

**RECHERCHE DES REGIONS GENOMIQUES RESPONSABLES DU DIABETE
INSIPIDE ET DE LA COULEUR DE COUILLE CELADON CHEZ LA CAILLE
JAPONAISE PAR RADSEQ.**

**Leroux Sophie^{1*}, Marissal Anna^{1*}, Bernard Maria^{2,6}, Esquerré Diane³, Barbieri
Johanna³, Gourichon David⁴, Rivière Sandrine⁴, Robic Annie¹, Hochu Isabelle¹,
Lagarrigue Sandrine⁵, Klopp Christophe⁶, Zerjal Tatiana², Bed'hom Bertrand²,
Minvielle Francis², Pitel Frédérique¹**

*¹GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 CASTANET TOLOSAN;
²GABI, INRA, AgroParisTech, Université Paris-Saclay, 78352 JOUY-EN-JOSAS; ³INRA,
GeT-PlaGe Genotoul, 31326 Castanet Tolosan; ⁴INRA, PEAT, 37380 NOUZILLY; ⁵ UMR
INRA-Agrocampus Ouest, PEGASE, 35000 RENNES; ⁶ INRA, Siganae - BIA, 31326
CASTANET-TOLOSAN*

**co-premiers auteurs*

frederique.pitel@inra.fr

RÉSUMÉ

La caille japonaise est une espèce pour laquelle il n'existe pas de puce de génotypage commerciale. Nous avons testé l'utilisation de la technique de RADSeq (Restriction site Associated DNA Sequencing), pour la première fois sur cette espèce. Le RADSeq est une technique de génotypage par séquençage d'une sous-représentation du génome - la même pour tous les individus - obtenue par sélection de fragments d'ADN issus de la digestion du génome par des enzymes de restriction. Les séquences obtenues sont ensuite comparées pour détecter des polymorphismes de type SNP (Single Nucleotide Polymorphism) et analyser le génotype de chaque individu en ces positions variables. Cette technique a été appliquée à l'étude de deux phénotypes : "céladon" et "diabète insipide", dont le caractère mendélien est connu ou soupçonné. Le premier est responsable de la couleur de coquille "céladon" (bleu-vert) ; le second, qui est présent à faible fréquence dans les populations commerciales, se manifeste par une polydipsie et une polyurie très marquées chez les femelles, et par une diminution de l'hématocrite. Un dispositif familial dans lequel ségrègent ces deux phénotypes a été mis en place : 158 femelles F2, informatives pour les deux caractères, ont été obtenues par croisement de lignées fixées pour l'un ou l'autre des phénotypes. Plus de 230 000 SNP ont été identifiés, et 36 244 ont été utilisés pour des analyses d'association avec les caractères étudiés, ce qui a permis d'identifier la région de "céladon" sur le chromosome 16. La région du "diabète insipide" n'a pas pu être localisée, des analyses complémentaires sont en cours.

ABSTRACT

Using RADSeq to search genomic regions causing diabetes insipidus and eggshell color "celadon" in Japanese quail.

Japanese quail is a species for which there is no commercial genotyping chip. We tested the RADSeq (Restriction site associated DNA Sequencing) technique for the first time in this species. RADSeq is a technique for genotyping by sequencing a reduced representation of the genome - the same for all individuals - obtained by selecting DNA fragments derived from restriction enzyme digestion of the genome. The sequences obtained are then compared to detect Single Nucleotide Polymorphisms (SNPs) and to analyze the genotype of each individual in these variable positions. This technique has been applied to the study of two phenotypes: "celadon" and "diabetes insipidus". The former is responsible for the shell color called "celadon" (blue-green); the latter, which is present at low frequency in commercial populations, is characterized by polydipsia and polyuria, very marked in females, and by a decrease of hematocrit levels. A family design in which both phenotypes are segregating has been established: 158 F2 females informative for both traits were obtained by crossing quail lines fixed for either phenotype. More than 230,000 SNPs were identified, and 36,000 were used for association analyzes with the traits under study, identifying the "celadon" region on chromosome 16. The "diabetes insipidus" region could not be localized, and additional analyzes are under way.

INTRODUCTION

Les nouveaux outils de génotypage à haut débit qui ont émergé il y a quelques années, comme les puces à SNP (Single Nucleotide Polymorphism), ont largement permis d'améliorer les capacités d'analyse des génomes des espèces agronomiques d'intérêt. Cependant, les espèces moins étudiées, comme la caille, ne disposent pas encore de puce de génotypage. L'étude menée ici a pour but de localiser deux gènes majeurs chez la caille japonaise *Coturnix japonica*, ce qui nécessite une bonne couverture du génome par des marqueurs polymorphes. Nous avons donc choisi de mettre en œuvre une technique de génotypage par séquençage, le RADSeq (Restriction-site Associated DNA Sequencing), qui permet de détecter efficacement un grand nombre de marqueurs sur l'ensemble du génome, que celui-ci soit connu ou non.

Diabète insipide

Des travaux menés il y a quelques années ont montré que le diabète insipide chez la caille japonaise était un caractère mendélien déterminé par une mutation récessive autosomale : *di* (Minvielle *et al.* 2007). La proximité phylogénétique entre *Gallus* et *Coturnix*, le travail antérieur sur l'hérédité de *di* chez la poule et les recherches récentes menées sur diverses mutations communes aux deux taxa laissent penser qu'un même gène, qui reste à découvrir, pourrait être impliqué dans les deux espèces. La lignée de cailles diabétiques (*di/di*) que nous avons développée et caractérisée (Minvielle *et al.* 2007) constitue une ressource unique pour trouver une mutation responsable du diabète insipide chez les oiseaux. Cette mutation (*di*) qui se maintient à faible fréquence dans les élevages commerciaux se manifeste par une polydipsie et une polyurie, très marquées chez les femelles, ainsi que par une diminution de l'hématocrite. Le syndrome n'est pas associé à une altération significative des performances zootechniques mesurées habituellement, sauf pour la consommation résiduelle qui est augmentée, certainement en relation avec le coût métabolique du phénomène polydipsie/polyurie. Chez l'homme et la souris, le diabète insipide est associé à une mutation du gène du récepteur 2 à la vasopressine, lié au sexe, ou à une mutation dans le gène autosomal de l'aquaporine 2 (Bichet 2009). Chez les oiseaux, l'hormone vasotocine AVT, homologue de la vasopressine des mammifères, joue un rôle central dans la régulation de l'équilibre hydrique mais nous n'avons trouvé de différence de concentration de l'hormone ni dans le cerveau ni dans le plasma des cailles *di/di*. En revanche, des travaux préliminaires menés sur la morphologie des structures rénales et sur le gène de l'aquaporine 2, qui est régulé par la vasotocine, ont montré des différences au niveau de la zone médullaire du rein et une diminution d'expression de ce gène, mais pas de différence dans sa séquence (Yang *et al.* 2008). A ce stade, si l'on

peut penser que c'est au niveau du rein qu'il faut chercher les effets de la mutation *di* aviaire, aucun gène candidat particulier n'est privilégié et c'est donc une recherche au niveau du génome complet que nous avons conduite.

Celadon

Cette mutation (*ce*), autosomale récessive, supprime complètement le dépôt de la pigmentation brune de surface de l'œuf qui donne le patron de coloration usuel chez la caille, et produit donc un œuf qui montre la couleur verte sous-jacente de la coquille. Or le patron de coloration de l'œuf de caille constitue un exemple typique du camouflage des œufs très répandu chez les oiseaux, et la connaissance de la mutation *ce* est donc une possible clé d'entrée dans la compréhension de l'architecture génétique de ce caractère complexe. Un précédent projet visait à localiser chez la caille quatre mutations particulièrement intéressantes car sans homologue connu chez la poule : deux mutations du plumage, "rusty" et "curly", et deux mutations de la couleur de la coquille des œufs, "white" et "celadon", qui produit un œuf bleu-vert (Leroux *et al.* 2013). L'analyse d'association conduite a mis en évidence des associations très significatives pour 3 des phénotypes, mais l'association avec *ce* est seulement suggérée, sur l'équivalent chez la Caille du chromosome 16 de la poule, dont la séquence de référence est très incomplète.

1. MATERIELS ET METHODES

1.1. Dispositif animal et phénotypage

Les cailles ont été élevées au Pôle Expérimental Avicole de Tours (PEAT). Cinq mâles F0 de la lignée céladon (homozygotes *ce/ce*) et cinq femelles F0 de la lignée diabète insipide (homozygotes *di/di*) ont été croisés (1 mâle pour 1 femelle), pour obtenir une génération F1 hétérozygote pour les deux mutations. Puis, 5 mâles F1 ont été croisés avec 20 femelles F1 (1 mâle pour 4 femelles), en évitant les croisements entre apparentés. Au total 158 femelles F2 ont été obtenues, sur lesquelles nous avons évalué la couleur de l'œuf et la polyurie (caractérisée par des fientes liquides). Ces F2 présentaient bien les 4 types de phénotypes attendus : 10 [CE, DI] (*ce/ce* et *di/di*), 33 [CE, WT] (*ce/ce*, et *di+/di* ou *di+/di+*), 33 [WT, DI] (*ce+/ce* ou *ce/ce*, et *di/di*), et 82 [WT, WT] (*ce+/ce* ou *ce+/ce+*, et *di+/di* ou *di+/di+*), avec "WT" : phénotype sauvage, "ce+" : allèle sauvage au locus *ce*, "di+" : allèle sauvage au locus *di*. Les proportions observées (10, 33, 33, 82) sont compatibles avec les proportions mendéliennes attendues (1/16, 3/16, 3/16, 9/16 soit 10, 30, 30 et 88 individus respectivement). Ainsi, nous disposons au sein de la même population de 43 femelles de phénotype mutant et 115 femelles de phénotype sauvage pour chacun des deux caractères CE et DI.

Des prélèvements de tissus (reins en particulier) ont été réalisés pour des analyses fonctionnelles ultérieures, et les échantillons de sang prélevés ont permis d'effectuer les extractions d'ADN.

1.2. Génotypage et analyses d'association

Au total, 192 individus ont été analysés par la technique de RADSeq (Baird *et al.*, 2008), soit l'ensemble des individus F0, F1 et F2, à l'exception d'une mère F1 dont le sang n'était pas disponible pour l'extraction d'ADN. Les ADN génomiques ont été digérés par deux enzymes de restriction (*EcoRI* et *TaqI*) et les fragments obtenus ont subi une ligation d'adaptateurs universels. Les résultats de ligation ont été soumis à une PCR effectuée grâce à des amorces Illumina indexées, permettant l'identification ultérieure de l'individu d'origine de chaque lecture. Les 192 échantillons ont été regroupés pour être séquencés ensemble sur trois lignes de HiSeq 3000 (2x150 bp). La fraction de l'amplifiat comprise entre 150 et 850 bp environ a préalablement été sélectionnée, en récupérant la fenêtre correspondant à la taille souhaitée après migration sur gel. Nous avons utilisé un logiciel dédié à l'analyse des résultats de RADSeq pour détecter les SNP et en déduire le génotypage des animaux : Stacks (Catchen *et al.*, 2011). Des scripts ont été développés pour en faciliter l'utilisation sur nos données. L'assemblage du génome de la caille (https://www.ncbi.nlm.nih.gov/assembly/GCF_00157_7835.1) ayant été rendu public en mars 2016, il a été possible d'utiliser l'option de Stacks qui permet un alignement des lectures sur le génome de référence avant la détection des SNP. Les analyses d'association (détails en Figure 2) ont été réalisées avec le package R GenABEL (Aulchenko *et al.*, 2007).

2. RESULTATS ET DISCUSSION

2.1. Détection de SNP et génotypage

Pour les 192 individus et les 3 lignes réunies, 1 275 116 654 lectures ont été obtenues et analysées avec Stacks. Le contrôle qualité a permis de conserver 78,4% des lectures. En moyenne, 99 469 loci (fragments du génome obtenus après digestion et séquençage) par individu ont été obtenus. Parmi ces loci, 27,6% présentaient des SNP, à hauteur de 2,3 SNP par locus en moyenne. Ainsi, 88 006 loci polymorphes pour 233 339 SNP uniques ont été obtenus. Sur l'ensemble de ces loci génotypés, la profondeur moyenne est de 33,1X. Deux individus comportaient très peu de séquences et ont donc été écartés lors des contrôles qualité de GenABEL, à cause d'une trop faible proportion de génotypes couverts sur l'ensemble des marqueurs. Parmi les 233 339 SNP obtenus, 36 244 SNP uniques sont conservés avec un taux de génotypage de 0,95, c'est-à-dire 95% des individus génotypés pour chaque SNP.

La répartition des SNP obtenus pour un taux de génotypage de 0,95 est représentée sur la Figure 1. On voit que les SNP sont répartis de façon assez homogène sur le génome de la caille. La distance moyenne entre marqueurs, hors chromosomes sexuels, est de 22,6 kb, avec une distance médiane de 35 kb, et une distance maximale de 1,2 Mb (sur le chromosome 3). La grande majorité (98,6%) des SNP sont présents sur des chromosomes identifiés, le reste se trouvant sur des "scaffolds" non assignés. Mis à part les chromosomes sexuels et le microchromosome 16 - dont l'assemblage est très fragmentaire -, la quantité de SNP sur un chromosome ou un scaffold est proportionnelle à la taille de celui-ci, ce qui confirme une bonne répartition des SNP à travers tout le génome. Enfin, le nombre de SNP reste homogène entre chromosomes avant et après les filtres qualité utilisés avec GenABEL.

2.2. Analyses d'association

La Figure 2 présente les graphiques de type "Manhattan-plots" obtenus pour chacun des caractères avec les SNP présentant un taux de typage de 0,95. Pour le caractère "céladon", un pic majeur apparaît au niveau du chromosome 16 (CM003796.1) (Fig. 2A), où 16 SNP ont une p.valeur inférieure au seuil de $1,38 \cdot 10^{-6}$ ($\alpha=0,05$, correction de Bonferroni). Le marqueur le plus associé (p.valeur = $6,24 \cdot 10^{-23}$) est localisé sur le scaffold 492. Ce résultat suggère que ce scaffold pourrait appartenir au chromosome 16. Nous avons confirmé cette hypothèse par analyse de liaison (résultats non présentés). L'assemblage du génome de la caille dans la région étant incomplet, tout comme celui de la poule, aucun gène candidat n'a pu être mis en évidence. Un séquençage des individus recombinants dans la région est en cours.

Aucune association significative n'est apparue pour le caractère "diabète insipide" (Figure 2B). Parmi les hypothèses permettant d'expliquer ce résultat, on peut mettre en avant les difficultés de phénotypage de ce caractère, et la possibilité de n'avoir pas généré de marqueurs dans la région portant le gène impliqué. Les données de RADSeq vont être complétées par le séquençage tout génome d'un mélange d'individus de la lignée d'origine, porteurs à l'état homozygote de l'allèle défavorable.

CONCLUSION

La technique de RADSeq a été utilisée avec succès chez la caille. Elle a permis de générer plus de 36 000 SNP de bonne qualité pour les analyses d'association dans ce dispositif sur l'ensemble du génome de la caille, et de localiser la région du génome impliquée dans la couleur céladon de la coquille sur le chromosome 16. Des analyses complémentaires sont en cours (en particulier le séquençage d'individus recombinants) pour tenter d'identifier les causes moléculaires des phénotypes étudiés.

REFERENCES BIBLIOGRAPHIQUES

- Aulchenko, Y. S., D. J. de Koning and C. Haley, 2007. *Genetics* (177): 577-85.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko and E. A. Johnson, 2008. *PLoS One* (3): e3376.
- Bichet, D. G., 2009. *Prog Mol Biol Transl Sci* (89): 15-29.
- Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko and J. H. Postlethwait, 2011. *G3* (Bethesda) (1): 171-82.
- Leroux S., Dehais P., Vignal A., Bouchez O., Faraut T., Rossignol M., Arnould C., Leterrier C., Pitel F., Minvielle F., 2013. *Dixièmes Journées de la Recherche Avicole et Palmipèdes à Foie Gras*, La Rochelle, France.
- Minvielle, F., R. Grossmann and D. Gourichon, 2007. *Poult Sci* (86): 249-54.
- Yang, Y., F. Minvielle, R. J. Kuykindoll, J.-M. Gasc, T. Yamamoto and H. Nishimura, 2008. *FASEB J* (22): 24.

Figure 1 : Répartition des 36 244 SNP utilisés sur le génome de référence de la caille (Coja2)
En abscisse : position sur le chromosome (en bases) ; en ordonnée : chromosome (ou groupe de liaison)

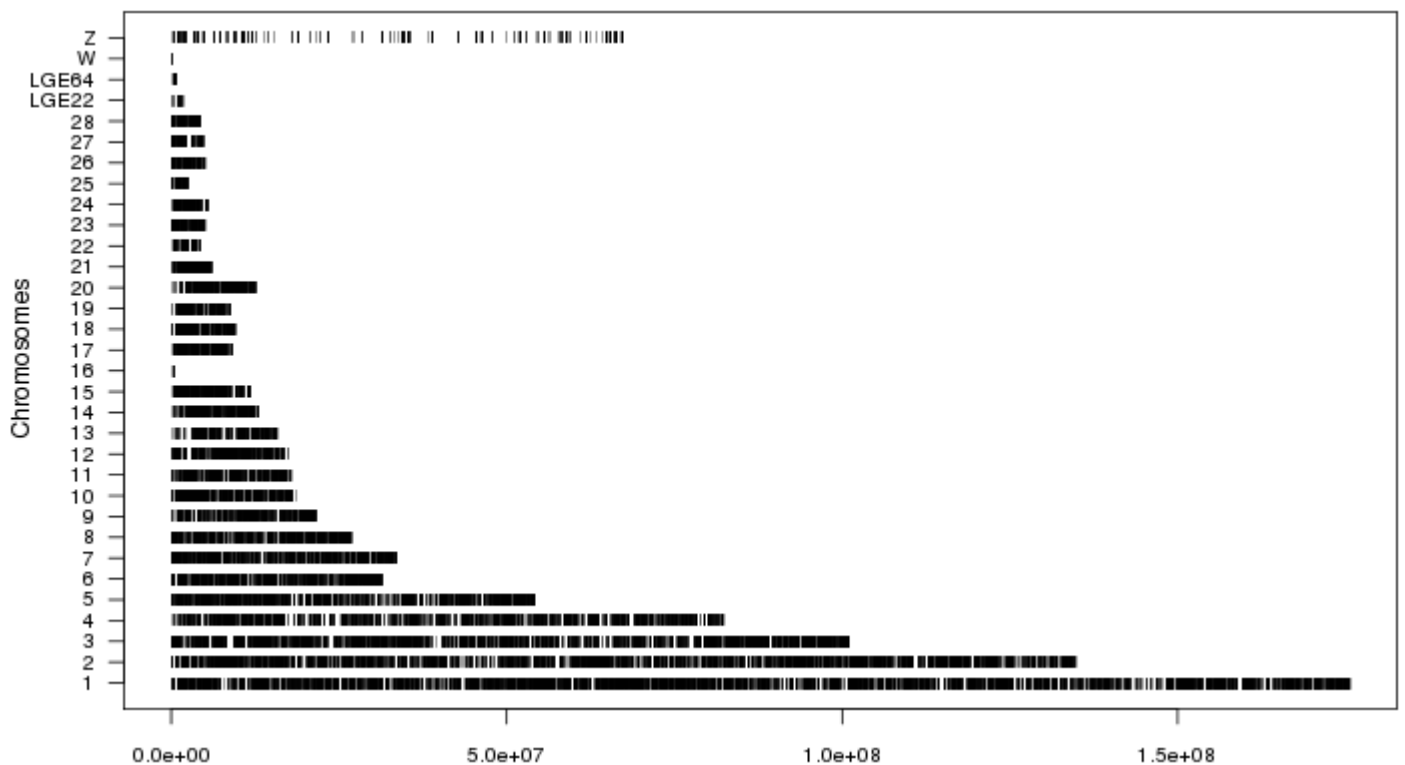


Figure 2 : Résultats d'association des marqueurs avec le caractère étudié

(A) Caractère "couleur de la coquille céladon " (CE). (B) Caractère "diabète insipide " (DI).

Les analyses ont été réalisées sur l'ensemble des individus génotypés. Chaque point, en fonction de la position du SNP sur le génome, indique le degré de significativité du test d'association d'un SNP avec le caractère en question ($-\log_{10}(p\text{-valeur})$). Nous avons appliqué une correction de Bonferroni, conservative, pour établir le seuil de significativité, en considérant 36 244 tests indépendants. Chaque changement de chromosome est illustré par un changement de couleur. La p.valeur pour chaque marqueur est obtenue par un modèle linéaire mixte incluant l'effet additif du SNP à chaque position, tenant compte de l'apparentement (FASTA, Family-based Score Test for Association). Un test utilisant un modèle linéaire récessif a également été utilisé pour DI, sans plus de succès (résultats non présentés).

