chaque cellule nucléée d'un individu est identique et que (si on excepte les vrais jumeaux) cette séquence est spécifique de l'individu, la traçabilité génétique pourrait venir au secours de la traçabilité administrative. L'estimation du polymorphisme d'une séquence d'ADN est basée sur l'étude des marqueurs génétiques comme les microsatellites ou les SNP. Les microsatellites sont plus polymorphes, mais le patron obtenu est plus difficile à interpréter alors que les SNP sont moins informatifs mais le signal obtenu est plus facile à analyser.

Les présents résultats indiquent qu'il est possible d'obtenir un génotype spécifique d'un individu en séquençant deux régions d'ADN riches en SNP (AF355774 et SLADQA). De plus, la possibilité d'obtenir un amplicon hybride regroupant les deux séquences permet d'obtenir le génotype par séquençage d'un seul produit de PCR. Cette approche pourrait être utile dans le contexte de la traçabilité car il est possible de comparer les génotypes obtenus à partir de tissus différents du même animal. Néanmoins, la méthode pourrait être améliorée en utilisant un logiciel informatique permettant de déduire automatiquement le génotype à partir de la séquence. En effet, à l'heure actuelle, cette étape est encore manuelle et requiert beaucoup de temps et de concentration. Cependant, le séquençage permet de détecter de nombreux SNP en une seule réaction, ce qui réduit les coûts d'analyse par rapport aux autres méthodes de génotypage actuellement disponibles (primer extension, OLA) (CHINA et al., 2001).

REMERCIEMENTS

Ce travail est soutenu financièrement par la Direction Générale des Technologies, de la Recherche et de l'Energie (DGTRE) du Ministère de la Région wallonne (Conventions n° EP1A320501R0092/9914330 et n° 114880). Nous remercions le Dr F. GOFFAUX pour sa lecture attentive et critique du manuscrit.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ANDO A., KAWATA H., MURAKAMI T., SHIGENARI A., SHIINA T., SADA M., TSUJI T., TORIU A., NAKANISHI Y., MITSUHASHI T., SEKIKAWA K., INOKO H., 2001. *Anim. Genet.*, 32, 73-77.
- CHINA B., EVRARD V., NOIRFALISE R., CLINQUART A., DAUBE G., 2001. *Ann. Méd. Vét.*, 145, 15-24.
- EVRARD V., CHINA B., NOIRFALISE R., DAUBE G., CLINQUART A., 2001. *Ann. Méd. Vét.*, 145, 5-12.
- GONG Y., 2002. *Anim. Genet.*, 33, 80-81.
- PEELMAN L.J., MORTIAUX F., VAN ZEVEREN A., DANSERCOER A., MOMMENS G., COOPMAN F., BOUQUET Y., BURNY A., RENAVILLE R., PORTETELLE D., 1998. *Anim. Genet.*, 29, 161-167.