

AGENAE-CHICK, UN OUTIL D'ANALYSE DU TRANSCRIPTOME DE LA POULE

Hérault Frédéric¹, Le Meuth-Metzinger Valérie^{1,2}, Désert Colette¹, Demeure Olivier¹,
Douaire Madeleine¹

¹ INRA-Agrocampus, UMR de génétique animale, 35042 Rennes Cedex.

² Présente adresse Université Paris 13, 93017 Bobigny Cedex.

AGENAE-chick, un outil d'analyse du transcriptome de la poule

Dans le cadre du programme d'analyse du génome des animaux d'élevage (AGENAE) une banque d'ADNc multi-tissus normalisée (collection de fragments d'ADN correspondant à des gènes exprimés) a été construite pour l'espèce poule. Plus de 13000 clones ont ainsi été obtenus. Leur séquençage a permis d'obtenir 23273 séquences partielles (EST pour étiquette de séquence exprimée) publiées dans la base de données publiques de l'EMBL. Ces séquences correspondent à environ 6800 gènes exprimés différents. Leur analyse effectuée grâce à des outils développés par l'équipe informatique SIGENAE a permis de confirmer la représentativité des différents tissus utilisés pour la construction de la banque. Elle a aussi montré la bonne représentativité des fonctions moléculaires associées aux gènes représentés. AGENAE-chick constitue donc un outil générique adéquat pour l'analyse du transcriptome de la poule. Cette collection permet de générer des micro-réseaux d'ADN (puces à ADN) indispensables aux études de génomique fonctionnelle projetées dans différents domaines de la biologie avicole (croissance, développement, reproduction, santé, comportement, qualité des produits). Leur production est assurée par le centre de ressources biologiques GADIE de Jouy en Josas.

Agenae-chick, a tool for chicken transcriptome analyses

In the course of the French research programme on farm animal genomes (AGENAE) a chicken multi-tissues normalised cDNA library was constructed. More than 13,000 clones were isolated and sequenced from both ends. It led to 23,273 Express Sequence Tag (EST) which have been delivered in the EMBL nucleotide sequence database. After assembling, these EST represent approximately 6,800 different expressed genes. Analyses of these EST, performed with tools generated by the informatic team SIGENAE, shown that tissues included in the cDNA library were actually well represented in AGENAE-chick. Besides, ontology analyses demonstrated that gene-associated molecular functions were fairly accounted for in that cDNA collection. Accordingly AGENAE-chick is a valuable tool for studies in chicken functional genomics. Selected unigene clones will be printed on microarray by the Biological Resource Centre "GADIE" at Jouy en Josas.

INTRODUCTION

Depuis plusieurs années une partie de la génétique moléculaire a évolué vers la génomique, synonyme de génétique à grande échelle, s'intéressant aux génomes dans leur ensemble. Cette évolution a été rendue possible par l'apparition de techniques permettant des analyses à haut débit. On a l'habitude de distinguer les études de la structure des génomes et celles de leur fonctionnement, ces dernières prenant le nom de génomique fonctionnelle (Hatey et al., 2000 ; Douaire et al., 2001). Dans ce contexte, un programme national de recherche a été mis en place, concernant la génomique pour 4 espèces animales, dont la poule (programme AGENAE, <http://www.inra.fr/agenae>). L'un des axes de ce programme concerne la

génomique fonctionnelle. Dans ce cadre, une banque multi-tissus d'ADNc (collection d'ADNc correspondant aux gènes exprimés) de l'espèce poule a été réalisée (Hérault et al. 2003) et séquencée. Cette collection est destinée à la fabrication de micro-réseaux d'ADNc (puces à ADN) supports des études de génomique fonctionnelle. Nous présentons ici les caractéristiques de cette collection issues de l'analyse des séquences ainsi que quelques éléments de comparaison avec d'autres collections existantes utilisées pour la fabrication de micro-réseaux.

1. BANQUE D'ADNc ET ETIQUETTES (EST)

La génétique fonctionnelle s'intéresse à l'expression des gènes, et entre autres aux ARN messagers (ARNm), première étape de la cascade conduisant des gènes aux protéines puis aux caractères phénotypiques, par l'intermédiaire des différents métabolismes. L'étude des ARNm passe par leur clonage sous forme d'ADN complémentaires (ADNc) et la constitution de banques d'ADNc, collections de clones d'ADNc représentatifs du transcriptome dans un type cellulaire, un tissu, voire un organisme entier ou plutôt un ensemble de tissus d'un organisme entier. Nous avons choisi de construire une banque d'ADNc à partir des ARNm de 25 tissus différents prélevés chez des animaux d'âge, de sexe et de conditions physiopathologiques différents afin de disposer du plus grand nombre possible de gènes exprimés (Héroult et al., 2003).

L'assemblage réalisé sur l'ensemble des EST de poule disponibles dans les bases de données publiques a conduit à 86971 contigs, parmi lesquels 6820 contiennent au moins une EST AGENAE-chick.

Autrement dit 6820 ARN messagers différents sont représentés dans AGENAE-chick.

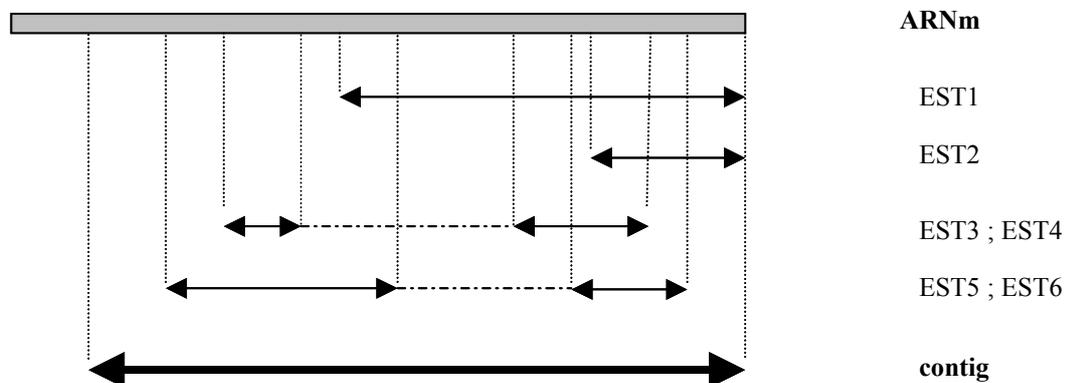
Le séquençage partiel (extrémités 5' et 3' des ADNc) des clones obtenus génère des "étiquettes" ou EST (pour Expressed Sequence Tag ou étiquette de séquence exprimée). Le séquençage de 13359 clones a ainsi permis d'obtenir 23273 EST intégrées dans la base de données SIGENAE et dans la base de séquences publique de l'EMBL (<http://www.embl-heidelberg.de/>).

Ces EST sont donc de petits fragments de séquence de gènes exprimés, plusieurs EST pouvant correspondre au même gène exprimé. Une étape de traitement informatique est alors réalisée pour regrouper (on parle d'"assemblage") les EST qui se ressemblent, sur des critères de similitude de séquences. On génère ainsi des "contigs" (figure 1). Chaque contig est censé représenter un gène exprimé (ARN messenger) mais il apparaît, d'après le nombre de contigs obtenus, que l'on n'est probablement pas encore capable de regrouper en un seul "contig" toutes les EST correspondant à un même ARNm. Nous allons cependant conserver cette analogie : un contig représente un ARN messenger (ou un gène exprimé).

Figure 1 - Correspondance entre ARNm, EST et contig

Les EST sont représentées par les traits fins double-fléchés. La correspondance avec la séquence de l'ARN est visionnée par les traits verticaux. Les EST sur des lignes horizontales différentes sont obtenues sur des clones différents (quatre dans le cas présent).

Le contig résultant de l'assemblage des EST est représenté par le trait gras double-fléché. Au contraire des clones il n'a pas de réalité physique ; il n'existe que par sa séquence

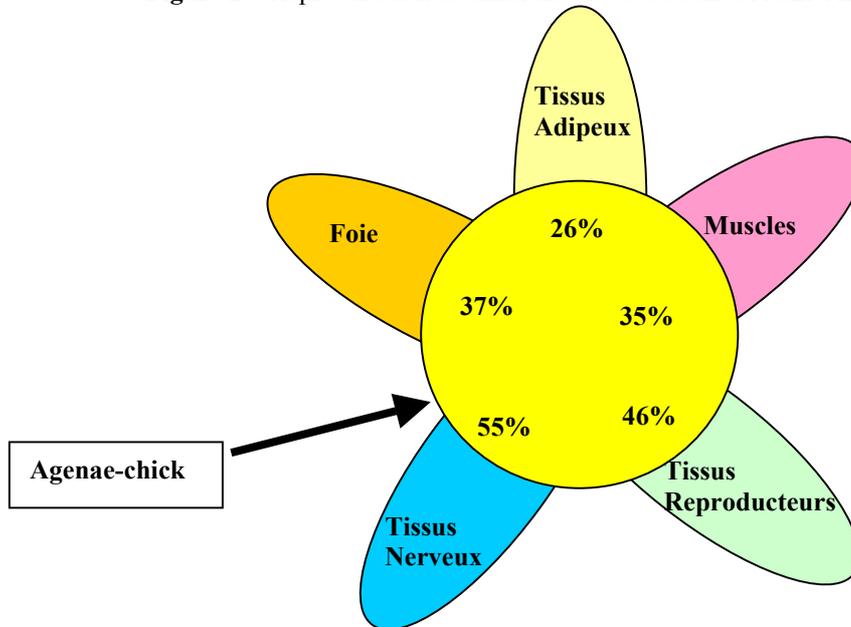


2. CARACTERISATION DES GENES PRESENTS DANS AGENAE-chick

Les ADNc présents dans AGENAE-chick proviennent de différents tissus. Les comparaisons entre AGENAE-chick et des banques mono-tissu

permettent de valider la représentativité d'AGENAE-chick pour des gènes exprimés dans différents tissus (figure 2)

Figure 2 – Représentation de différents tissus au sein d'AGENAE-chick



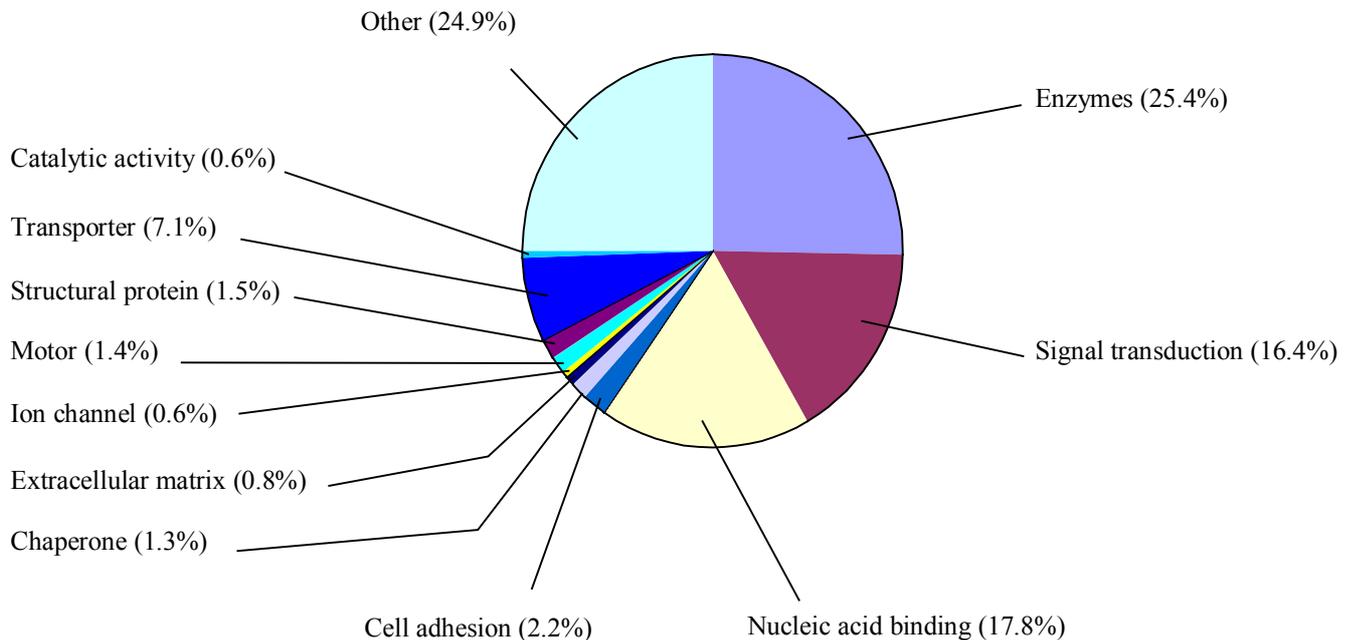
Ainsi par exemple 26 % des gènes représentés dans AGENAE-chick sont des gènes exprimés dans le tissu adipeux. Ces résultats valident le caractère multi-tissus de la collection. On peut en effet supposer que les autres tissus ayant contribué à la construction de la banque sont également bien représentés.

La somme des pourcentages obtenus pour les 5 tissus analysés dépasse 100, ce qui illustre bien la redondance qui existe entre différentes banques mono-tissu (un même gène peut être exprimé dans différents tissus), redondance que nous avons partiellement évitée avec la stratégie multi-tissus adoptée.

Par ailleurs, une analyse partielle des fonctions moléculaires associées aux gènes représentés dans la collection a été réalisée. Pour des raisons pratiques elle n'a pu porter que sur les gènes ayant une forte ressemblance avec des gènes humains soit environ 33

% des gènes représentés dans AGENAE-chick. Les fonctions associées aux gènes ont été automatiquement générées par un logiciel adapté (onto-miner, <http://vortex.cs.wayne.edu:8080>). Des groupes de gènes de fonction moléculaire similaire ont ensuite été constitués (Venter et al., 2001). Malgré la faible proportion de gènes analysés, l'analyse effectuée montre la variété des fonctions représentées (figure 3), avec une répartition similaire à celle obtenue par l'analyse du génome humain (Venter et al., 2001). Ce résultat confirme d'une autre manière qu'AGENAE-chick assure une bonne représentativité du génome aviaire. Il prouve aussi l'intérêt de cette collection pour la fabrication de micro-réseaux d'ADNc (puces à ADN) destinés aux analyses de génomique fonctionnelle (Hatey et al., 2001, Douaire et al., 2003).

Figure 3 - Distribution des fonctions moléculaires associées aux gènes représentés dans AGENAE-chick



3.COMPARAISONS AVEC LES COLLECTIONS D'ADNc EXISTANTES

En même temps qu'AGENAE, d'autres programmes de génomique fonctionnelle de la poule ont été développés dans le monde et ont conduit à l'élaboration de collections d'ADNc pour constituer des micro-réseaux d'ADN. Actuellement, un consortium regroupant des scientifiques académiques britanniques, allemands et américains propose (pour 150 \$) une puce contenant 13000 fragments de gènes différents. En comparaison, environ la moitié des séquences représentées dans AGENAE-chick (3080 sur 6820) ne figurent pas dans la collection du consortium.

Par ailleurs, un ensemble de banques d'ADNc constituées dans plusieurs universités américaines (Delaware et Maryland) ont permis l'élaboration d'une autre puce contenant près de 12000 gènes. En comparaison près de 60% des gènes représentés dans AGENAE-chick (3946 sur 6820) ne figurent pas dans cette collection.

Ces résultats confirment la richesse d'AGENAE-chick et sa complémentarité par rapport à d'autres collections. Des discussions sont engagées pour constituer une collection plus complète en association avec les collections américaines.

CONCLUSION

Nous disposons maintenant d'une collection d'ADNc représentative d'un grand nombre de gènes exprimés chez *Gallus gallus*. Les résultats présentés ici confirment sa richesse, notamment en terme de nombre

de fonctions biologiques représentées, et donc sa pertinence dans l'analyse de réseaux de régulation génique.

La fabrication de micro-réseaux d'ADN utilisant cette collection débutera en janvier 2005 au centre de ressources biologiques GADIE à Jouy en Josas qui assure déjà l'entretien des clones d'ADNc. Les micro-réseaux produits permettront de poursuivre, avec des outils plus performants, les études de génomique entreprises dans les domaines de la qualité des viandes et des carcasses, du contrôle du sexe, des résistances aux pathogènes, du comportement. Les nouveaux programmes développés en partenariat avec la profession avicole maintenant intégrée dans AGENAE bénéficieront de ce nouvel outil de génomique fonctionnelle.

REMERCIEMENTS

Ce travail a été réalisé en collaboration étroite avec l'équipe SIGENAE qui a conçu et réalisé tous les logiciels permettant les comparaisons de banque d'ADNc ainsi que la base de données de séquences utilisée.

REFERENCES BIBLIOGRAPHIQUES

- Douaire M., Vignal A., Duclos M. J., Dambrine G., 2001. Quatrièmes Journées de la Recherche Avicole, Nantes, 27-29 mars
- Hatey F., Martin P., Douaire M., Le Gac F., Dambrine G., Herpin P., Monget P., 2000. INRA Prod. Anim., HS octobre 2000, 175-180.
- Héroult F., Le Meuth-Metzinger V., Désert C., Retout E., Piumi F., Govoroun M., Duclos M. J., Douaire M., 2003. Cinquièmes Journées de la Recherche Avicole, Tours, 26 et 27 mars
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X. , 2001. Science, 291, 1304-51.