

ANALYSE DE TRANSCRIPTOMES DE FOIES DE CANARDS PAR SEQUENÇAGE D'ARN (RNA-SEQ)

Diot Christian¹, Houée-Bigot Magalie¹, Demeure Olivier¹, Baéza Elisabeth², Vignal Alain³, Pitel Frédérique³, Marie-Etancelin Christel⁴, Robert-Granié Christèle⁴, Molette Caroline^{3,5}, Bouchez Olivier⁶, Esquerré Diane⁶, Marsaud Nathalie⁶, Klopp Christophe^{6,7}, Peterlongo Pierre⁸ et Lemaitre Claire⁸

¹INRA-Agrocampus Ouest, UMR1348 PEGASE 35042 RENNES

²INRA, UR83 URA 37380 NOUZILLY

³INRA-ENVIT, UMR444 LGC 31326 CASTANET-TOLOSAN

⁴INRA UR631 SAGA 31326 CASTANET-TOLOSAN

⁵INRA-ENSAT-ENVIT, UMR1289 TANDEM 31326 CASTANET-TOLOSAN

⁶INRA, GeT-PlaGe Genotoul 31326 CASTANET-TOLOSAN

⁷INRA, UR875 SIGENAE 31326 CASTANET-TOLOSAN

⁸INRIA Rennes - Bretagne Atlantique, Irisa, EPI GenScale, 35000 RENNES

christian.diot@rennes.inra.fr

RÉSUMÉ

La connaissance et la maîtrise des mécanismes mis en œuvre lors de la stéatose hépatique sont des préalables essentiels à la définition de prédicteurs précoces expliquant une part de la variabilité des caractères exprimés chez le mulard. C'est dans ce contexte que le projet DLiverS a été entrepris. Il visait en premier lieu à analyser les transcriptomes hépatiques de canards Pékin, de Barbarie et de leurs hybrides (mulard et hinny), nourris *ad libitum* ou gavés afin de mettre en évidence les gènes présentant des différences d'expression. Le second objectif visait à identifier l'origine parentale (Pékin ou Barbarie) des ARNm exprimés chez les mulards et hinnys et appréhender ainsi les mécanismes sous-tendant les effets d'hétérosis observés. Les ARN ont été séquencés à haut débit sur la plateforme GeT-PlaGe de Genotoul. Les assemblages, annotations et comptages des séquences ont été réalisés par l'équipe bioinformatique Sigenae. Les analyses de transcriptomes et de polymorphismes sont en cours. Elles vont permettre d'identifier les gènes différemment exprimés et leur origine parentale. Au final, ce projet devrait donc améliorer la connaissance des gènes impliqués dans le développement de la stéatose hépatique du canard mulard et préciser les origines parentales des ARNm exprimés. Ces résultats devraient trouver des applications intéressantes pour la filière.

ABSTRACT

Transcriptome analyses of duck livers by high throughput RNA sequencing

Characterization of the mechanisms involved in the development of hepatic steatosis is a prerequisite to the definition of predictive markers explaining part of the variability of liver traits in mule ducks. The DLiverS project was developed for this aim with two major objectives: analyze hepatic transcriptomes from Pekin, Muscovy, mule and hinny ducks fed *ad-libitum* or overfed to highlight differentially expressed genes and identify genome specific expression of the genes in mules and hinnys in order to decipher heterosis effects observed in these hybrids. Liver RNAs were sequenced using high throughput technology on GeT-PlaGe Genotoul facilities. Assemblies, annotations and counting were conducted by Sigenae bioinformatics' team.

Transcriptomics analyses and characterization of polymorphisms are underway. The first results allow identifying differentially expressed genes and their parental origins. The analyses of this huge amount of data must be further conducted before to definitely conclude. However, we expect that the results should greatly improve our knowledge of the mechanisms and genes involved in the development of steatosis and their parental origin. They could also constitute the first milestones of genetic improvement of duck liver production.

INTRODUCTION

Les canards sont des espèces de rente importantes en France, notamment au regard de la production du foie gras qui représente plus de 75 % du total mondial. En fait, 95% du foie gras produit provient de mâles mulards gavés, hybrides résultant du croisement intergénérique entre des femelles communes (*Anas platyrhynchos*) telles que les canes Pékin et des mâles Barbarie (*Cairina moschata*). La sélection directe de caractères du canard mulard est coûteuse puisque nécessitant la procréation de canetons, leur gavage et abattage. De plus, certaines mesures de qualité des foies (notamment, l'évaluation du rendement technologique à la cuisson) sont destructives. La filière est demandeuse de prédicteurs précoces, marqueurs ou gènes, chez les espèces parentales expliquant une part de la variabilité des caractères exprimés chez le canard mulard. La connaissance et la maîtrise génétique des mécanismes mis en œuvre lors de la stéatose hépatique est donc nécessaire. Quelques travaux ont été entrepris pour améliorer la connaissance des mécanismes impliqués dans le développement de la stéatose hépatique du canard : estimation des paramètres génétiques relatifs à l'aptitude au gavage et à la qualité et composition des foies gras (Marie-Etancelin *et al.*, 2011), recherche de QTL (Kileh-Wais *et al.*, 2013), séquençage d'ARN à moyen débit (Pitel *et al.*, 2010) et expression de quelques gènes candidats (Hérault *et al.*, 2008, 2010).

Notre projet, nommé DLiverS, vise à compléter les résultats obtenus par séquençage à haut débit des ARN (RNA-Seq) de foies de canards Pékin et de Barbarie et de leurs hybrides réciproques, mulard et hinny, nourris *ad libitum* ou gavés. Il comporte deux objectifs majeurs : 1) Analyser les transcriptomes hépatiques afin de mettre en évidence des différences d'expression entre génotypes et/ou statuts nutritionnels et ainsi mieux comprendre les mécanismes impliqués dans le développement de la stéatose hépatique ; 2) Identifier l'origine parentale (Pékin ou Barbarie) des ARN exprimés chez les mulards et les hinnys, ces derniers résultant du croisement de mâles Pékin et de femelles Barbarie. Ces hybrides réciproques sont en effet d'excellents modèles pour l'étude des effets d'hétérosis observés notamment dans la production de foie gras, le canard mulard ayant une aptitude à la production de foie gras similaire à celle du canard de Barbarie et supérieure à celle observée chez les canards Pékin et hinny. Comme ces mécanismes pourraient impliquer des phénomènes d'empreinte parentale, décrite chez les mammifères, et sachant que l'existence de tels phénomènes chez les oiseaux est controversée, étudier l'origine parentale des gènes exprimés chez les hybrides devrait apporter de nouveaux

éclairages sur la compréhension de ces mécanismes.

Enfin, sachant l'importance de la stéatose hépatique dans les syndromes métaboliques chez l'homme, il n'est pas exclu que les résultats obtenus puissent trouver des applications en santé humaine.

1. MATERIELS ET METHODES

1.1. Dispositif expérimental et séquençages

Quatre-vingt échantillons de foies de 4 génotypes différents (Pékin, Barbarie, mulards et hinny), gavés ou non (soit 10 échantillons par condition) et de pédigrées connus (Chartrin *et al.*, 2006) ont été utilisés dans cette étude. Les ARN totaux ont été extraits de ces échantillons selon le protocole déjà décrit (Hérault *et al.*, 2008) à l'aide de kits NucleoSpin® RNA L (Macherey-Nagel EURL).

La préparation des banques à partir des ARN totaux et les séquençages « paired-ends » en multiplex de 6 ont été réalisés sur la plateforme Génome et Transcriptome - PlaGe de la Génompolé Toulousaine (<http://get.genotoul.fr>) avec la technologie HiSeq 2000 et à l'aide du kit TruSeq v3 (Illumina), nécessitant 14 lignes de séquençage (figure 1).

1.2. Analyses des données

Les travaux d'alignement des séquences et de découverte des transcrits et des gènes ont été réalisés par l'équipe Sigeneae en utilisant deux séquences de référence, celle encore grossière du canard commun ou celle du poulet.

L'analyse des SNP est en cours, réalisée à l'aide de deux outils. Le premier a été développé dans le cadre d'un programme analysant l'empreinte chez le poulet (voir la communication de Frésard *et al.*, JRA-JRPFG2013/000144). Les scripts (perl et python), écrits en collaboration avec l'équipe Sigeneae, permettent d'identifier des SNP à expression biaisée et de réaliser les analyses statistiques des résultats. Le deuxième outil, KisSnp, permet de trouver des SNPs en se passant d'assemblage ou de génome de référence (Peterlongo *et al.*, 2010). Il prend en entrée 1, 2 ou n jeux de données NGS (fasta ou fastq) et propose en sortie l'ensemble des couples de séquences différenciées par une unique substitution. Chacune de ces séquences est alors fournie avec des informations concernant la qualité moyenne des séquences et leur couverture dans chacun des jeux de données d'entrée.

2. RESULTATS ET DISCUSSION

Comme le montre la figure 2, environ 340 millions de séquences ont été obtenues par ligne de séquençage, soit une moyenne de 56,7 millions de séquences par échantillon. Au total, il s'agit donc de 4,760 milliards de séquences qui ont été produites dans le cadre de ce projet. Ces données sont gérées sous NG6, environnement permettant le stockage et la mise à disposition des données (Mariette *et al.*, 2012). Plus de 70% de ces séquences s'alignent sur le transcriptome de référence « Duck V4 s.ap.4 » réalisé par Sigenae (<http://www.sigenae.org/index.php?id=197>).

Des travaux préliminaires, menés sur quelques échantillons séquencés, ont permis d'identifier quelques SNP spécifiques d'espèces selon l'approche décrite par Pastinen (2010) (figure 3). Les travaux visent maintenant à optimiser les outils, tant en terme informatique que statistique, afin d'identifier les SNP spécifiques des génomes

parentaux et de quantifier leur abondance au sein des ARN exprimés par les hybrides.

CONCLUSION

Une fois les analyses élargies à l'ensemble des séquences obtenues, il est raisonnable de penser que nous aurons des informations particulièrement utiles à l'identification des gènes présentant des différences d'expression entre génotypes et/ou statuts nutritionnels et de l'origine parentale des gènes (différentiellement) exprimés chez les hybrides.

Au-delà de leur portée cognitive tout à fait intéressante, ces résultats pourraient aussi constituer les premiers jalons d'une sélection précoce, chez les parents du mulard, de canards gras présentant des caractères intéressant la profession.

REFERENCES BIBLIOGRAPHIQUES

- Chartrin P., Bernadet M., Guy G., Mourot J., Hocquette J., Rideau N., Duclos M.J., Baéza E., 2006. Comp. Biochem. Physiol. A, (145), 390-396.
- Hérault F., Robert E., Diot C., 2008. Anim. Genet., (39), 325-327.
- Hérault F., Saez G., Robert E., Al Mohammad A., Davail S., Chartrin P., Baéza E., Diot C., 2010. Anim. Genet., (41), 12-20.
- Kileh-Wais M., Elsen J.M., Vignal A., Feves K., Vignoles F., Fernandez X., Manse H., Davail S., André J.M., Bastianelli D., Bonnal L., Filangi O., Baéza E., Guéméné D., Genêt C., Bernadet M.D., Dubos F., Marie-Etancelin C., 2013. J. Anim. Sci. Sous presse
- Marie-Etancelin C., Basso B., Davail S., Gontier K., Fernandez X., Vitezica Z.G., Bastianelli D., Baéza E., Bernadet M.D., Guy G., Brun J.M., Legarra A., 2011. J. Anim. Sci., (89), 669-679.
- Mariette J., Escudié F., Allias N., Salin G., Noirot C., Thomas S., Klopp C., 2012. BMC Genomics, (13), 462.
- Peterlongo P., Schnel N., Pisanti N., Sagot M.F., Lacroix V., 2010. Proceedings of String Processing and Information Retrieval (SPIRE) 2010, Springer LNCS (6393), 147-158.
- Pastinen T. 2010. Nat. Rev. Genet., (11), 533-538.
- Pitel F., Leroux S., Fève K., Vignoles F., Duby C., Bernadet M.D., Marie-Etancelin C., Basso B., Klopp C., Vignal A., Diot C., 2010. 9èmes Journées de la Recherche Filière Palmipèdes à Foie Gras, Bordeaux, France, 07-08 octobre 2010.

Figure 1. Schéma de séquençage.

Les ARN provenant des foies de canards des 4 génotypes (jaune : Pékin ; rouge : Barbarie ; orange : mulard ; saumon clair : hinny), gavés (G, vert vif) ou non (NG, vert clair), ont été répartis de manière aléatoire avant d'être étiquetés (Tag1 à Tag6) et répartis dans les lignes de séquençage (L1 à L14).

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14
Tag1	1 G	82 NG	3 G	84 NG	2 G	81 NG	4 G	83 NG	25 G	106 NG	27 G	108 NG	37 G	119 NG
Tag2	89 NG	5 G	86 NG	7 G	85 NG	8 G	87 NG	6 G	107 NG	28 G	105 NG	26 G	118 NG	40 G
Tag3	11 G	92 NG	9 G	90 NG	12 G	91 NG	10 G	89 NG	30 G	109 NG	32 G	111 NG	39 G	
Tag4	94 NG	15 G	96 NG	13 G	95 NG	14 G	93 NG	16 G	112 NG	31 G	110 NG	29 G	120 NG	
Tag5	18 G	99 NG	20 G	97 NG	19 G	98 NG	17 G	100 NG	35 G	116 NG	33 G	114 NG	38 G	
Tag6	101 NG	22 G	103 NG	24 G	102 NG	21 G	104 NG	23 G	113 NG	34 G	115 NG	36 G	117 NG	

Figure 2. Nombre de séquences obtenues par ligne de séquençage (valeurs exprimées en millions)

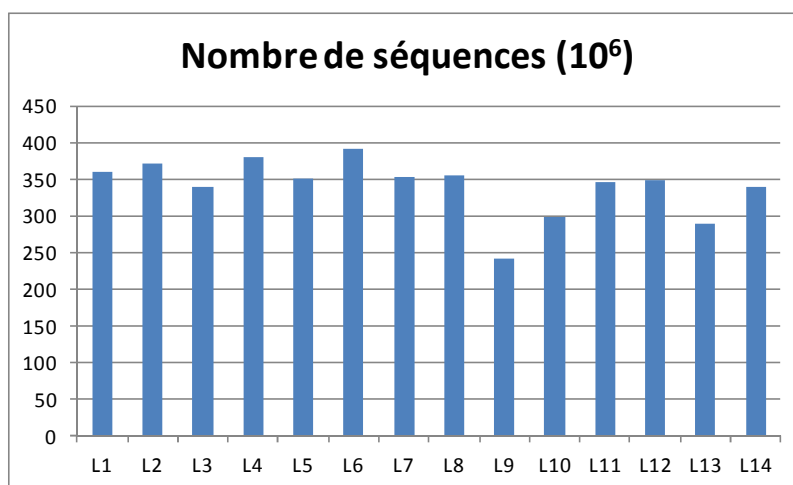


Figure 3. Principe de l'expression spécifique de génome (d'après Pastinen, 2010).

Le principe vise en premier lieu à rechercher dans les espèces parentales les SNP ayant une forme spécifique de l'une des espèces et l'autre forme spécifique de l'autre espèce. Dans l'exemple, A est spécifique du Pékin et G du Barbarie. Ensuite sont quantifiés chez les hybrides les ARN portant l'une et/ou l'autre forme de ces SNP afin d'identifier les expressions spécifiques de génome ou les déséquilibres d'expression. Dans l'exemple, chez l'hybride le gène Pékin est trois fois plus exprimé que le gène Barbarie.

