

DETECTION DE SNP CHEZ LA CAILLE : PRODUCTION D'UN OUTIL COMMUN POUR DES PROGRAMMES SCIENTIFIQUES MULTIPLES.

**Leroux Sophie¹, Dehais Patrice¹, Vignal Alain¹, Bouchez Olivier^{1,2}, Faraut Thomas¹,
Gourichon David³, Rivière Sandrine³, Calandreau Ludovic⁴, Bertin Aline⁴, Bed'hom
Bertrand⁵, Recoquillay Julien⁶, Duval Elisabeth⁶, Beaumont Catherine⁶, Rossignol
Marie-Noëlle⁷, Arnould Cécile⁴, Leterrier Christine⁴, Pitel Frédérique¹, Minvielle
Francis⁵**

¹INRA-ENVT - UMR444 Laboratoire de Génétique Cellulaire - 31326, Castanet-Tolosan,

²INRA - GeT-PlaGe Genotoul - 31326, Castanet-Tolosan,

³INRA - PEAT Pôle d'Expérimentation Avicole de Tours - 37380, Nouzilly,

⁴INRA, UMR85 Physiologie de la Reproduction et des Comportements; CNRS, UMR7247 ;
Université de Tours, IFCE - 37380, Nouzilly,

⁵INRA-AgroParisTech - UMR1313 Génétique Animale et Biologie Intégrative - 78350, Jouy-
en-Josas,

⁶INRA - UR83 Recherche Avicoles - 37380, Nouzilly,

⁷LABOGENA - 78350, Jouy-en-Josas

sophie.leroux@toulouse.inra.fr

RÉSUMÉ

Pour une espèce non modèle comme la caille, accéder aux outils actuels de génotypage utilisant un grand nombre de SNP (Single Nucleotide Polymorphism) nécessite la mise en commun de ressources provenant de programmes différents pour identifier un jeu de marqueurs pertinent et mutualiser les génotypes. Cette démarche a été appliquée à trois études. La première vise à détecter des QTL (Quantitative Trait Loci) responsables de la motivation sociale chez les oiseaux. Le second programme a pour but d'identifier deux mutations du plumage, "rusty" et "curly", et deux mutations de la couleur de la coquille, "celadon" et "œuf blanc". Le dernier dispositif cherche à évaluer les performances et la fertilité des hybrides caille des blés x caille japonaise dans un but d'évaluation des risques de pollution génétique des populations de cailles des blés par la caille japonaise. Pour obtenir des SNP utilisables dans les trois dispositifs, un séquençage complet du génome a été effectué sur les parents F0 de chaque lignée divergente et sur des cailles des blés. Une moyenne de 20 milliards de bases par lignée ont été générées. Les analyses ont été effectuées en alignant les séquences obtenues sur le génome de la poule. Plus de 11 millions de SNP ont été détectées. A partir des SNP les plus discriminants entre les lignées, 6000 marqueurs ont été choisis, le mieux répartis possible sur le génome, pour effectuer le génotypage des individus issus de chaque programme.

ABSTRACT

SNP detection in quail: sharing resources to build a collaborative genotyping tool.

In quail, a non model species, accessing modern genotyping tools, using a large number of SNPs (Single Nucleotide Polymorphism), requires pooling resources from different programs. We have joined three studies. The first project aims to detect QTL responsible for social motivation in birds. The second program aims to identify two plumage mutations, "rusty" and "curly", and two shell color mutations, "celadon" and "white egg". The last study seeks to evaluate the performance and fertility of hybrids between Japanese and common quail, to assess the risk of "genetic pollution" of common quail by Japanese ones. To obtain SNPs usable in the three devices, whole genome sequencing was performed on F0 parents of each line and common quail samples. An average of 20 billion bases per line was obtained. Analyzes were performed by aligning the obtained sequences on the chicken genome. More than 11 millions SNP were detected. From the most discriminating SNPs between the lines, 6,000 markers were chosen, for a genome coverage as uniform as possible, to perform genotyping in each program.

INTRODUCTION

Grâce à l'avancée des technologies de séquençage et de génotypage, les SNP (Single Nucleotide Polymorphism) sont maintenant les marqueurs de choix pour la plupart des programmes de génétique. Pour une espèce non modèle comme la caille, contrairement au poulet qui est très étudié, accéder aux outils actuels de génétique moléculaire nécessite la mise en commun de ressources provenant de programmes différents. Nous avons ainsi regroupé trois études distinctes pour produire un outil de génotypage mutualisé.

Le premier projet vise à détecter, à partir d'un croisement entre lignées divergentes pour un caractère de comportement, des QTL (Quantitative Trait Loci) responsables de la motivation sociale chez les oiseaux, et à déterminer les liens génétiques entre les caractéristiques de différents comportements sociaux et les principaux paramètres de production (ponte, poids, consommation alimentaire...).

Le second a pour but d'identifier les locus responsables de deux phénotypes de plumage et deux phénotypes de couleur de la coquille.

Le dernier programme cherche à évaluer les performances et la fertilité d'animaux d'une population issue d'hybridation entre caille des blés (*Coturnix c. coturnix*) et caille japonaise (*Coturnix c. japonica*), notamment dans un but d'évaluation des risques de pollution génétique des populations de cailles des blés par la caille japonaise.

Pour obtenir des SNP utilisables dans les trois dispositifs, un séquençage complet du génome a été effectué sur les parents F0 de chaque lignée divergente - afin de détecter des SNP spécifiques de lignée - et sur des cailles des blés. Les analyses ont été effectuées en alignant les séquences obtenues sur le génome de la poule. A partir des SNP les plus discriminants entre les lignées, 6000 marqueurs ont été choisis pour effectuer le génotypage (Infinium iSelect) des individus issus de chaque programme.

1. MATERIELS ET METHODES

1.1. Dispositifs animaux

- Motivation sociale

Un dispositif F2 a été produit à partir de deux lignées expérimentales sélectionnées de manière divergente pour leur motivation sociale (Mills et al., 1991) : la lignée HSR (High Social Reinstatement) et la lignée LSR (Low social Reinstatement). Les descendants F2 (n=910) ont été mesurés pour des caractères de comportement et de production (voir Recoquilly et al., 2013).

- Mutations (type de plumage et couleur de coquille)

Deux populations F2 d'environ 200 cailles chacune ont été produites et phénotypées, chacune présentant deux des quatre locus en ségrégation : les deux phénotypes de plumage (RU : couleur "rusty" et CU : forme "curly" ; (Minvielle et al., 2005), les deux phénotypes de couleur de coquille (CE : "celadon" et

WE : "œuf blanc"). La population totale est composée de 425 animaux.

- Hybrides (caille japonaise x caille des blés)

Les performances de production de deux populations, japonaise pure et issue d'hybridation entre caille japonaise et caille des blés (Chazara et al., 2010, Minvielle et al., 2012), ont été suivies jusqu'à la F12. Le dispositif animal comprend 255 individus.

1.2. Séquençage

Le séquençage a été réalisé à partir d'animaux fondateurs des trois populations d'étude.

Trois lignes de séquenceur HiSeq (Illumina) ont été utilisées, pour des fragments de 100 paires de bases (pb), en "paired-ends", à raison d'une ligne par population de cailles (mélange de fondateurs : 10 HSR, 10 LSR et 2 cailles des blés).

1.3. Détection de SNP

En l'absence de génome de référence chez la caille, les séquences obtenues ont été alignées sur le génome de la poule, en utilisant le programme d'alignement de séquences bwa, option bwasw pour optimiser l'alignement hétérologue (Li et al., 2010). Plusieurs filtres ont été appliqués pour sélectionner les séquences à partir desquelles ont été détectés les SNP :

- absence de contamination par de l'ADN hétérologue
- qualité des séquences et de leur alignement sur le génome de référence
- absence de bases non déterminées (N)
- positionnement unique sur le génome de la poule
- pourcentage d'identité supérieur à 70%.

1.4. Elaboration d'un set de marqueurs

La sélection des SNP potentiellement utilisables (de bonne qualité et optimisant la possibilité de discrimination entre lignées) a été effectuée selon plusieurs critères :

- nature du SNP (pas d'insertion/délétion, hors régions de répétition, seulement SNP bialléliques, qualité, profondeur de lecture)
- spécificité par rapport aux lignées
- adéquation avec les impératifs technologiques (pas de SNP "ambigu" (A/T ou C/G), pas de SNP additionnel dans les 120 paires de bases flanquantes).

Les SNP filtrés ont ensuite été sélectionnés en les triant par ordre décroissant de score Illumina, pouvoir de discrimination et profondeur de lecture. Enfin, la sélection finale a été effectuée de manière à optimiser la répartition des SNP sur le génome, en se basant sur la carte génétique de la poule (Groenen et al., 2009).

1.5. Génotypages et analyses

Le set de marqueurs défini a été utilisé pour le génotypage de 1632 individus par la méthode iSelect d'Illumina. Les génotypes ont été filtrés en fonction de leur qualité (GenomeStudio V2010.1, Illumina).

- Motivation sociale

Le logiciel CRI-MAP (Green et al., 1990), version 2.503 (version révisée du programme CRI-MAP v2.4

de Phil Green, modifiée par Jill Maddox et Ian Evans) est utilisé pour la construction des cartes génétiques à partir du dispositif F2.

- Mutations

Seuls les marqueurs informatifs (pour lesquels différents allèles étaient observés chez les reproducteurs) ont été conservés. Une analyse d'association entre les phénotypes et les génotypes a été conduite sur ces SNP informatifs chez les 425 individus concernés, à l'aide du logiciel PLINK (Purcell *et al.*, 2007).

- Hybrides

Les résultats de génotypage concernant cette partie du programme restent à analyser.

2. RESULTATS ET DISCUSSION

2.1. Séquençage et détection de SNP

Des fondateurs des trois populations d'étude (lignées HSR et LSR et caille des blés) ont été séquencés pour sélectionner les SNP les plus informatifs dans notre dispositif, c'est-à-dire ceux discriminant au maximum les trois populations de caille. Le nombre de bases obtenues par séquençage est équivalent dans les trois lignées : 19,206,067,484 pour la lignée HSR, 23,311,378,932 pour la lignée LSR, et 17,721,519,792 pour la caille des blés. Une moyenne de 42% des lectures s'alignent sur le génome de la poule (Figure 1). Ce résultat, inférieur à celui obtenu précédemment (72%, Frésard *et al.*, 2012), s'explique en partie par la différence entre les méthodologies utilisées.

Les données (après filtres sur la profondeur, le caractère biallélique non ambigu - hors AT et CG - et la localisation sur un chromosome connu de poule) ont généré 7,178,545 marqueurs utilisables chez la caille japonaise (représentée par les lignées HSR et LSR) et 11,371,040 SNP en intégrant les données issues de la caille des blés. Un total de 39% des SNP issues des trois populations ne sont pas retrouvées au sein des seules cailles japonaises. En se limitant aux SNP de la même espèce, japonaise, on observe un ratio transitions/transversions de 2,3, similaire à celui établi préalablement dans cette espèce sur des données génomiques (2,2, Frésard *et al.*, 2012). Cette valeur est comparable à celles obtenues chez le poulet (2,2), le canard (2,3) et, dans une moindre mesure, la mésange charbonnière (1,7) (Kraus *et al.*, 2011, Van Bers *et al.*, 2010). On trouve également un ratio supérieur à 1 dans d'autres espèces comme le navet, la truite arc-en-ciel ou l'homme (Boussaha *et al.*, 2012, Park, 2009, Park *et al.*, 2010).

REMERCIEMENTS

Le séquençage haut-débit a été réalisé sur la plate-forme GeT-PlaGe Genotoul de Toulouse Midi-Pyrénées. Nous remercions Céline Chantry-Darmon (Labogena, Jouy-en-Josas) pour la gestion des génotypes, et Jill Maddox pour la mise à disposition de la version adaptée de CRI-MAP. Ce projet a bénéficié des financements ANR "SNP-BB", AIP INRA Bioressources "Pyresavi" et AAP INRA Département de Génétique animale "Genocot".

2.2. Génotypages

Un total de 5472 SNP constitue le panel de marqueurs finalement utilisés (Figure 2). Des problèmes ont été rencontrés dans l'interprétation des génotypes, et près de la moitié des SNP se sont avérés inexploitable. Il semble que la méthode de définition des séquences flanquantes en vue du génotypage par la technique Illumina "iSelect" soit en cause. Néanmoins, près de 4 millions de génotypes ont été obtenus et sont en cours d'analyse.

- Motivation sociale

Un total de 2145 SNP ont été correctement analysés. La construction des cartes génétiques est en cours.

- Mutations

Au total, 2090 SNP étaient informatifs dans ce dispositif. L'analyse d'association réalisée a permis de localiser des régions du génome candidates pour chaque locus (Figure 3): RU, CU et WE sont respectivement associés à des régions des chromosomes 1, 5 et 6 de manière très hautement significative ($P < 10^{-20}$). CE est associé au chromosome 16 de manière faiblement significative ($P < 10^{-5}$), ce qui peut s'expliquer par la faible couverture de ce chromosome, riche en séquences répétées, dont l'assemblage est encore à améliorer.

CONCLUSION

Cette étude a permis d'obtenir la séquence d'individus issus des trois populations de caille (deux lignées de caille japonaise et une lignée de caille des blés). Ces données permettront d'une part de détecter des signatures de sélection entre lignées sélectionnées de manière divergente sur leur motivation sociale, et éventuellement de détecter des régions du génome spécifiques de chacune des espèces (caille japonaise et caille des blés).

Ce travail a également abouti à la définition d'un set de marqueurs génétiques utilisable pour plusieurs programmes indépendants chez la caille. Les résultats de génotypage ont d'ores et déjà permis de localiser les 4 mutations recherchées. Une cartographie plus fine est désormais envisageable pour aboutir à leur identification. La construction de cartes génétiques de moyenne densité est en cours sur le dispositif F2 pour permettre la recherche de QTL intervenant dans le comportement social. Ces cartes constitueront les premières cartes génétiques de SNP dans cette espèce. Ce programme souligne l'intérêt, pour les "petites espèces" - pour lesquelles les outils de génotypage haut-débit ne sont pas disponibles - de fédérer nos efforts pour générer des outils utilisables par la communauté.

REFERENCES BIBLIOGRAPHIQUES

- Boussaha M, Guyomard R, Cabau C, Esquerre D et Quillet E, 2012. BMC Genomics, (13), 238.
- Chazara O, Minvielle F, Roux D, Bed'hom B, Feve K, Coville J-L, Kayang BB, Lumineau S, Vignal A, Boutin J-M et Rognon X, 2010. Conserv. Genet., (11), 1051-1062.
- Frésard L, Leroux S, Dehais P, Servin B, Gilbert H, Bouchez O, Klopp C, Cabau C, Vignoles F, Feve K, Ricros A, Gourichon D, Diot C, Richard S, Leterrier C, Beaumont C, Vignal A, Minvielle F et Pitel F, 2012. BMC Genomics, (13), 551.
- Green P, Falls K et Crooks S, 1990. Washington University School of Medicine, St. Louis, MO., (
- Groenen MA, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans RP, Besnier F, Lathrop M, Muir WM, Wong GK, Gut I et Andersson L, 2009. Genome Res, (19), 510-519.
- Kraus RH, Kerstens HH, Van Hooft P, Crooijmans RP, Van Der Poel JJ, Elmsberg J, Vignal A, Huang Y, Li N, Prins HH et Groenen MA, 2011. BMC Genomics, (12), 150.
- Li H et Durbin R, 2010. Bioinformatics, (26), 589-95.
- Mills AD et Faure JM, 1991. J Comp Psychol, (105), 25-38.
- Minvielle F, Gourichon D et Moussu C, 2005. BMC Genet, (6), 14.
- Minvielle F, Gourichon D, Rivière S et Rognon X, (2012), *XXIV World's Poultry Congress*, Salvador, Bahia, Brazil, 5 - 9 August 2012.
- Park L, 2009. Genet Res (Camb), (91), 293-303.
- Park S, Yu HJ, Mun JH et Lee SC, 2010. Mol Genet Genomics, (283), 135-45.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et Sham PC, 2007. Am J Hum Genet, (81), 559-75, <http://pngu.mgh.harvard.edu/purcell/plink/>.
- Recoquilly J, Leterrier C, Pitel F, Bertin A, Calandreau L, Gourichon D, Rivière S, Beaumont C, Le Bihan-Duval E et Arnould C, (2013), *10èmes Journées de la Recherche Avicole et Palmipèdes à Foie Gras, 26 au 28 mars 2013*, La Rochelle,
- van Bers NE, van Oers K, Kerstens HH, Dibbitts BW, Crooijmans RP, Visser ME et Groenen MA, 2010. Mol Ecol, (19 Suppl 1), 89-99.

Figure 1. Alignement des séquences sur le génome de la poule
Pour chaque chromosome de poule sont représentés les taux de couverture par les séquences de chacune des trois lignées de caille.

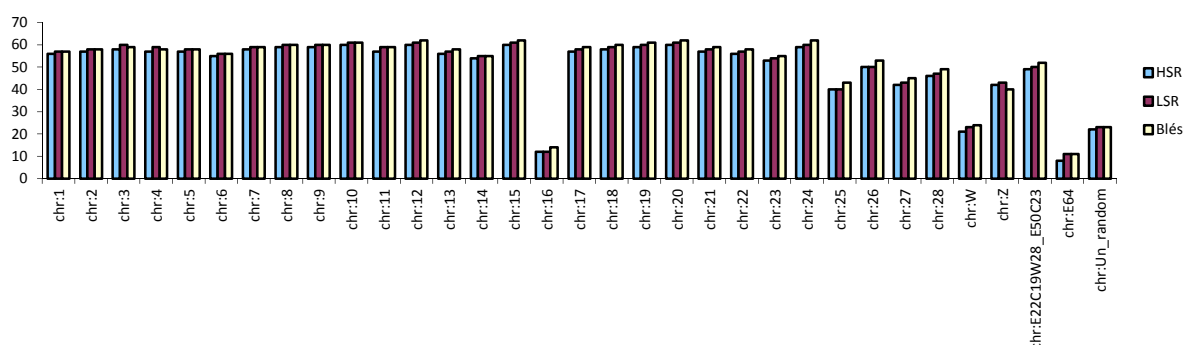


Figure 2. Set de marqueurs final pour des analyses chez la caille.
La densité de SNP couvrant le génome est représentée par chromosome de poule (SNP/cM).
Les rapports SNP/cM du chromosome W et du groupe de liaison E64 ne sont pas connus.

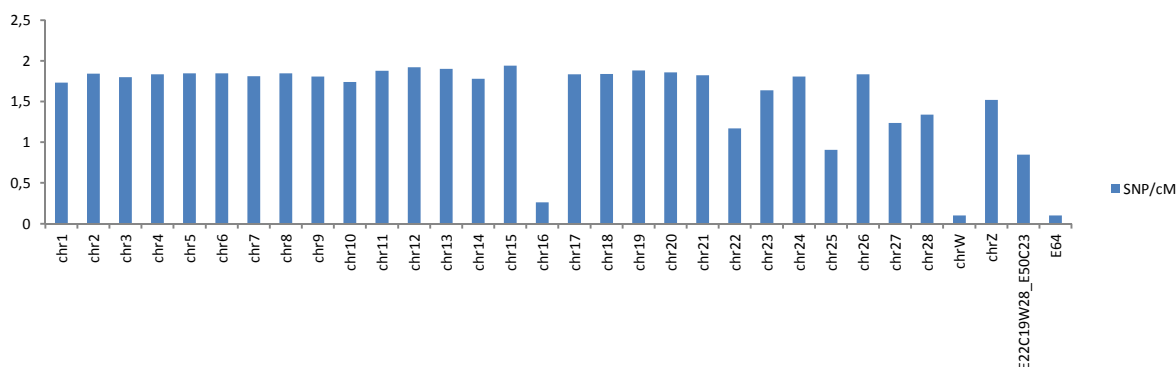


Figure 3. Analyse d'association pour les 4 mutations

Chaque point indique le degré de signification du test d'association pour un SNP en fonction de sa position orthologue sur le génome de poule. Les différents chromosomes sont représentés par différentes couleurs (premier en noir : SNP non assignés à un chromosome de poule, puis chromosome 1, 2, etc).
Plumage : RU (couleur "rusty") et CU (forme "curly") ; couleur de coquille : CE ("celadon") et WE ("œuf blanc").

