

VARIATIONS STRUCTURALES DU GENOME ENTRE DES POPULATIONS SAUVAGES ET DOMESTIQUES DE POULES *GALLUS GALLUS*

Piégu Benoît¹, Arensburger Peter², Beauclair Linda¹, Chabault Marie³, Raynaud Emilie³, Coustham Vincent³, Brard-Fudulea Sophie⁴, Guémené Daniel⁴, Guizard Sébastien^{1,5}, Le Bihan-Duval Elisabeth³, Bigot Yves¹

¹ PRC UMR INRA0085, CNRS 7247, Centre INRA Val de Loire, 37380 Nouzilly - France

² Biological Sciences Department, California State Polytechnic University, Pomona, CA 91768 - United States of America

³ BOA UMR INRA 0083, Centre INRA Val de Loire, 37380 Nouzilly, France

⁴ SYSAAF, Centre INRA Val de Loire, 37380 Nouzilly, France

⁵ Florimond Desprez, 3 Rue Florimond Desprez, 59242 Cappelle en Pélève, France

sophie.brard-fudulea@inra.fr

RESUME

Les travaux réalisés jusqu'à présent pour comprendre au niveau génomique les causes des différences morphologiques, physiologiques et comportementales entre les poules sauvages et les races et lignées de poules domestiques se sont principalement basés sur des données moléculaires de type SNP. Pourtant, il a été montré que d'autres types de variations du génome jouent un rôle dans l'évolution des oiseaux. Cette étude compare les tailles de génome de la Poule rouge de jungle et de plusieurs races et lignées de poule, et recherche l'origine des variations en se concentrant sur les séquences dupliquées (CNV, Copy Number Variation). Des données de séquence ont été obtenues sur 3 espèces de poules sauvages et sur 18 races et lignées de poules domestiques. L'analyse des CNV a été réalisée à l'aide de CNVator, en se limitant aux régions pouvant être alignées sur le génome de référence de la poule. Les résultats montrent que les variations de taille de génome ne peuvent être causées par des éléments transposables. En revanche, des variations structurales significatives sont causées par des ADN_r, des répétitions au niveau des télomères, des centromères, et de l'ADN satellite subtélomérique, ainsi que par des duplications de segments chromosomiques. Une partie de ces variations serait due à la domestication de la poule, et une autre à la sélection réalisée intra-race ou lignée. Compte-tenu de la présence de gènes dans ces régions, il est possible que ces modifications structurelles du génome aient joué un rôle dans l'adaptation des races et lignées de poules aux environnements humains.

ABSTRACT

Variations in genome organization between wild and domesticated lineages of fowls belonging to the *Gallus gallus* species

Efforts to elucidate the causes of morphological, physiological and behavioral differences between wild fowls and their domesticated relatives, the chicken, have so far mainly focused on the identification of single nucleotide mutations. Other types of genomic variations have however been demonstrated to be important in avian evolution, including chicken domestication and associated variations in phenotype. Such variations include several types of sequences duplicated in tandem that can each vary in their level of repetitions. These may be at the origin of genome size differences between related species, sub-species, populations and lines. Here, we report on genome size differences between the red jungle fowl and several domestic chicken breeds and selected lines, and explore the origins of these differences within the repeated portions of these genomes. No differences were found with regard to transposable element abundance. However, sequences duplicated in tandem such as rDNA units, telomere repeats, centromeric and subtelomeric satellite DNA and segmental duplications were found to have been significantly re-shaped, first during domestication and subsequently by human-mediated breeder selection.

INTRODUCTION

Depuis la domestication de la Poule dans le Sud de l'Asie (il y a 8000 ans d'après les données archéologiques, voire 42 à 74 000 ans d'après des données moléculaires, Sawai *et al.*, 2010), sa diffusion dans le reste du monde a conduit à la création de races dites anciennes liées à la localisation géographique des animaux. Les races modernes sélectionnées actuellement sont apparues plus récemment, au cours du 19^e siècle. La domestication peut être considérée comme une adaptation de l'animal à un environnement humain, où la capacité à survivre en milieu sauvage peut être perdue. En plus de ce processus, la sélection mise en œuvre depuis la fin du 19^e siècle s'est initialement concentrée sur l'amélioration de performances de production. Au niveau génétique et génomique, des signatures de sélection (fréquences alléliques très basses/très élevées pour certains gènes) ont été observées, et par comparaison avec la Poule rouge de jungle il a été montré que la diversité génétique avait diminué, et que la proportion de mutations non-silencieuses fixées était élevée. Il y a par contre eu peu de travaux de comparaison structurale du génome de la Poule rouge de jungle et des races domestiques. L'objectif de cette étude est (i) de regarder si de la variabilité de taille de génome existe entre la Poule rouge de jungle et les races domestiques, (ii) de rechercher les causes des diminutions de taille de génome observées, et (iii) de comparer spécifiquement les CNV (copy number variations) et leurs contenus entre Poule rouge de jungle, races anciennes, et lignées sélectionnées.

1. MATERIEL & METHODES

1.1 Echantillons biologiques, mesure de la quantité d'ADN dans les globules rouges et séquençage

Des échantillons de sang ont été prélevés sur 3 races sauvages de poule (Poule rouge de jungle, Poule grise de jungle, Poule sri-lankaise de jungle), et 18 races/lignées de poules domestiques (Alsacienne, Araucana, Barbu, Brahma, Faverolle, Fayoumi, Gournay, Houdan, Orpington, Serama, Silky, lignées de pondeuses à œufs bruns ou blancs). La taille du génome dans les cellules sanguines a été évaluée par cytométrie, avec une taille arbitraire de 1 unité pour la Poule rouge de jungle qui servira de référence pour la comparaison des tailles de génome.

Pour 10 individus (2 Poules rouge de jungle, 2 Araucana, 2 Alsacienne, 2 White Leghorn et 2 Poules grise de jungle), l'ADN a été purifié à partir

de cellules sanguines, et la concentration en ADN des cellules a été estimée par spectrophotométrie et fluorescence. Pour 5 échantillons (un de chacune des races citées précédemment), 5 librairies ont été préparées (longueur moyenne de 550 à 600 pb), et le séquençage MiSeq a été réalisé par la plateforme I2BC (Gif-sur-Yvette, France). La qualité des lectures a été analysée avec FastQC, et elles ont ensuite été filtrées à l'aide de TrimmomaticSE.

1.2 Quantifications des composants des séquences

A partir de ces données, 2 approches ont été mises en œuvre pour identifier les variations de séquences répétées entre les 5 races. Pour la 1^{ère} approche les répétitions ont été recherchées avec les logiciels RepeatExplorer (Novák *et al.*, 2010, 2013) et dnaPipeTE. Pour la 2^{ème} approche, les lectures ont été alignées sur le génome de référence de la Poule avec le logiciel bowtie2 (Langmead *et al.*, 2012). Le comptage du nombre de répétitions pour chaque race par rapport à celles présentes chez la Poule rouge de jungle a été obtenu avec Bioconductor featureCounts (Guizard *et al.*, 2016).

Les variations des niveaux d'ARN ribosomiaux transcrits entre la Poule rouge de jungle et l'Araucana, l'Alsacienne et la White Leghorn ont été déterminés par qPCR.

La variation du nombre de répétitions télomériques du motif TTAGGG présentes chez la Poule rouge de jungle, l'Araucana, l'Alsacienne et la White Leghorn a été estimée par hybridation dot blot (quantification par fluorométrie proche infrarouge). Pour les répétitions en tandem dans les régions sub-télomériques et dans les centromères, une librairie des séquences de ces régions a été créée à partir des séquences disponibles dans les bases de données publiques afin de palier à leur absence dans le génome de référence de la Poule. Les lectures générées dans ce projet ont été alignées sur cette librairie avec le logiciel Bioconductor featureCounts.

1.3 Analyse des CNV

Les CNV sont des fragments d'ADN d'au moins 50 paires de bases dont la perte ou le gain contribue significativement à la variabilité de la taille du génome. L'outil CNVator a été utilisé pour détecter des CNV. Pour tous les génomes utilisés, les données de séquences proviennent de bases de données. Les filtres suivants ont été appliqués pour choisir ces données : alignement d'au moins 90% des séquences sur le génome de référence de la Poule, au moins 90% de lectures uniques après suppression des duplicats de PCR, profondeur de

séquençage moyenne de 8X. Les lectures des jeux de données passant ces filtres ont été alignées sur la version 5 du génome de référence de la Poule à l'aide de bowtie2.

Le calcul de la densité normalisée des lectures a été réalisé par CNVator sur des fenêtres de 100 pb, avec correction pour le taux local en GC. L'analyse avec CNVator a été réalisée en fixant comme paramètres une p-value de détection inférieure à 0,01, et une longueur de CNV d'au moins 0,5 kb, ces paramètres devant limiter le risque de détecter de faux CNV (Yi *et al.*, 2014). Les CNV incluant des régions non-connues du génome de référence ont été exclus, ainsi que les lectures s'alignant sur des régions non-assemblées du génome. De même les chromosomes Z et W n'ont pas été pris en compte car la plupart des séquences venaient d'individus mâles. A l'issue de ces tris sur le génome de référence, il restait 0,9 Gpb dans le génome de référence (1,28 Gpb initialement).

2. RESULTATS ET DISCUSSION

2.1 Longueur de l'ADN

La longueur de l'ADN a été estimée dans 3 races de poules sauvages et dans 18 races et lignées de poules domestiques. La Figure 1 montre qu'il est possible de distinguer 2 groupes. Le 1^{er} groupe est celui des races ayant un génome 5 à 15% plus petit que celui de la Poule rouge de jungle: il s'agit des 2 espèces proches de *Gallus gallus*, de 7 races de poules européennes et américaines, d'une lignée asiatique, et de 2 lignées commerciales de pondeuses. Les tailles de génome des 8 autres races et lignées n'étaient pas significativement différentes de celle de la Poule rouge de jungle. A noter que la taille de génome varie significativement entre les lignées White Leghorn. Ces résultats montrent qu'il existe de la variabilité de taille de génome entre la Poule rouge de jungle et les races anciennes de poules, et qu'il est possible d'observer de la variabilité de taille de génome pour un même type génétique de poule. Pour comprendre les causes de ces variations, la composition structurale des séquences a été étudiée.

2.2 Variations des séquences répétées

Les logiciels RepeatExplorer et dnaPipeTE ont été utilisés pour regarder s'il existe de la variabilité dans le nombre de séquences répétées, quelle que soit leur structure ou leur origine, entre les séquences d'une Poule rouge de jungle, d'une Poule grise de jungle, d'une Araucana, d'une Alsacienne et d'une White Leghorn. L'analyse de ces 5 jeux de données n'a pas révélé de variation significative des proportions en répétitions. De même l'alignement

de ces 5 jeux de séquences sur le génome de référence de la poule n'a pas permis de détecter de différence significative entre les nombres de lectures des différentes races s'alignant sur des séquences répétées.

La qPCR réalisée pour détecter les répétitions en tandem des ADN ribosomiaux chez une Poule rouge de jungle, une Poule grise de jungle, une Araucana, une Alsacienne et une White Leghorn ont mis en évidence des différences significatives : la quantité d'ADN ribosomal est moitié moins importante chez l'Alsacienne et l'Araucana comparées à la Poule rouge de jungle. Chez la Poule grise de jungle et la White Leghorn, la quantité d'ADN ribosomal est ¼ moins importante que celle de la Poule rouge de jungle. Le génome de la Poule rouge de jungle étant composé de 0,5 à 2% d'ADN ribosomal (Su & Delany, 1998 ; Delany *et al.*, 1999), ces différences de quantité d'ADN ribosomal pourraient expliquer en partie les différences de taille de génome observées entre les races.

Chez la Poule, les répétitions télomériques TTAGGG représentent 3 à 4% du génome. L'hybridation dot-plot de l'ADN génomique de 3 Poules rouges de jungle, 2 Poules grises de jungle, 2 Araucana, 2 Alsaciennes et 2 White Leghorn a été réalisé sur un oligonucléotide (TTAGGG)₇. Par rapport à la Poule rouge de jungle, les résultats montrent une part de répétitions télomériques plus faible chez les races domestiques : 1,3% chez l'Alsacienne, 1,2 % chez l'Araucana et les White Leghorn, contre 1,9% chez la Poule rouge de jungle. Chez la Poule grise de jungle, ce taux de répétitions télomériques est un peu plus élevé (2,7%).

L'alignement des lectures d'ADN sur la librairie de séquences répétées en tandem permet de supposer que 5,5 à 6,5% du génome de la Poule rouge de jungle sont composés de ce type de répétitions. Chez les autres races (à l'exception de la White Leghorn), 2 fois moins de lectures alignées ont été obtenues.

A l'issue de l'ensemble de ces analyses, les variations des séquences répétées étudiées pourraient expliquer environ 5% de la réduction de taille de génome observée entre la Poule rouge de jungle et les races domestiques anciennes (environ 2% dus à l'ADN ribosomal, 2 à 4% aux répétitions TTAGGG, et 6% aux séquences dupliquées en tandem dans les centromères et les télomères). Ces variations sont suffisantes pour justifier les différences de tailles de génome constatées entre la Poule rouge de jungle et les White Leghorn, en revanche elles n'expliquent que partiellement les différences observées avec les autres races.

2.3 Analyse des CNV

Après application des filtres, 15 jeux de données publics et privés ont été retenus pour l'analyse des CNV : 3 Araucanas, 3 Poules coréennes, 3 Poules rouges de jungle, 5 White Leghorn, 1 White Plymouth Rock. Deux types de profils pour les CNV ont pu être distingués à l'issue de ces analyses.

Le 1^{er} profil est celui des 3 Poules rouges de jungle : par comparaison avec le génome de référence de *Gallus gallus*, la délétion de 11 à 14 millions de paires de bases a été observée, ainsi que la duplication de 2 à 3 millions de paires de bases. En terme de quantité d'ADN, ces CNV représentent une variation totale de 35 à 45 millions de paires de bases (90% due à des délétions, 10% à des duplications).

Le 2nd profil est celui des races domestiques de poules, chez lesquelles CNVator permet d'observer des délétions de 14 à 30 millions pb, et des duplications de 2,5 à 5 millions pb. En terme de quantité d'ADN, ces CNV représentent une variation totale de 20 à 30 millions de paires de bases (70% due à des délétions, 30% à des duplications). Les duplications de CNV sont 2 fois plus importantes chez la Poule rouge de jungle par rapport aux races domestiques, tandis que les délétions de CNV représentent 2 à 3 fois plus d'ADN chez les races de poules domestiques comparées aux Poules rouges de jungle. Ces résultats montrent que les causes de variation de la quantité d'ADN dans les CNV diffèrent entre les Poules rouges de jungle d'une part et les races domestiques d'autre part.

Cette étude s'est limitée aux autosomes (70% du génome de la Poule rouge de jungle). A l'échelle du génome entier, les CNV devraient expliquer au total 4 à 5% de la différence de quantité d'ADN entre la Poule rouge de jungle et les races domestiques (45 à 55 millions pb, même ordre de grandeur que pour les répétitions en tandem). Au total, les séquences dupliquées et les répétitions en tandem seraient donc responsables de 8 à 10% des différences de tailles de génomes observées entre la Poule rouge de jungle et les races domestiques. Par ailleurs, les résultats montrent que chez les White Leghorn et chez la White Plymouth Rock, il existe des variations interindividuelles significatives (part des CNV dans le génome et nombre de loci affectés par des délétions), suggérant que la variabilité existe inter-races mais aussi intra-race (Figure 2a, 2b).

2.4 Contenu des CNV

Pour l'analyse du contenu des CNV, les échantillons ont été répartis en 5 groupes : les

Poules rouges de jungle, les Poules coréennes domestiques, 3 White Leghorn issue d'une même lignée, et 3 White Leghorn issues de lignées différentes et obtenant des résultats similaires en terme de taux de CNV. Les 4 derniers groupes de cette liste sont considérés comme des poules domestiques, par opposition à la Poule rouge de jungle. Il a été observé 50 à 70% de délétions et duplications de CNV communes aux 5 groupes. Sur ces CNV communs aux 5 groupes, 385 délétions sont communes à la Poule rouge de jungle et aux races domestiques. Au sein des 4 groupes de poules domestiques, 614 délétions communes ont été identifiées, dont 29 non-présentes chez la Poule rouge de jungle. Ce résultat suggère que la domestication et la sélection ont conduit à un très faible nombre de nouvelles délétions dans les CNV par rapport à celles existant déjà chez la Poule rouge de jungle.

La composition des CNV a été étudiée, afin de vérifier si des gènes contenus dans des CNV ont vu leur nombre diminuer ou ont disparu avec la domestication, et si au contraire il existe des gènes qui seraient présents dans les CNV des Poules domestiques et absents des CNV de la Poule rouge de jungle.

Les processus biologiques concernés par les gènes présents dans les CNV ont été étudiés avec le plugin ClueGO du logiciel Cytoscape, utilisé sur la base de donnée *Gallus gallus* 2018 de QuickGO. Pour les CNV ayant subi des délétions, ces analyses ont permis de constater un enrichissement des CNV de la Poule rouge de jungle comparés à ceux des poules domestiques pour 22 gènes impliqués dans 7 voies métaboliques (méiose ovocytaire, signalisation AGE-RAGE dans les complications du diabète, signalisation de l'apeline, récepteur du glutamate, transport lysosomal). Chez les poules domestiques, les CNV sont enrichis pour plusieurs centaines de gènes impliqués dans la régulation de la transcription de l'ADN, le développement des cellules épithéliales, le système immunitaire et le système nerveux. Pour les CNV ayant subi des duplications, un enrichissement significatif a été observé chez l'Araucana pour 10 gènes, dont certains impliqués dans le fonctionnement du système immunitaire. Chez les Poules coréennes, l'enrichissement concerne le fonctionnement du cytosquelette.

CONCLUSION

Les résultats de la cytométrie ont montré que des différences de 5 à 15% pouvaient être observées entre la taille des génomes de la Poule rouge de jungle et des races domestiques de poules, anciennes et récentes. L'analyse des séquences a

confirmé que cette variabilité était principalement due à des répétitions en tandem (8 à 10%). Les délétions constatées dans les CNV permettent de supposer que la domestication a conduit à une modification du nombre de copies de gènes, avec des fragments homozygotes présents chez la Poule rouge de jungle d'une part, et homozygotes absents ou hétérozygotes chez les Poules domestiques d'autre part. Les voies métaboliques concernées par ces gènes sont principalement la régulation de la transcription, le système immunitaire et le développement du cerveau. La réduction du nombre de séquences répétées dans les régions juxtaposées

aux télomères et dans les centromères pourrait, comme chez les mammifères, être la cause de maladies (la réduction des régions sub-télomériques altérant la transcription des gènes voisins d'après Mefford & Trask, 2002). Des études complémentaires basées sur de plus larges jeux de données seraient nécessaires, afin de valider l'hypothèse selon laquelle la perte de séquences répétées en tandem au cours de la domestication aurait réduit la capacité des animaux à s'adapter à de nouveaux environnements et/ou à résister à certaines maladies.

REFERENCES BIBLIOGRAPHIQUES

Delany, M.E., Krupkin, A.B., 1999. Genome. 42, 60-71
Guizard, S., Piégu, B., Arensburger, P., et al., 2016. BMC Genomics. 17, 659
Langmead, B., Salzberg, S. 2012. Nat. Methods. 9, 357-359
Mefford, H.C., Trask, B.J., 2002 Nat. Rev. Genet. 3, 91-102
Novák, P., Neumann, P., Macas, J., 2010. BMC Bioinformatics. 11, 378
Novák, P., Neumann, P., Pech, J., et al., 2013. Bioinformatics. 29, 792-793
Sawai, H., Kim, H.L., Kuno, K., et al. 2010. PLoS One. 5, e10639
Su, M.H., Delany, M.E., 1998. Poult. Sci. 77, 1748-1754
Yi, G., Qu, L., Liu, J., et al., 2014. BMC Genomics.15, 962

Figure 1. Taille de génome en unités arbitraires chez différentes races et lignées de poules (référence : Poule rouge de jungle). WLL : White Layer Line, BLL : Brown Layer Line

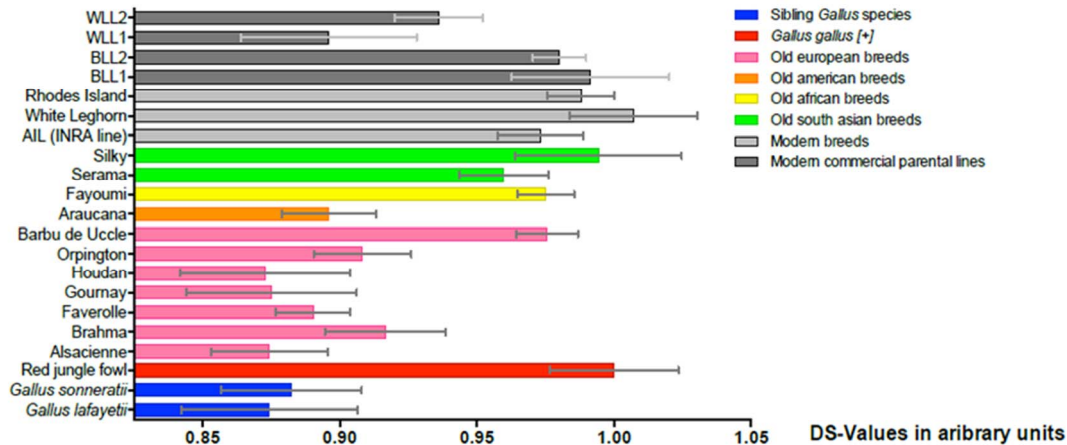


Figure 2. (a) Part des CNV dans les chromosomes de différentes races de poules et (b) nombre de CNV avec délétions d'allèles. WR : White Plymouth Rock, WL : White Leghorn, KD : Poule coréenne, AR : Araucana, RJF : Poule rouge de jungle

