Utilisation du transcriptome pour préciser la localisation d'un QTL responsable de la variabilité d'engraissement chez le poulet de chair

Guillaume Le Mignon, Frederique Pitel, Colette Désert, Sophie Leroux, Olivier Demeure, Grégory Guernec, Madeleine Douaire, Pascale Le Roy, Sandrine Lagarrigue

▶ **To cite this version:**

Guillaume Le Mignon, Frederique Pitel, Colette Désert, Sophie Leroux, Olivier Demeure, et al.. Utilisation du transcriptome pour préciser la localisation d'un QTL responsable de la variabilité d'engraissement chez le poulet de chair. Huitièmes Journées de la Recherche Avicole, Mar 2009, Paris (FR), France. hal-00730108

HAL Id: hal-00730108

https://institut-agro-rennes-angers.hal.science/hal-00730108

Submitted on 4 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UTILISATION DU TRANSCRIPTOME POUR PRECISER LA LOCALISATION D'UN QTL RESPONSABLE DE LA VARIABILITE D'ENGRAISSEMENT CHEZ LE POULET DE CHAIR

Le Mignon Guillaume[a,d], Pitel Frédérique[b], Desert Colette [a], Leroux Sophie[b], Demeure Olivier[a], Guernec Gregory[c], Douaire Madeleine[a], Le Roy Pascale[a] , Lagarrigue Sandrine[a]

[a]UMR INRA Agrocampus Génétique Animale 35042 RENNES, [b]INRA Laboratoire de Génétique Cellulaire 31326 AUZEVILLE, [c]INRA UR1037 SCRIBE IFR140 Ouest-Genopole 35000 RENNES, [d] ITAVI 28 rue du rocher 75008 PARIS

RESUME

Alors que de nombreux QTL pour différents caractères ont été cartographiés dans les espèces d'élevage, l'identification des mutations causales reste encore laborieuse.
Nous proposons dans cette étude de combiner analyse de liaison et données transcriptomiques pour cartographier plus finement une région QTL de l'engraissement chez le poulet de chair préalablement détecté sur le chromosome 5. Trois stratégies ont été utilisées. Ces trois stratégies permettent d'affiner l'intervalle de localisation de ce QTL initialement de 160-191 cM (7Mb) à 182-185 cM (0,6Mb) en accord avec l'intervalle de localisation de 165-184 cM préalablement détecté dans un dispositif F2 de plus de 1200 individus. Le nombre de gènes candidats localisés dans cet intervalle est maintenant réduit de 118 à 19 gènes. Nos résultats démontrent la faisabilité et l'intérêt de la stratégie de génomique génétique chez les animaux d'élevage.

ABSTRACT

Although many QTL for various traits have been mapped in livestock, identification of causative mutations remains difficult. The aim of this study was therefore to combine transcriptome data and linkage analysis to refine a chicken QTL region with a significant effect on abdominal fatness (AF) previously detected on chromosome 5 (GGA5). Three approaches were used. These three approaches made it possible to refine the GGA5 AF QTL from 160-191cM (7 Mb) to a most probable position confidence interval of 182-185 cM (0.6 Mb) in accord with the 165-184 cM confidence interval for GGA5 AF QTL previously detected in F2 design of more than 1200 birds. Thus the number of positional candidate genes is now reduced from 118 to 19 genes. Our results show the feasibility and the interest of a "genetical genomics" strategy in livestock species.

INTRODUCTION

Malgré le succès de la détection de région QTL ces 20 dernières années chez les animaux d'élevage, les intervalles de localisation restent encore très larges pouvant contenir une centaine de gènes. Il est donc aujourd'hui encore très difficile d'identifier les gènes responsables de la variabilité d'un caractère.

Dans les espèces modèles (souris, levure…), plusieurs équipes de recherche ont récemment proposé de coupler des programmes de détection de QTL avec les données transcriptomiques pour identifier les gènes candidats (Ghazalpour *et al.*, 2006; Schadt *et al.*, 2003). Cette nouvelle approche également appelée « genetical genomics » considère l'expression d'un gène comme une variable quantitative et utilise les marqueurs génétiques pour identifier des régions génomiques régulant le niveau d'expression de ces gènes. Dans ces travaux, nous nous proposons d'utiliser cette méthodologie dans le but d'affiner la localisation de QTL dans une espèce d'élevage, le poulet, pour laquelle la structure de population et la densité de marqueurs sont moins idéales que dans les espèces modèles. Bien que de nombreuses régions QTL aient été identifiées chez le poulet, aucune mutation causale n'a encore été clairement identifiée excepté les récents travaux d'une équipe française (URA-INRA-Tours E. Lebihan-Duval *et al.,* communication personnelle). Nous décrivons ici trois approches (résumées en figure 1) qui utilisent les données transcriptomiques dans le but de réduire l'intervalle de confiance d'un QTL responsable de la variabilité d'engraissement abdominal préalablement détecté sur le chromosome 5 avec une forte significativité ($p<0,01$) et un large effet (1,02 écart type phénotypique). Les approches 1 et 2 ont déjà été décrites dans la littérature (Lan *et al.*, 2006; Ponsuksili *et al.*, 2005; Schadt *et al.*, 2003) alors que la troisième n'a encore jamais fait l'objet de publication.

1. MATERIELS ET METHODES

1.1. Animaux et procédures expérimentales

Un individu mâle hétérozygote pour le QTL d'intérêt sur le chromosome 5 est croisé avec 8 femelles d'une lignée maigre pour le poids de gras abdominal pour générer 71 mâles back-cross. La présence du QTL est confirmée après une analyse de liaison sur le caractère d'intérêt sur ces descendants. A 4 semaines, des échantillons de sang sont prélevés pour une extraction d'ADN en vu d'un génotypage. Après le sacrifice des animaux à 9 semaines, les foies sont collectés, rapidement congelés en azote liquide puis stockés à -80°C avant une extraction des ARN pour les analyses transcriptomiques. Après éviscération, les carcasses sont stockées toute la nuit à 4°C puis découpées et mesurés pour le poids de gras abdominal.

1.2. Extraction de l'ADN et génotypage des marqueurs

L'ADN génomique est extrait à partir des échantillons de sang (100µl) pour les 71 mâles selon la méthode phénol/chloroforme modifiée par Assaf *et al.* (Assaf *et al.*, 2003). L'ADN est ensuite quantifié par densité optique puis dilué à 10ng/µl. Dix marqueurs sont ensuite génotypés sur le chromosome 5. Les amplifications PCR sont réalisées dans un mélange de 15 µl contenant 25 ng d'ADN génomique, 0,2 µM de chaque amorce, 0,5 U de Taq polymerase, 20 mM de Tris-HCl pH 8,4, 50 mM de KCl, 2 mM de MgCl2, 0,2 mM de dNTP. Après un cycle de dénaturation à 94°C pendant 5 minutes, 30 cycles d'amplification ont été effectués (dénaturation : 94°C, 30 s ; hybridation : 55°C ou 58°C en fonction des marqueurs, 30 s ; élongation : 72°C, 30 s). La taille des fragments amplifiés a été analysée sur un séquenceur ABI 3100 (Perkin-Elmer) à l'aide des logiciels GeneScan 3.7 et Genotyper 3.6 (PE Biosystems).

1.3. Extraction de l'ARN

L'ARN total est extrait avec la méthode trizol (Invitrogen, Cergy Pontoise, France) en accord avec les instructions du fournisseur. La qualité et la concentration des ARN extraits sont ensuite estimées à l'aide du 2100 Bioanalyzer (Agilent Technologies, Massy, France).

1.4. Procédures appliquées aux puces oligonucléotides

- ***Marquage des ARN et hybridation***. Cinq µg de chaque échantillon d'ARNm est rétro-transcrit puis marqué avec un fluorochrome Cy5 en utilisant le protocole de marquage ChipShot™ Direct Labeling kit (Promega, Charbonnieres, France). Chaque échantillon d'ARNm marqué en Cy5 est ensuite hybridé sur une puce de 20461 oligonucléotides de poulet (collection EADGENE) parallèlement à un échantillon référence marqué en Cy3 (mélange d'ARNm de l'ensemble des échantillons de foie prélevés) en accord avec la procédure de la plateforme d'hybridation Genopole "Toulouse Midi-Pyrénées" (France). Nous n'avons retenu pour les analyses transcriptomiques que 46 animaux, les 20 animaux les plus extrêmes pour le niveau de caractère et 6 animaux intermédiaires. Le ratio de fluorescence pour chaque gène reflète l'abondance relative des ARNm pour chaque échantillon par rapport au niveau d'ARN de la référence. La référence permet ainsi de prendre en compte toutes interactions possibles entre les spots et les puces.
- ***Acquisition des données***. Les signaux fluorescents sont détectés à l'aide d'un scanner

(GenePix 4000A, Axon Instrument, CA). Les images sont alors analysées avec le logiciel GenepixPro 4.0 (Axon instruments, Inc., Union City, CA).

- Normalisation des données. Pour l'étape de normalisation, les données sont filtrées selon trois critères : i) le critère genepix effectué par GenepixPro 4.0 qui détecte automatiquement les spots dits anormaux, ii) le critère SNR qui estime l'intensité du signal par rapport au bruit de fond local, et iii) le critère d'asymétrie qui détecte les spots asymétriques. Sur la base des spots précédemment sélectionnés, les valeurs log2 du ratio Cy5/Cy3 sont alors normalisées par régression non linéaire (Lowess fitness) avec la procédure LIMMA de l'environnement Bioconductor. L'hypothèse utilisée lors de cette normalisation est que la majorité des expressions géniques ne diffère pas entre deux échantillons.

1.5. Détection de QTL et eQTL

Les valeurs de gras abdominal des animaux backcross sont ajustées en fonction des effets lot ou mère par une analyse de covariance, incluant le poids vif des animaux en covariable (procédure GLM de SAS). Pour les analyses eQTL, aucun ajustement des niveaux d'expression de l'ARNm des gènes n'est effectué pour les effets lots ou mères à cause de l'effectif réduit de la population. Le logiciel QTLMAP est utilisé pour la détection de QTL et des eQTL. Les seuils de signification sont empiriquement calculés après 2000 simulations des données.

2. RESULTATS

Les stratégies pour affiner l'intervalle de confiance du QTL pour l'engraissement sur GGA5 sont résumées sur la figure 1.

- Sélection des 693 gènes corrélés au caractère

Une analyse de corrélation entre les niveaux d'expression de gènes et le caractère révèle 693 gènes significativement corrélés ($p< 0.05$).

- Approche 1: Décomposition du caractère engraissement à partir des 693 profils transcriptomiques

Cette approche propose de "disséquer" le caractère d'engraissement en séparant les individus selon leurs profils transcriptomiques pour les 693 gènes. Une double classification ascendante hiérarchique est ainsi réalisée pour les 46 animaux en fonction des niveaux transcriptionnels de ces gènes. Nous pouvons clairement identifier 4 sous groupes d'animaux. Une analyse QTL réalisée sur uniquement 3 de ces sous groupes (35 animaux) permet d'améliorer très nettement la détection du QTL ($p<0,001$). Le nouvel intervalle de confiance est indiqué sur le tableau 1.

-Approche 2: Sélection des gènes corrélés au caractère et régulé par une région eQTL co-localisante avec la région QTL gras de GGA5

Parmi les 693 gènes sélectionnés pour leur corrélation avec le caractère, 91 gènes présentent une région eQTL dans l'intervalle de localisation du QTL gras sur le chromosome 5 (GGA5) (160-191 cM). Nous pouvons observer une surreprésentation de gènes ayant un eQTL dans la région d'intérêt parmi les gènes corrélés au caractère.

- Sélection des gènes possédant un eQTL proche de la position la plus probable du QTL gras. Afin de sélectionner les gènes régulés par un eQTL proche de la position la plus vraisemblable du QTL, 91 nouvelles détections QTL ont été réalisées sur les niveaux d'engraissement résiduels corrigés pour chaque niveau d'expression de ces 91 gènes. Pour un gène donné, la diminution importante du maximum de vraisemblance du QTL gras est signe i) d'une covariance entre le niveau du transcrit et le niveau du caractère élevée et/ou ii) d'une proximité entre les maximum de vraisemblance de l'eQTL et du QTL importante. Finalement, 37 gènes parmi les 91 gènes permettent une correction efficace du QTL gras.

- Analyse QTL combinant les 37 niveaux d'expression de gène et le caractère. Nous avons ensuite généré une nouvelle variable combinant les niveaux d'expression de ces 37 gènes. Déjà utilisée dans le même contexte par Lan *et al* (2003), l'analyse en composante principale (ACP) a été utilisée pour réduire la dimension de l'ensemble des données expressionnelles et ainsi créer cette nouvelle variable. Cette variable est la combinaison linéaire des 37 gènes pondérés par leurs coordonnées sur le premier axe de l'ACP, axe qui discrimine bien les descendants gras des maigres. Une nouvelle analyse QTL utilisant une analyse discriminante entre la variable combinée et le caractère gras pour chaque position du chromosome 5 est réalisée. Indiqué sur le tableau 1, nous détectons un QTL pour ces 37 variables combinés au gras à 180 cM ($p<0,01$). Le maximum de vraisemblance est considérablement augmenté (19,73) comparé au caractère engraissement seul (10,2) pour un seuil de signification identique. Le nouvel intervalle de confiance est ainsi réduit comme indiqué sur le tableau 1.

- Approche 3: Prédiction de l'allèle paternel Q versus q à la mutation causale chez les animaux recombinants

Dans un premier temps, nous avons déterminé sur la base des allèles aux marqueurs transmis par le père, l'haplotype paternel Q (versus q) sur l'ensemble de l'intervalle de localisation du QTL gras reçu par chaque descendant. Nous pouvons ainsi déterminer avec certitude l'haplotype paternel de 25 descendants : 11 q et 14 Q. A partir

des profils transcriptomiques des 37 gènes précédemment sélectionnés, nous avons cherché à distinguer par une analyse discriminante les deux groupes d'animaux Q versus q. Le modèle généré a alors été utilisé pour prédire l'allèle paternel Q versus q à la mutation causale reçu pour chacun des 6 animaux recombinants dans l'intervalle du QTL. Après avoir déterminé les points de recombinaison dans l'intervalle de localisation du QTL chez ces animaux recombinants, grâce au développement de 4 nouveaux marqueurs dans la région, l'intervalle de localisation pour le QTL gras a été encore réduit comme indiqué sur le tableau 1.

CONCLUSION

Les approches 1 et 2 fondées sur l'analyse de liaison classique mais ici utilisant l'information transcriptomique permettent une réduction significative mais modérée cependant de la région QTL (31cM versus 14cM pour l'approche 1 et de 27cM pour l'approche 2). Par contre la troisième approche fondée sur la prédiction chez les recombinants de l'allèle paternel au QTL couplée avec la détermination des points de recombinaison permet une réduction beaucoup plus drastique de la région d'intérêt (5cM). En prenant l'intersection des intervalles de confiance des régions QTL / eQTL obtenus par les 3 approches décrites ci-dessus appliquées sur une population de 46 animaux, la localisation la plus probable de la mutation causale responsable de notre caractère serait comprise dans une région de 3cM (182-185). Auparavant, une détection QTL classique réalisée sur 1200 individus nous avait permis d'identifier une région QTL de 31 cM.

Nos résultats montrent l'intérêt des approches de « génomique génétique » dans le contexte de réduction de régions QTL responsable de la variabilité de caractères chez les animaux d'élevage, première étape nécessaire pour identifier le polymorphisme causal sous-jacent.

REFERENCES BIBLIOGRAPHIQUES

Assaf S., Hazard D., Pitel F., Morisson M., Alizadeh M., Gondret F., Diot C., Vignal A., Douaire M., and Lagarrigue S. (2003). *Poult Sci* **82:** 54-61.

Ghazalpour A., Doss S., Zhang B., Wang S., Plaisier C., Castellanos R., Brozell A., Schadt E. E., Drake T. A., Lusis A. J., and Horvath S. (2006). *PLoS Genet* **2:** e130.

Lan H., Chen M., Flowers J. B., Yandell B. S., Stapleton D. S., Mata C. M., Mui E. T., Flowers M. T., Schueler K. L., Manly K. F., Williams R. W., Kendziorski C., and Attie A. D. (2006). *PLoS Genet* **2:** e6.

Ponsuksili S., Murani E., Schellander K., Schwerin M., and Wimmers K. (2005). *Biochim Biophys Acta* **1730:** 31-40.

Schadt E. E., Monks S. A., Drake T. A., Lusis A. J., Che N., Colinayo V., Ruff T. G., Milligan S. B., Lamb J. R., Cavet G., Linsley P. S., Mao M., Stoughton R. B., and Friend S. H. (2003). *Nature* **422:** 297-302.

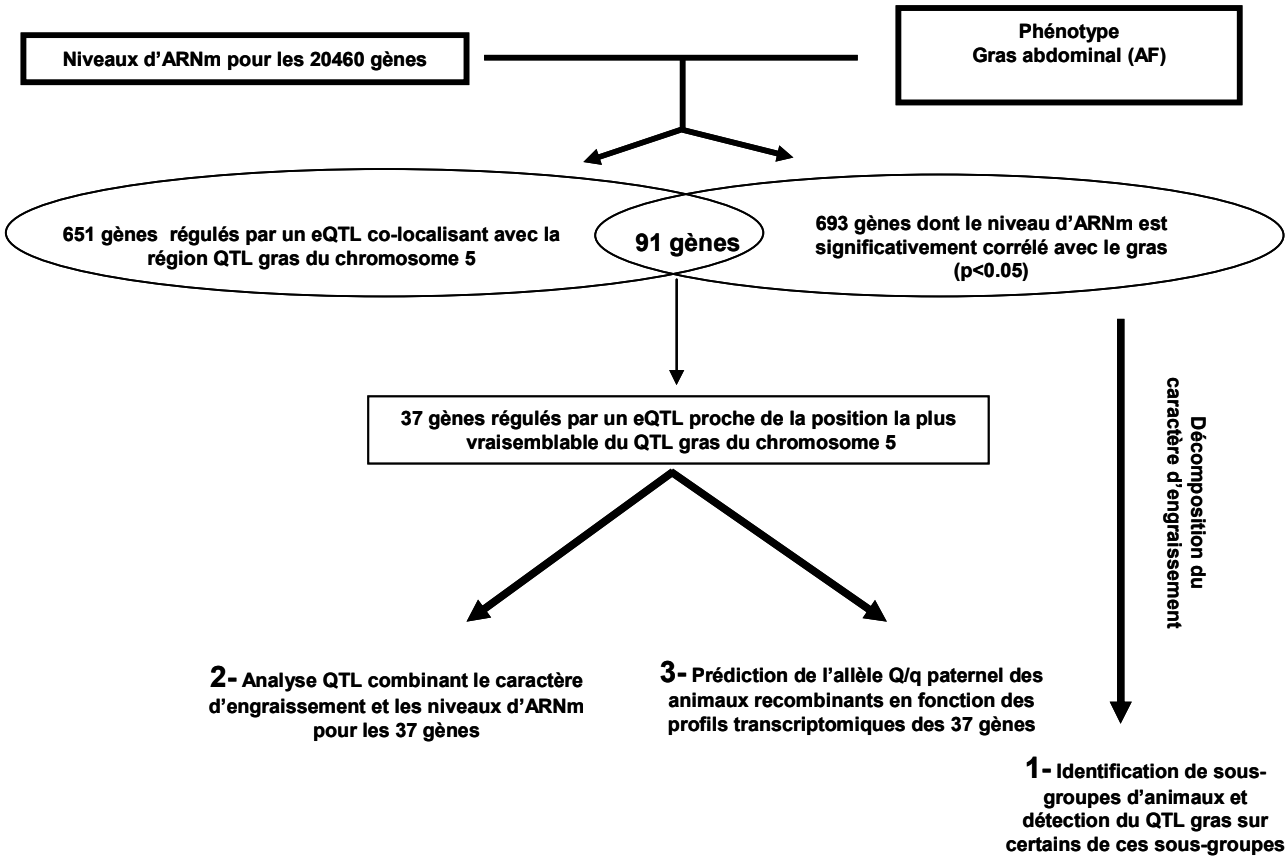

Figure 1. Synthèse des 3 approches réalisées dans le but de réduire l'intervalle de localisation du QTL pour l'engraissement localisé sur le chromosome 5.

Position de l'intervalle de localisation du QTL GGA5 selon les approches utilisées	Position la plus probable (cM)	Significativité (*:p<0.05; **:p<0.01)	Intervalle de localisation (cM)	Intervalle de localisation (Mb)	Nombre de gènes présence dans l'intervalle de localisation
Intervalle de localisation sans utilisation des approches transcriptomiques	179	*	**160-191 (31)**	**52,5-59,5 (7)**	**118**
Intervalle de localisation avec l'approche 1	180	**	171-185 (14)	55,5-57,5 (2)	38
Intervalle de localisation avec l'approche 2	177	**	155-182 (27)	51-57 (6)	95
Intervalle de localisation avec l'approche 3	/	/	182-187 (5)	56,9-57,8 (0,9)	24
Intervalle de localisation regroupant les 3 approches	/	/	**182-185 (3)**	**56,9-57,5 (0,6)**	**19**

Tableau 1. Réduction de l'intervalle de localisation du QTL pour l'engraissement localisé sur le chromosome 5 par trois approches utilisant des données transcriptomiques. La base de données ncbi nous permet de relever le nombre de gènes présents dans chaque intervalle de localisation.