

DEVELOPPEMENT DE MARQUEURS SNP CHEZ LE CANARD COMMUN POUR LA RECHERCHE DE QTL

Alain Vignal¹, Olivier Rué², Christel Marie-Etancelin³, Christophe Klopp⁴

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

²INRA, UMR 444 LGC, SIGENAE, 31326 Castanet-Tolosan, France

³INRA, UR 631 SAGA, 24 Chemin de Borde Rouge, 31326 Castanet Tolosan Cedex, France

⁴INRA, Plateforme Genotoul, 31326 Castanet-Tolosan, France

alain.vignal@toulouse.inra.fr

RÉSUMÉ

La détection et l'utilisation de SNP est de nos jours indispensable à la recherche en génétique : étude et gestion de populations, recherche de QTL dans des croisements ou par test d'association dans des populations. Une première détection de QTL dans le dispositif Back-Cross GeneCan a été réalisée en utilisant 91 marqueurs microsatellites ne couvrant que 778 cM en 16 groupes de liaison (Kileh-Wais et al., 2012). Or, le génome du canard comporte 40 paires de chromosomes (y compris les gonosomes Z et W) et si l'on estime que la taille de la carte génétique attendue pourrait être proche de celle de la poule, une longueur de l'ordre de 3000 cM est attendue (Groenen et al., 2009). Afin de compléter cette étude en étendant la carte génétique par ajout de marqueurs, nous avons effectué une recherche de SNP de manière ciblée pour le protocole GeneCan par séquençage des ADN des 7 pères du dispositif.

Le choix d'un jeu final de 384 marqueurs a été fait en maximisant l'informativité des pères F1 et en optimisant la répartition des SNP le long du génome, en tenant compte de la différence de taux de recombinaison existant entre les macrochromosomes et les microchromosomes.

ABSTRACT

Development of SNP in common duck for QTL detection

The detection and utilisation of SNP is nowadays essential for research in genetics: study and managing of populations, QTL detection in crosses or in populations by association testing. A first QTL detection study in the Genecan Back-Cross design was performed using 91 microsatellite markers covering 778 cM in 16 linkage groups (Kileh-Wais et al., 2012). The duck genome contains 40 chromosome pairs, including the Z and W gonosomes, and if the estimates for the length of the genetic map are close to that of chicken, around 3000 cM are expected (Groenen et al., 2009). To complete our QTL scan, we detected SNP specifically for our cross by sequencing the 7 F1 sires.

The choice of a final set of 384 markers was done by maximizing the heterozygosity of the 7 F1 sires and the distribution of the SNP along the genome, taking into account the difference in recombination rates between macrochromosomes and microchromosomes.

INTRODUCTION

Quelles que soient les avancées de la génomique en matière de séquençage des génomes, la localisation de caractères phénotypiques sur le génome et le développement de marqueurs génétiques moléculaires par l'approche de la détection de QTL (Quantitative Trait Loci) passent par la réalisation préalable d'une carte génétique. Le canard comporte 40 paires de chromosomes et à ce jour, seules des cartes génétiques rudimentaires 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, réalisée en France, comporte 91 marqueurs microsatellites répartis en 16 groupes de liaison (Kileh-Wais *et al.*, 2012). Cette dernière a été réalisée en utilisant le dispositif Genecan de détection de QTL et sera la seule qui nous intéressera ici, notre but étant de détecter des QTL dans ce dispositif.

Ces cartes génétiques sont donc loin de couvrir les chromosomes et une autre estimation de leur couverture partielle du génome vient de la longueur de 778 cM de la carte du dispositif Genecan, alors que si l'on se base sur la carte génétique de la poule, une longueur bien plus importante de 3000 cM environ est attendue (Groenen *et al.*, 2009).

Afin d'étendre la détection de QTL réalisée dans le

dispositif Genecan, il donc est nécessaire de développer de nouveaux marqueurs. Tous les marqueurs microsatellites publiés et/ou développés à ce jour au laboratoire ayant été testés, nous avons décidé de nous tourner vers les marqueurs de type SNP. Un premier jeu de 384 SNP développé dans l'équipe de Groenen (Kraus *et al.*, 2011) et testé dans notre population n'ayant révélé que très peu de polymorphisme, il s'est avéré plus intéressant de développer un jeu spécifique.

Pour ce faire, nous avons séquencé les ADN des pères des 7 familles du dispositif Genecan, afin de détecter les SNP les plus informatifs et de pouvoir les choisir en fonction de leur position sur le génome, afin de le couvrir au mieux.

1. MATERIELS ET METHODES

1.1. Séquençage et détection des SNP

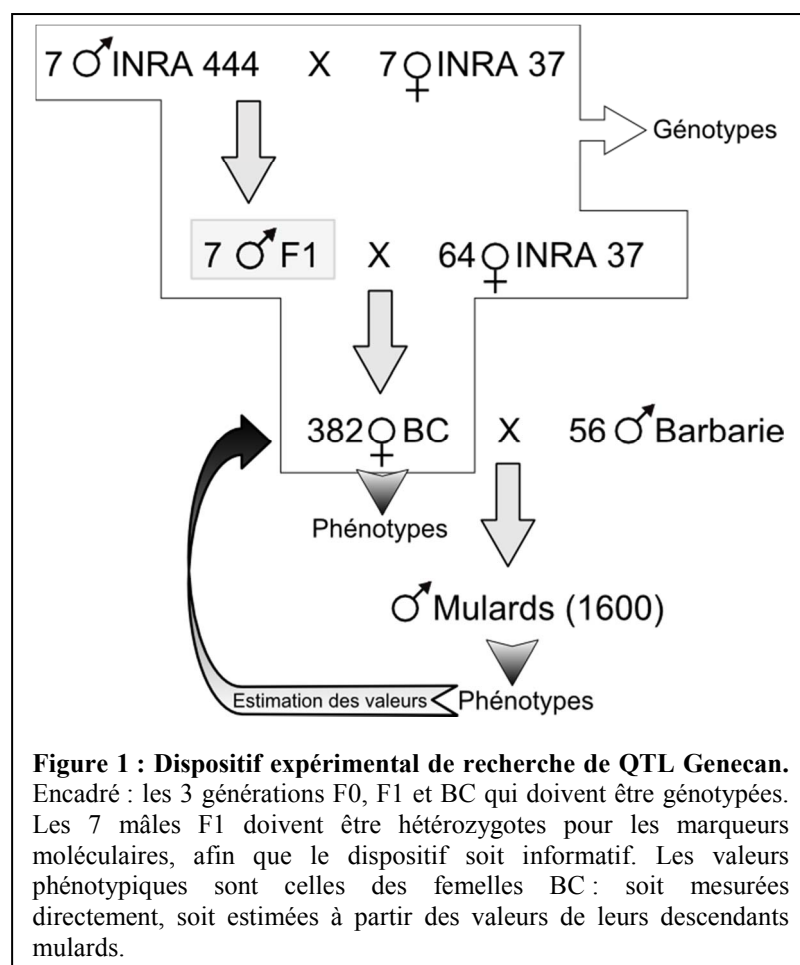
Les ADN des 7 pères F1 de chaque famille de demi sœurs du dispositif ont été séquencés chacun dans une piste d'un séquenceur Illumina HiSeq, avec des lectures de 100 pb en paired-end (longueur des lectures = 100 nucléotides et les deux extrémités de fragments de 300 pb environ d'ADN sont lues). La qualité des séquences est estimée par alignement sur la séquence de référence du génome du canard (scaffolds de plus de 1 kb) et vérification que les deux

lectures paired-ends sont bien présentes et proches l'une de l'autre.

Afin de détecter les SNP, l'alignement des séquences est affiné à l'aide de GATK (Genome Analysis Tool Kit) (McKenna *et al.*, 2010) et les lectures s'alignant avec une qualité de mapping < 30 (1 chance sur 1000 de mauvais alignement) sont éliminées. Le processus élimine notamment les lectures s'alignant en plusieurs endroits du génome, susceptibles de générer de nombreux faux SNP.

1.2. Positionnement sur le génome de la poule

Le génome du canard n'étant pas à ce jour assemblé en chromosomes, l'estimation de la position des SNP est réalisée par alignement sur le génome de la poule, qui présente une très forte conservation de synténie avec celui du canard (Fillon *et al.*, 2007; Skinner *et al.*, 2009). Les SNP de canard sont alignés sur le génome de la poule en utilisant BWASW (Li & Durbin, 2010) et 150 pb de part et d'autre des SNP.



1.3. Jeu de SNP final

Le but étant de réaliser les génotypages par la technologie Golden Gate® VeraCode® d'Illumina, 384 SNP, soit le nombre nécessaire pour un jeu VeraCode®, ont été choisis en tenant compte (i) de leur score Illumina (ii) de leur informativité chez les 7 pères F1 et (iii) de leur espacement le long du génome, en utilisant celui de la poule comme référence.

2. RESULTATS ET DISCUSSION

2.1. Séquençage et détection de SNP

Un total de 3 milliard de séquences a été obtenu, soit 300 Gbases ou 250 équivalents-génome, en estimant la taille du génome du canard à 1,2 Gb. Chaque individu présente une couverture moyenne de 35 équivalents-génome. Après les filtres de qualité (alignement sur le génome de référence), 80 % des séquences sont jugées utilisables pour la détection de SNP et un total de 11 million de SNP a été trouvé après l'analyse par GATK. Cependant, 90 % ne sont informatifs que dans un seul mâle F1 et seul 1 % des SNP sont informatifs pour les 7 (Tableau 1).

Nb. F1 hét.	Nombre de SNPs	Pourcentage
7	119 203	1,09 %
6	539 047	4,93%
5	1 488 143	13,60%
4	3 053 588	27,91%
3	5 086 133	46,49%
2	7 432 296	67,93%
1	9 870 810	90,22%

Tableau 1 : Nombre de SNP détectés et informativité. Nb. F1 hét. : nombre de mâles F1 hétérozygotes pour les SNP. Nombre de SNP : nombre ayant au moins X mâles hétérozygotes. Pourcentage : par rapport à tous les SNP détectés sur le génome de référence

2.2. Positionnement sur le génome de la poule

La position de 1 million de SNP a pu être prédite sur le génome du canard en utilisant les résultats d'alignement de 150 pb de séquence de part et d'autre des SNP sur le génome de la poule. En ne gardant que les SNP présentant du polymorphisme pour au moins 5 des 7 pères F1 du dispositif Genecan, il reste 157 436 SNP. La répartition de ces SNP le long du génome de la poule est montrée pour 6 chromosomes (Figure 2). La répartition inégale des SNP le long des chromosomes peut soit être due à une variation régionale de l'hétérozygotie entre les 2 lignées à

F1 hét	SNP alignés poule	SNP panel
5	101 267	1
6	43 563	145
7	12 606	238
Total	157 436	384

Tableau 2 : Informativité des SNP positionnés sur le génome de la poule.

l'origine du croisement, soit à une similarité plus faible entre les génomes de la poule et du canard. Une fois les cartes génétiques construites, l'ordre réel des marqueurs chez le canard sera établi.

2.3. Jeu de SNP final

Afin de couvrir le génome de la manière la plus uniforme possible, tout en gardant un niveau élevé d'informativité, un jeu de 384 SNP a été choisi en tenant compte de la position sur le génome de la poule et du nombre de pères hétérozygotes.

Nous avons utilisé un critère de distance moyenne entre marqueurs différent en fonction de la taille des chromosomes, afin de tenir compte des taux de recombinaison plus élevés pour les microchromosomes. Seule une minorité des 157 436 SNP positionnés sur le génome de la poule sont informatifs pour les 7 pères du dispositif (Tableau 2). Cependant, nous avons réussi à obtenir plus de 60 % de SNP informatifs pour les 7 pères dans le panel final. La répartition des SNP du panel est montrée pour 6 chromosomes de poule figure 2. On peut voir que la répartition des SNP n'est pas uniforme, avec des régions pauvres en SNP informatifs localisés sur le génome de la poule, comme par exemple sur GGA6. C'est pour s'assurer de la couverture correcte de telles

régions, qu'un séquençage complet des pères, tel que réalisé ici, est nécessaire. La distance moyenne entre marqueurs pour les 3 catégories de chromosomes est montrée dans le tableau 3.

Chromosomes	Distance entre SNP (kb)
Macros (GGA1, 2, 3, 4q, 5 et Z)	3 000
Moyens (GGA6-10)	1 200
Microchromosomes et GGA4p	640

Tableau 3 : Distance entre SNP dans le panel des 384 SNP finaux, en référence au génome de la poule.

CONCLUSION

Le séquençage des génomes des 7 pères du dispositif de recherche de QTL Genecan a permis de choisir un jeu de marqueurs pour le génotypage, qui va permettre d'améliorer grandement la qualité de la carte génétique du canard. En plus d'une détection de QTL plus complète, tenant compte notamment des microchromosomes connus chez la poule, la carte génétique générée va permettre de contribuer à l'assemblage du génome du canard. Les marqueurs développés sont spécifiques des lignées impliquées dans le croisement QTL. Afin de développer un panel de marqueurs généraliste, une étude approfondie d'un plus grand nombre de lignées sera nécessaire.

REFERENCES BIBLIOGRAPHIQUES

Fillon V., Vignoles M., Crooijmans R. P., Groenen M. A., Zoorob R., and Vignal A. (2007). FISH mapping of 57 BAC clones reveals strong conservation of synteny between Galliformes and Anseriformes. *Anim Genet* **38**: 303-7.

Groenen M. A., Wahlberg P., Foglio M., Cheng H. H., Megens H. J., Crooijmans R. P., Besnier F., Lathrop M., Muir W. M., Wong G. K., Gut I., and Andersson L. (2009). A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. *Genome Res* **19**: 510-9.

Huang Y., Zhao Y., Haley C. S., Hu S., Hao J., Wu C., and Li N. (2006). A genetic and cytogenetic map for the duck (*Anas platyrhynchos*). *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., and Marie-Etancelin C. (2012). Detection of QTL controlling metabolism, meat and liver quality traits of the overfed interspecific hybrid mule duck. *J Anim Sci*, sous presse.

Kraus R. H., Kerstens H. H., van Hooft P., Crooijmans R. P., van der Poel J. J., Elmberg J., Vignal A., Huang Y., Li N., Prins H. H., and Groenen M. A. (2011). Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC Genomics* **12**: 150.

Li H., and Durbin R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589-95.

McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernysky A., Garimella K., Altshuler D., Gabriel S., Daly M., and DePristo M. A. (2010). The Genome Analysis Toolkit: a MapReduce

framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297-303.

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., and Volker M. (2009). Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. *BMC Genomics* **10**: 357.

