



HAL
open science

Avancées récentes en génomique des canards

Alain Vignal, Christian Diot, Caroline Molette, Christel Marie-Etancelin

► **To cite this version:**

Alain Vignal, Christian Diot, Caroline Molette, Christel Marie-Etancelin. Avancées récentes en génomique des canards. 10. Journées de la Recherche Avicole et Palmipèdes à Foie Gras, Institut Technique de l'Aviculture et des Elevages de Petits Animaux (ITAVI). Paris, FRA.; Comité Interprofessionnel des Palmipèdes à Foie Gras (CIFOG). Paris, FRA.; Institut National de la Recherche Agronomique (INRA). Paris, FRA., Mar 2013, La Rochelle, France. 279 p. hal-01210409

HAL Id: hal-01210409

<https://hal.science/hal-01210409>

Submitted on 3 Jun 2020

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.

AVANCEES RECENTES EN GENOMIQUE DES CANARDS

Alain Vignal¹, Christian Diot², Caroline Molette^{1,3}, Christel Marie-Etancelin⁴

¹INRA-ENVT, UMR 444 LGC, 31326 Castanet-Tolosan, France

²INRA, UMR 1348 PEGASE, 65 rue de Saint-Brieuc, 35042 Rennes, France

³INRA, INPT-ENSAT, INPT-ENVT, UMR TANDEM, chemin de Borde Rouge – 31326
CASTANET-TOLOSAN

⁴INRA, UR 631 SAGA, 24 Chemin de Borde Rouge, 31326 Castanet Tolosan Cedex, France
alain.vignal@toulouse.inra.fr

RÉSUMÉ

Avec la démocratisation des outils de la génomique et plus particulièrement du séquençage à haut débit, la séquence du génome du canard commun a été réalisée et va bientôt être publiée. Cependant, bien que ce soit une étape importante et un tournant dans les connaissances de la biologie de cette espèce, des projets complémentaires sont nécessaires et pour beaucoup déjà initiés : recherche de SNP (Single Nucleotide Polymorphism) pour les études et gestion de populations ; cartes d'hybrides irradiés pour ordonner la séquence le long des chromosomes ; carte génétique pour la détection des QTL (Quantitative Trait Locus) ; séquençage de transcrits pour l'annotation structurelle et fonctionnelle du génome. La première détection de QTL influençant les performances du mulard, réalisée à l'aide de marqueurs microsatellites chez la cane commune, sera complétée par une seconde étude utilisant des marqueurs SNP développés spécifiquement et permettant une bien meilleure couverture du génome. Les modèles de transcrits et de protéines construits grâce à la séquence faciliteront les études en protéomique et transcriptomique : identification de protéines, mise au point de marqueurs pour la qRT-PCR ou tout simplement existence d'un génome de référence pour l'approche par RNA-Seq. Le canard mulard, d'intérêt agronomique, étant le produit du croisement du canard commun avec le canard de Barbarie, devra également être étudié.

ABSTRACT

Recent advances in duck genomics

As the genomics becomes commonplace and sequencing throughputs increase, the duck genome sequence will be soon published. However, although this is an important step and will increase drastically knowledge on the biology of this species, complementary projects are needed, some of which initiated: SNP (Single Nucleotide Polymorphism) detection for population studies and management; radiation hybrid maps for ordering the sequence along chromosomes; genetic maps for QTL (Quantitative Trait Locus) detection; transcript sequencing for structural and functional annotation. The first QTL detection for traits influencing the performances of the mule duck, performed using microsatellite markers in the common duck, will be completed by a second study using specifically designed SNP, allowing for an improved coverage of the genome. Transcript and protein models constructed from the sequences will facilitate proteomics and transcriptomics studies: protein identification, primer design for qRT-PCR or using a reference genome for RNA-Seq approaches. The mule duck, of agronomic interest, the product of a cross between the common duck and the Muscovy duck will also have to be studied.

INTRODUCTION

La génomique est une science provenant de l'association de la biologie moléculaire et cellulaire d'une part, et de la génétique classique d'autre part. Son essor tient cependant beaucoup aux développements importants des technologies de séquençage de l'ADN et de la bioinformatique, permettant l'analyse d'un volume croissant de données. De nos jours, elle regroupe un ensemble d'analyses qui vont de l'établissement de cartes du génome (cartographie) à l'identification de nouveaux gènes, à l'étude de leurs fonctions et au séquençage des molécules d'ADN. Ainsi, on parlera de génomique structurale pour la cartographie et le séquençage des génomes et de génomique fonctionnelle pour l'étude de l'expression et de la fonction des gènes, principalement via le transcriptome et le protéome. Les applications sont par exemple la gestion des populations et l'identification de marqueurs génétiques par recherche de QTL ou par association génétique pour la génomique structurale, ou la recherche de marqueurs moléculaires de phénotypes complexes et de mise en évidence de voies métaboliques pour la génomique fonctionnelle. Nous allons ici récapituler l'état des recherches en génomique du canard, avec une vision appliquée à la filière des palmipèdes gras.

1. UN SEUL GENOME DE L'HYBRIDE EST SEQUENCE

L'animal le plus souvent utilisé (>95%) dans la filière des palmipèdes gras est le mulard, un individu hybride et stérile entre deux espèces de canards, qui sont distantes phylogénétiquement d'environ 20 millions d'années : *Anas platyrhynchos* (le canard commun) et *Cairina moschata* (le canard de Barbarie). A titre de comparaison, la distance phylogénétique entre le cheval (*Equus caballus*) et l'âne (*Equus asinus*) est estimée entre 6,4 et 12,7 millions d'années (<http://www.timetree.org/>). Quand ce n'est pas le canard mulard qui est utilisé pour le gavage, c'est le canard de Barbarie, le canard commun présentant de faibles performances et n'étant jamais utilisé. Pour le reste du monde, le terme générique « canard » se réfère au canard commun, qui est donc, et de loin, le plus étudié : soit comme modèle pour la résistance aux virus aviaires, soit en raison d'intérêts agronomiques.

2. GENOMIQUE DU CANARD COMMUN

2.1. Séquence du génome

Le séquençage du génome a été réalisé au Beijing Genomics Institute (BGI), mais n'est pas encore publiée (N. Li et Y. Huang, communication

personnelle). Le séquençage à l'aide de la technologie Illumina™, permet une diminution des coûts, mais la longueur des lectures individuelles de séquence est plus courte qu'avec la technologie « traditionnelle », dite de Sanger. Il en résulte une séquence morcelée en contigs (portions de séquences continue) et scaffolds (ou échafaudages, portions de séquence discontinue, résultant de l'assemblage de contigs) (figures 1 et 2). Le plus grand contig fait 264 kb et le plus grand scaffold 5,9 Mb, alors que la taille du plus grand chromosome est de l'ordre de 200 Mb (APL1). Une statistique couramment utilisée, pour estimer la qualité de la continuité de la séquence et de la couverture du génome est le N50 (ou le N90), utilisés pour les contigs et les scaffolds. Ce sont les valeurs pour lesquelles au moins 50% (90%) du génome sont couverts par des scaffolds ou contigs de tailles supérieures à N50 (N90). Pour le canard commun, le N50 contig est de 264 kb et le N50 scaffold de 1,2 Mb (figure 2).

2.2. Assemblage en chromosomes : de l'ordre de la poule, vers des cartes spécifiques

Le canard commun comporte 40 paires de chromosomes. A ce jour, seules des cartes génétiques rudimentaires avec des marqueurs facilement échangeables entre laboratoires ont été publiées : l'une réalisée en Chine comporte 115 marqueurs microsatellites répartis en 19 groupes de liaison, dont 10 assignés à des chromosomes (Huang *et al.*, 2006) et l'autre en France comporte 91 marqueurs microsatellites répartis en 16 groupes de liaison (Kileh-Wais *et al.*, 2012). Une troisième carte a été réalisée à Taïwan, comportant 260 marqueurs AFLP

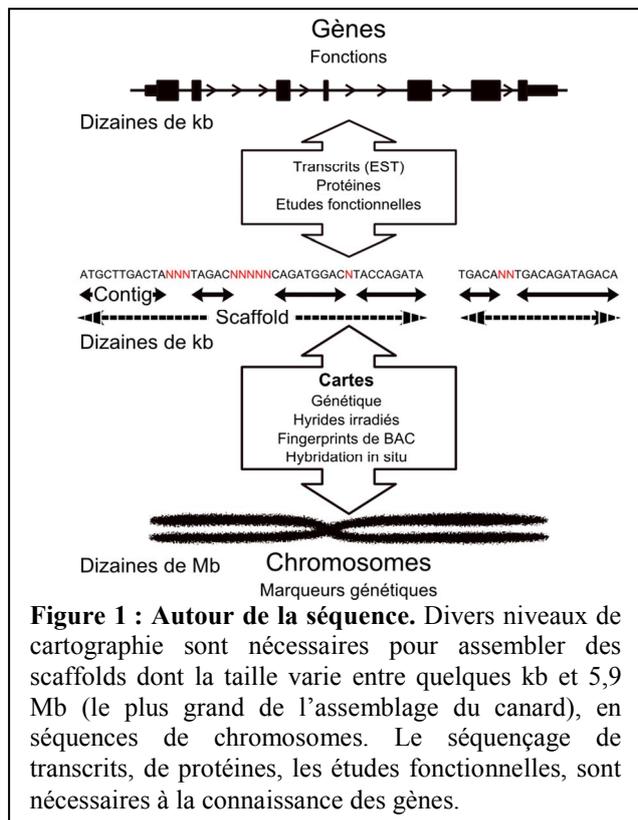
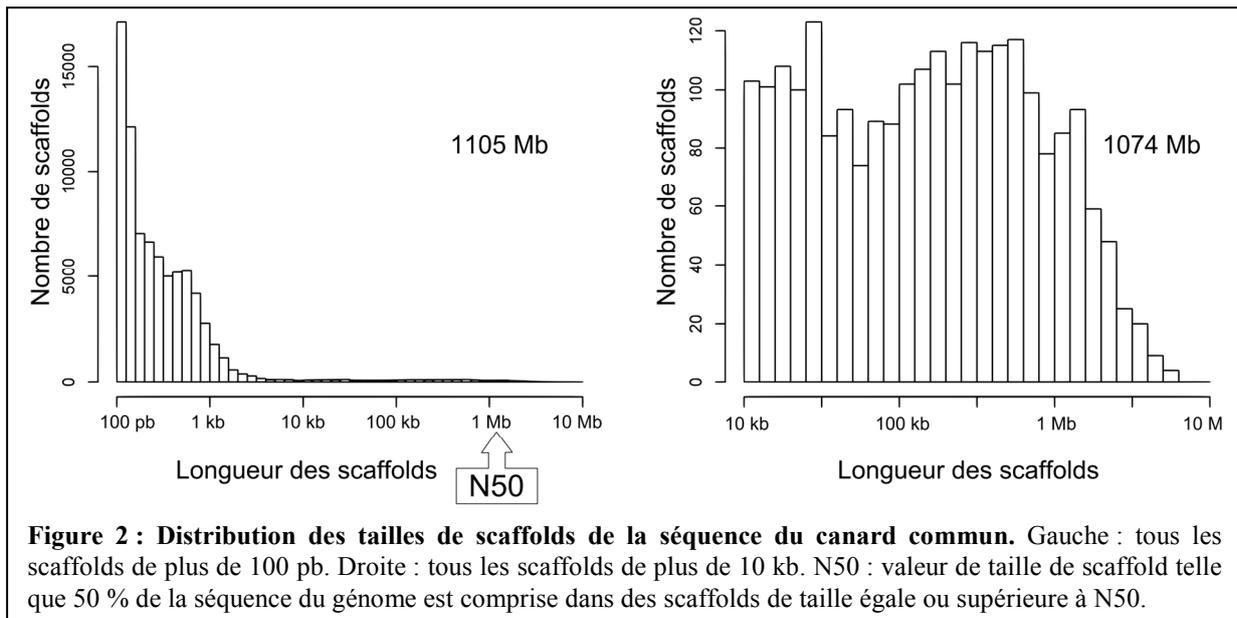


Figure 1 : Autour de la séquence. Divers niveaux de cartographie sont nécessaires pour assembler des scaffolds dont la taille varie entre quelques kb et 5,9 Mb (le plus grand de l'assemblage du canard), en séquences de chromosomes. Le séquençage de transcrits, de protéines, les études fonctionnelles, sont nécessaires à la connaissance des gènes.



répartis en 32 groupes de liaison (Huang *et al.*, 2009). Si cette dernière carte semble plus complète, il faut noter que les marqueurs AFLP sont difficiles à transférer d'un laboratoire à un autre.

Le seul réarrangement inter-chromosomique connu à ce jour entre les génomes de la poule et du canard concerne le chromosome 4 de la poule, présent sous la forme de 2 chromosomes chez le canard : GGA4 = APL4 + APL10 (GGA = *Gallus gallus* ; APL = *Anas platyrhynchos*). De plus, seul un faible nombre de réarrangements intra-chromosomiques ont été détectés (Fillon *et al.*, 2007; Skinner *et al.*, 2009). Une première approximation de l'ordre des scaffolds du canard a donc été obtenue par alignement de ces derniers sur la séquence de la poule.

Afin de mieux déterminer l'ordre des scaffolds le long des chromosomes, un panel d'hybrides irradiés a été réalisé, permettant de produire des premières cartes de chromosomes. Quelques réarrangements entre les chromosomes de poule et de canard ont pu ainsi être mis en évidence (Rao *et al.*, 2012).

Un niveau de cartographie plus fin nécessiterait la réalisation d'une carte de contigs de BAC. Bien que des banques de BAC (Bacterial Artificial Chromosome ou chromosome artificiel de bactérie) (Yuan *et al.*, 2006) et de fosmidés (Moon & Magor, 2004) aient été produites, aucun projet de cartographie n'existe à ce jour à notre connaissance.

2.3. Annotation structurale et fonctionnelle

Une fois la séquence du génome produite, il est nécessaire de repérer les gènes. Pour cela, le séquençage des transcrits, produisant des EST (Expressed Sequence Tags, ou étiquettes de séquences transcrites) et l'alignement sur la séquence du génome, permet de repérer les séquences codantes (figure 3). Bien que non publiées, des données provenant de plusieurs études de séquençage d'ARN dans divers tissus (foie, muscle, cerveau, tissu, rate, intestin, poumon) ont été utilisées pour réaliser

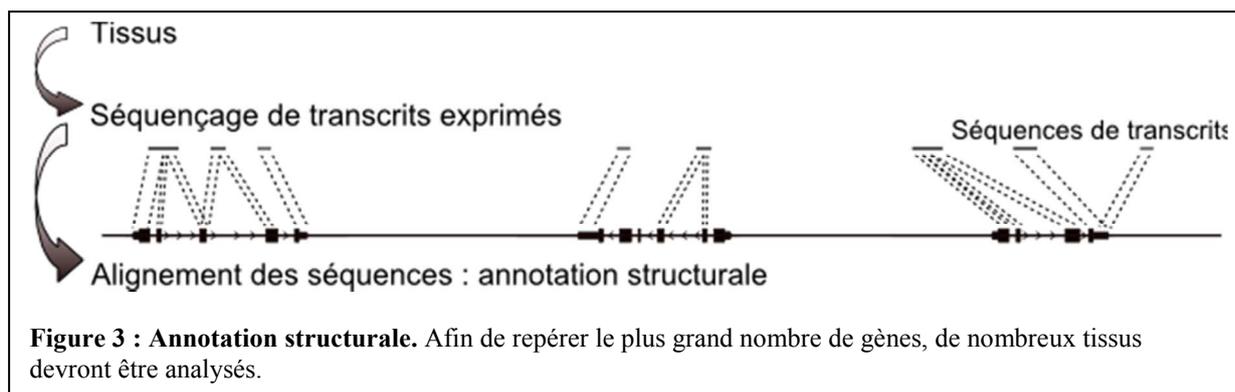
l'annotation structurale du génome, accessible sur le site [ensembl](http://pre.ensembl.org/Anas_platyrhynchos/Info/Index) (http://pre.ensembl.org/Anas_platyrhynchos/Info/Index), comportant 16 450 modèles de gènes, 17 169 modèles de transcrits et 16 353 modèles de protéines. A titre de comparaison, pour le canard de Barbarie, seul un séquençage plus restreint (cerveau, foie, muscle) de transcrits a été réalisé (Pitel *et al.*, 2009). Finalement, un projet en cours vise à analyser le transcriptome hépatique par RNA-Seq (séquençage d'ARN, pour estimer le niveau d'expression de l'ensemble des gènes après alignement sur la séquence de référence) des deux espèces de canards et leurs produits hybrides (mulards et hinnys) dans deux conditions : gavés et non gavés (Diot *et al.*, 2012).

En ce qui concerne l'annotation fonctionnelle, la seule possibilité d'assignation d'une fonction aux protéines, gènes et transcrits actuellement prédits, est par similitude de séquence avec des gènes annotés dans d'autres espèces. L'espèce la plus proche est la poule, mais l'annotation de son génome étant loin d'être complète, ou étant elle-même très dépendante des similitudes trouvées avec d'autres espèces, les résultats obtenus devront être traités avec précautions.

3. TRAVAUX EN COURS

3.1. Génétique

Les marqueurs microsatellites sont maintenant utilisés depuis plus de vingt ans et dans certaines espèces animales, sont de plus en plus remplacés par les SNP (Single Nucleotide Polymorphism ou polymorphisme de base nucléotidique ponctuel). En effet, la poule bénéficie maintenant d'une puce à SNP permettant le génotypage simultané de plus de 580 000 SNP, tandis que généralement, quelques centaines de microsatellites seulement sont utilisées dans le meilleur des cas.



Cependant, les puces à SNP ont un coût élevé et les microsatellites sont parfois plus rapides à mettre en œuvre, surtout dans le cas d'une étude restreinte à une région spécifique. En fait, un problème majeur concernant les microsatellites, est la difficulté de les développer par séquençage des régions flanquantes dans le cas d'un génome inconnu. Cependant, la disponibilité de la séquence du génome a maintenant permis d'accéder directement à plus de 6 000 microsatellites, qui sont maintenant immédiatement disponibles. Une carte du chromosome APL12, nécessaire à une étude QTL en cours, a ainsi été réalisée.

Afin d'élargir les études au génome entier, le développement de SNP est en cours de réalisation, par séquençage et alignement sur le génome de référence (Vignal *et al.*, 2012), mais les coûts actuels du génotypage avec des puces comportant plusieurs milliers de marqueurs peuvent encore être prohibitifs dans le cadre d'une étude comportant moins d'un millier d'animaux. La réalisation d'une carte génétique comportant plus de 350 marqueurs SNP, choisis de manière à optimiser la couverture du génome est prévue à très court terme. Cette carte étant construite à partir du dispositif de détection de QTL Genecan (Marie-Etancelin *et al.*, 2008), la détection de QTL, eQTL et pQTL pourra être réalisée dans des conditions optimales (François *et al.*, 2012).

3.2 Génomique fonctionnelle

Des progrès indéniables ont été faits, mais comme en génétique, les stratégies doivent être adaptées aux moyens et outils disponibles. L'approche récente de RNA-Seq est très prometteuse, mais ne peut être réalisée pour le moment que sur un nombre restreint d'échantillons. Pour des études nécessitant un plus grand nombre d'échantillons, telles que rencontrées en génétique pour la recherche de eQTL (QTL d'expression génique), le séquençage est donc exclu. L'option d'utiliser des puces d'expression est nettement préférable, mais est pour le moment écartée pour des raisons de prix. La miniaturisation des volumes réactionnels ayant permis de réduire les coûts de la qRT-PCR, une étude pour une centaine de gènes sur plusieurs centaines d'individus est en cours (François *et al.*, 2012).

L'étude des profils d'expression protéique dans les tissus d'intérêt (principalement le foie) revêt un intérêt particulier pour la compréhension des mécanismes physiologiques. L'étape limitante de ces analyses reste l'identification des protéines par spectrométrie de masse. En effet, une faible annotation des génomes de canard conduit à un faible taux d'identification (50-70%) des protéines chez le canard mulard lié à l'obligation de travailler par homologie de séquences avec d'autres espèces animales telle que la poule.

4. PERSPECTIVES ET APPLICATIONS

4.1 Gestion des populations

Un premier jeu de marqueurs microsatellites a été défini pour de l'assignation de parenté à la fois chez le canard commun et le canard de Barbarie (Chapuis *et al.*, 2010a; Chapuis *et al.*, 2010b). Cependant, il n'avait pu être réalisé qu'à partir d'un nombre restreint de marqueurs disponibles à l'époque et nécessitait de tester au préalable tous les marqueurs dans les deux espèces. Le séquençage de canards de Barbarie va permettre de repérer les marqueurs microsatellites en commun entre les deux espèces et donc d'améliorer ce panel. Le séquençage des mêmes canards de Barbarie va aussi permettre la détection de SNP, en référence à l'assemblage de la séquence du canard commun.

4.2 Connaissance des génomes

L'analyse des résultats du programme de RNA Seq impliquant les deux espèces et leurs hybrides permettra d'une part d'avoir une première vue exhaustive des variations de l'expression génique entre le canard commun, le canard de Barbarie et leurs descendants hybrides, mais aussi une bonne idée de la divergence nucléotidique (SNP) des régions codantes et de structure primaire des protéines.

4.3 Liens entre phénotype-génotype

Les connaissances accumulées en génomique du canard commun sont bien plus importantes que pour le canard de Barbarie. Cependant, les outils proposés par la génomique n'ont d'intérêt pour l'analyse des caractères, que s'il existe en regard des dispositifs animaux expérimentaux correspondants, souvent

lourds à mettre en place. En matière de recherche de QTL, le seul dispositif existant ne permet d'aborder que la voie du canard commun pour la recherche de QTL. Cependant, il n'est pas inconcevable de considérer que le prix du génotypage à haute densité puisse diminuer dans un avenir proche, auquel cas les recherches pourraient être réalisées par tests d'association directement dans des populations en sélection.

CONCLUSION

Ne disposant pas des moyens pouvant être mis en œuvre pour les espèces majeures, le canard a bénéficié récemment des progrès réalisés sur les coûts et la rapidité des technologies de la génomique et notamment du séquençage. Une difficulté additionnelle, liée à la filière, est l'exploitation d'un produit final, fruit d'un croisement interspécifique, nécessitant une connaissance approfondie du génome de deux espèces.

Un point important à souligner est que la profusion de données générées par les technologies à haut débit est à mettre en regard des problèmes nouveaux posés par leur analyse, qui devient souvent l'étape critique. L'acquisition des données phénotypiques est également un point à ne pas négliger.

REFERENCES BIBLIOGRAPHIQUES

- Chapuis H., Boscher MY., Faugeras R., Genestout L., Meriaux JC., Feve K., Guemene D., 2010a. In "13. European Poultry Conference (EPC 2010)", pp. 309, French Branch of World's Poultry Science Association.
- Chapuis H., Faugeras R., Rossignol MN., Feve K., Genestout L., Chantry-Darmon C., Guemene D., 2010b. In "9. Journées de la Recherche sur les Palmipèdes à Foie Gras", pp. 1-4, ITAVI.
- Diot C., Houée-Bigot M., Demeure O., Baéza E., Vignal A., Pitel F., Marie-Etancelin C., Robert-Granié C., Molette C., Bouchez O., Esquerré D., Marsaud N., Klopp C., Peterlongo P., Lemaitre C., 2013. Dixièmes Journées de la Recherche Avicole et Palmipèdes à Foie Gras, La Rochelles, 26-28 mars 2013.
- François Y., Marie-Etancelin C., Molette, C., Vignal A., Davail S., 2013 Dixièmes Journées de la Recherche Avicole et Palmipèdes à Foie Gras, La Rochelles, 26-28 mars 2013.
- Fillon V., Vignoles M., Crooijmans R. P., Groenen M. A., Zoorob R., Vignal A., 2007. *Anim Genet* 38: 303-7.
- Huang C. W., Cheng Y. S., Rouvier R., Yang K. T., Wu C. P., Huang H. L., and Huang M. C. (2009). *Genet Sel Evol* 41: 28.
- Huang Y., Zhao Y., Haley C. S., Hu S., Hao J., Wu C., Li N., 2006. *Genetics* 173: 287-96.
- Kileh-Wais M., Elsen J. M., Vignal A., Feves K., Vignoles F., Fernandez X., Manse H., Davail S., Andre J. M., Bastianelli D., Bonnal L., Filangi O., Baeza E., Guemene D., Genet C., Bernadet M. D., Dubos F., Marie-Etancelin C., 2012. *J Anim Sci. sous presse*.
- Marie-Etancelin C., André JM., Baéza E., Basso B., Bastianelli D., Bernadet MD, Brun JM., Davail S., Dubos F., Fernandez X., Gontier K., Guémené D., Guy G., Manse H., Mialon MM., Larzul C., 2008. 8èmes JRPF, Arcachon, 30-31 octobre 2008, 17-20.
- Moon D. A., Magor K. E., 2004. *Anim Genet* 35: 417-8.
- Pitel F., Vignal A., Leroux S., Feve K., Vignoles F., Tircazes A., Morisson M., Marty A., Donnadiou C., Milan D., Gourichon D., F. M., Leterrier C., Arnould C., Bernadet M. D., Marie-Etancelin C., Basso B., Herault F., Lecerf F., Besnard J., Calenge F., Beaumont C., Klopp C., Diot C., 2009. In "8èmes Journées de la Recherche Avicole, St Malo, France, 25-26 mars 2009, 594-598."
- Rao M., Morisson M., Faraut T., Bardes S., Feve K., Labarthe E., Fillon V., Huang Y., Li N., Vignal A., 2012. *BMC Genomics* 13: 513.
- Skinner B. M., Robertson L. B., Tempest H. G., Langley E. J., Ioannou D., Fowler K. E., Crooijmans R. P., Hall A. D., Griffin D. K., Volker M. 2009. *BMC Genomics* 10: 357.
- Yuan X., Zhang M., Ruan W., Song C., Ren L., Guo Y., Hu X., Li N. 2006. *Anim Genet* 37: 599-600

