

ÉTUDE DU DÉSÉQUILIBRE DE LIAISON DANS DES LIGNÉES DE POULES DE TYPES GÉNÉTIQUES “PONTE” ET “CHAIR”

Hérault Frédéric¹, Varenne Amandine², Burlot Thierry², Recoquillay Julien³, Macé Camille³, Fagnoul Frédéric³, Picard-Druet David¹, Herry Florian^{1,2}, Allais Sophie¹ et Le Roy Pascale¹

¹PEGASE, INRA, Agrocampus Ouest, 16 Le Clos 35590 Saint-Gilles, France

²NOVOGEN, 5, rue des Compagnons, Secteur du Vau Ballier, 22960 PLEDREAN, France

³HUBBARD, Mauguierand 22800 Le Foeil, France

frederic.herault@inra.fr

RÉSUMÉ

La structure du déséquilibre de liaison (DL) au sein des populations en sélection impacte fortement la précision obtenue lors des études de cartographie de QTL ou lors de l'évaluation génomique des reproducteurs. Chez les oiseaux, la structure hétérogène du génome nécessite de décrire précisément le DL pour optimiser la sélection. L'utilisation des puces SNP haute densité pour le génotypage des populations de volailles est une opportunité pour approfondir notre connaissance de la structure du DL de ces populations. L'objectif de cette étude est d'acquérir une connaissance haute résolution de la structure du DL au sein de populations de poules de types ponte et chair. Nous avons analysé les génotypes (puce 600 K Affymetrix® Axiom® HD SNP) de 1541 animaux issus de 3 populations. L'étendue et le niveau du DL ont été estimés par le r^2 moyen à distance physique donnée entre SNP. Cette étude met en évidence des différences importantes de structure du DL entre lignées et entre chromosomes. L'étendue et le niveau du DL sont plus importants dans les lignées de type ponte ou pour les macro-chromosomes et le chromosome Z. Ce niveau important de DL peut faciliter la détection de QTL sur ces chromosomes, mais peut également compliquer la localisation fine de polymorphismes causaux. À l'inverse, le faible niveau de DL observé sur les micro-chromosomes nécessite l'utilisation d'une forte densité de SNP pour détecter une association avec un phénotype, mais devrait permettre la cartographie fine d'un polymorphisme causal. Ces différences sont à prendre en considération pour définir une stratégie de génotypage économique et efficace pour la cartographie fine de QTL ou l'évaluation génomique.

ABSTRACT

A Linkage disequilibrium study in layers and broiler commercial chicken populations.

Knowledge of the linkage disequilibrium (LD) pattern is useful in animal genetic studies as it underlies mapping studies and genomic selection. This is all the more important in birds given the heterogeneous structure of the avian karyotype. Recently, the availability of the high density 600 K Affymetrix® Axiom® HD SNP genotyping array allowed to assess an in-depth knowledge of the LD pattern in chicken genome. The aim of the present study was to assess a higher resolution of the LD pattern in chicken genome in layer and broiler lines. In this study, we analyzed genotypes of 1541 animals from layers and broiler commercial populations to characterize their LD pattern. LD was measured by the average r^2 value at a given physical distance between SNP. LD extended over a larger region for layer lines than for broiler line. Most differences between lines appeared at small interval distances (< 0.5Mb). LD extent and decay differed considerably between chromosomes categories. Average r^2 values were higher for Z chromosome than for macro, intermediates and micro-chromosomes. The extent of useful LD observed for autosomal chromosomes was at least ten-fold longer for layer lines than for broiler. Finally, this study shed light on high LD for Z chromosome. The differences in LD pattern observed between chromosomes and chicken lines should be taken into account to define an economically efficient genotyping strategy.

INTRODUCTION

La structure du déséquilibre de liaison (DL) au sein des populations en sélection impacte fortement la précision obtenue lors des études de cartographie de QTL ou lors de l'évaluation génomique des reproducteurs. En se basant sur le génotype de plusieurs centaines de milliers de marqueurs, les études d'association à l'échelle du génome (GWAS) reposent sur l'étendue et le niveau du déséquilibre de liaison (DL) entre marqueur et polymorphisme causal pour aboutir à une localisation fine de QTL. Le DL est également un des paramètres affectant la précision des évaluations génomiques. Une connaissance approfondie du DL est donc un préalable pour parvenir à une prédiction fiable de la valeur génétique des candidats à la reproduction et pour maximiser le gain génétique en sélection génomique (Meuwissen et al., 2001). D'autre part, l'étendue du DL est également un paramètre important pour la sélection d'un sous-ensemble de marqueurs pour la réalisation d'une puce basse densité nécessaire pour la mise en œuvre de stratégies de génotypage à large échelle. Depuis 2013, une puce haute densité (600 K Affymetrix® Axiom® HD SNP genotyping array) est disponible pour le génotypage des poules. Ce support de génotypage a été réalisé en se basant sur 24 lignées de poules (lignées expérimentales ou commerciales de types pontes et chairs) de manière à assurer une distribution uniforme (sur la base d'une carte génétique) de marqueurs informatifs quel que soit le type d'individu génotypé (Kranis et al., 2013).

Chez les oiseaux, la structure du génome est particulièrement hétérogène. Chez la poule le génome est composé de 38 autosomes et d'une paire de chromosomes sexuels. Les autosomes sont organisés en chromosomes de grande taille ou macro-chromosomes qui contiennent de 40 – 200 millions de paires de bases (Mb) (chromosomes 1 à 5), chromosomes de taille intermédiaire de 20 – 40 Mb (chromosomes 6 à 10) et de micro-chromosomes dont la taille est inférieure à 20 Mb (chromosomes 11 à 38). Le taux de recombinaison et la densité génique de ces micro-chromosomes sont supérieurs à ceux des autres catégories de chromosomes (International Chicken Genome Sequencing Consortium, 2004).

L'objectif de cette étude est d'acquérir une connaissance haute résolution de la structure du DL au sein de populations de poules de types génétiques ponte et chair en nous référant aux génotypes obtenus à partir d'une puce haute densité (600 K Affymetrix® Axiom® HD SNP genotyping array).

1. MATERIELS ET METHODES

1.1. Génotype

Les génotypages de 1541 animaux provenant de deux lignées commerciales de poules pondeuses et d'une lignée commerciale de poulet de chair ont été utilisés dans le cadre de cette étude. Les deux lignées de poule pondeuses proviennent de la société NOVOGEN. Il s'agit d'une lignée Leghorn à œuf blanc (L, 771 individus) et d'une lignée Rhode-Island (RI, 447 individus) à œuf brun. La lignée de type chair, lignée mâle à croissance rapide (Shaver Yellow Cornish (SYC), 383 individus), vient quant à elle de la société HUBBARD. Les génotypages ont été effectués avec le support 600 K Affymetrix® Axiom® HD SNP genotyping array. Ce support de génotypage contient 580 961 SNP répartis sur les chromosomes 1 à 28, le chromosome 33, le groupe de liaison LGE64 ainsi que sur les deux chromosomes sexuels. Parmi ces SNP, 3 724 n'ont pas pu être localisés. La localisation des marqueurs a été effectuée sur la version 5 de l'assemblage du génome de *Gallus gallus*.

Au sein de chaque lignée la qualité des données de génotypages a été contrôlée avec le logiciel PLINK (PLINK version 1.9; www.cog-genomics.org/plink/1.9/) (Chang et al., 2015). Les SNP dont le taux d'assignation de génotype (call rate) est inférieur à 95%, la fréquence de l'allèle minoritaire (MAF) est inférieure à 0,05 ou s'éloignant trop de l'équilibre de Hardy-Weinberg (HWE) (P-value <0,0001) ont été supprimés de l'analyse. Pour réaliser cette étude, seuls les SNP possédant une assignation chromosomique ont été utilisés. Les SNP localisés sur les chromosomes W, 16, 33 et sur le groupe de liaison LGE-64 ont également été supprimés de l'analyse, du fait du faible nombre de marqueurs localisés sur ces chromosomes. L'ensemble des individus de l'analyse ont été conservés et avaient un call rate > 95%.

1.2. Estimation du déséquilibre de liaison

Le DL, qui est caractéristique d'une population, décrit les associations non aléatoires entre allèles à des loci différents. Le niveau de DL peut être mesuré par le r^2 qui mesure la corrélation entre allèles de deux loci. Ainsi deux marqueurs ayant un r^2 de 1 sont totalement équivalents et l'un peut être utilisé comme substitut de l'autre (l'information contenue est équivalente). Dans le cadre de cette étude, le r^2 a été utilisé comme mesure du DL entre chaque paire de SNP au sein d'un chromosome pour chaque lignée étudiée. Cette mesure a été effectuée avec le logiciel PLINK. Le niveau et l'étendue du DL ont été calculés pour chaque catégorie de chromosome selon la classification établie par l'International Chicken Genome Sequencing Consortium, à savoir les macro-chromosomes, les chromosomes intermédiaires, les micro-chromosomes et le chromosome sexuel Z. Au sein de chaque lignée, la structure du DL a été estimée pour chaque catégorie de chromosome comme étant la valeur moyenne du r^2 pour une distance physique donnée entre SNP.

2. RESULTATS ET DISCUSSION

2.1. Marqueurs

Comme le montre le Tableau 1, la plupart des SNP non conservés dans l'étude ont été supprimés du fait de leurs faible MAF. Il apparaît clairement dans ce tableau que les lignées de type ponte ont une proportion de SNP à faible MAF plus importante que la lignée de type chair montrant une quasi-fixation d'un nombre plus important de SNP dans ces lignées. La lignée Leghorn est la lignée qui possède le plus petit nombre de marqueurs informatifs, tandis que la lignée SYC est la lignée possédant le plus grand nombre de marqueurs informatifs. Ce résultat démontre que la lignée chair est plus polymorphe que les lignées de type ponte. Cette différence de variabilité génétique peut s'expliquer d'une part par l'histoire de la lignée (création plus ou moins récente) et d'autre part par le nombre de reproducteurs non apparentés utilisés dans chacune des lignées. Ces résultats corroborent les résultats d'études précédentes (Megens et al., 2009; Qanbari et al., 2010). Malgré de grandes différences dans le nombre de marqueurs informatifs retenus pour chaque lignée, la couverture du génome est sensiblement la même pour les trois lignées considérées (environ 1,02 Gb). La distance moyenne entre deux marqueurs informatifs consécutifs est inférieure à 5 kb pour toutes les lignées.

2.2. Analyse du déséquilibre de liaison

L'étendue et la décroissance du DL ont été estimées pour chaque lignée en fonction de la distance physique séparant les marqueurs (Tableau 2).

Dans les trois lignées étudiées, le DL décroît à mesure que la distance séparant les marqueurs augmente, pour atteindre 0,02 quand la distance entre les SNP considérés est supérieure à 3 Mb pour la lignée chair ou supérieure à 5 Mb pour les lignées ponte. Le niveau du DL varie énormément entre les lignées considérées, et cela d'autant plus que les distances entre marqueurs sont petites. La lignée Leghorn et la lignée chair ont respectivement le niveau de DL moyen le plus fort et le plus faible quel que soit la distance inter-marqueurs considérée et ce malgré le fait que la distance moyenne entre deux marqueurs adjacents soit plus importante dans les lignées ponte que dans la lignée chair.

L'étendue et la décroissance du DL ont également été analysées intra-lignée en fonction des catégories de chromosomes. Comme le montre la Figure 1, l'étendue et la décroissance du DL dépendent grandement de la catégorie de chromosome considérée. Quelle que soit la lignée considérée, le niveau du DL est plus élevé et s'étend sur une plus grande distance pour le chromosome Z puis pour les macro-chromosomes que pour les chromosomes intermédiaires et les micro-chromosomes. Le niveau

et l'étendue du DL sont plus importants dans la lignée Leghorn que dans la lignée RI et la lignée SYC.

De manière générale, un DL moyen supérieur à 0,3 est considéré comme suffisant et donc utile pour mener des analyses d'association ou des évaluations en sélection génomique (Ardlie et al., 2002; Calus et al., 2008). De ce point de vue, comme le montre le Tableau 3, l'étendue du DL utile ($DL > 0.3$) pour l'ensemble des autosomes est très similaire pour les deux lignées ponte. Au niveau du génome l'étendue du DL utile se situe entre 200 et 300 kb pour les lignées ponte et entre 20 et 30 kb pour la lignée chair. A titre de comparaison, l'étendue du DL utile se situe entre 280 et 413 kb dans les races porcines occidentales (respectivement Large White et Duroc) et entre 33 kb et 40 kb pour les races bovines (Holstein et Angus) (Ai et al., 2013; Pérez O'Brien et al., 2014). Au niveau des chromosomes de grande taille (macro et intermédiaire) l'étendue du DL utile dans les lignées ponte est environ 10 fois plus grande que chez la lignée chair. Cet écart est plus important encore si l'on considère les micro-chromosomes pour lesquels apparaît également une légère différence entre les lignées ponte. Cette différence d'étendue du DL utile au niveau des micro-chromosomes pourrait s'expliquer par le fait que le taux de recombinaison est plus important pour ces chromosomes de petites tailles que pour les autres autosomes. Cette importante différence d'étendue du DL utile entre lignées n'est toutefois pas observée au niveau du chromosome Z. L'étendue du DL utile sur le chromosome Z est de l'ordre de 500-1500 kb pour la lignée L, 400-500 kb pour la lignée RI et 300-400 kb pour la lignée chair. La différence d'étendue du DL utile entre le chromosome Z et les chromosomes de grande taille (macro et intermédiaire) est relativement faible pour les lignées ponte, elle est d'un facteur 10 pour la lignée chair.

CONCLUSION

Dans le cadre de cette étude nous avons réalisé une caractérisation haute résolution de la structure du DL de trois lignées commerciales de poules de types génétiques ponte et chair en nous basant sur des génotypes obtenus à partir d'une puce haute densité (600K Affymetrix® Axiom® HD SNP genotyping array). L'étendue et le niveau de DL observés varient grandement entre lignées et entre catégories de chromosomes. Les différences observées entre lignées sont nettement plus marquées entre types génétiques ponte et chair qu'entre les deux lignées de type ponte. L'étendue importante du DL pour le chromosome Z ainsi que pour les macro-chromosomes peut faciliter la détection de QTL sur ces chromosomes. Cependant cette caractéristique du DL va rendre plus compliquée la localisation fine des polymorphismes causaux. A l'inverse, sur les micro-chromosomes, le faible niveau de DL observé impose l'utilisation d'une carte plus dense en marqueurs afin de pouvoir identifier une

association marqueurs / phénotype, mais devrait être un avantage pour la cartographie fine d'un polymorphisme causal. Ces différences dans la structure du DL entre différents types génétiques et catégories de chromosomes sont à prendre en considération pour définir une stratégie de génotypage économique et efficace pour la cartographie fine de

QTL ou l'évaluation génomique. Concrètement, dans le cadre de la réalisation d'une puce basse densité pour la mise en œuvre d'un génotypage de routine à grande échelle, ces résultats impliquent une sélection de marqueurs différenciée (nombre et localisation) en fonction des lignées à génotyper.

REFERENCES BIBLIOGRAPHIQUES

- Ai, H., Huang, L., Ren, J., 2013. PLOS ONE 8, e56001.
- Ardlie, K.G., Kruglyak, L., Seielstad, M., 2002. Nat Rev Genet 3, nrg777.
- Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W., Veerkamp, R.F., 2008. Genetics 178, 553–561.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. GigaScience 4, 7.
- International Chicken Genome Sequencing Consortium, 2004. Nature 432, nature03154.
- Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P., Hocking, P.M., Fife, M., Salmon, N., Fulton, J., Strom, T.M., Haberer, G., Weigend, S., Preisinger, R., Gholami, M., Qanbari, S., Simianer, H., Watson, K.A., Woolliams, J.A., Burt, D.W., 2013. BMC Genomics 14, 59.
- Megens, H.-J., Crooijmans, R.P., Bastiaansen, J.W., Kerstens, H.H., Coster, A., Jalving, R., Vereijken, A., Silva, P., Muir, W.M., Cheng, H.H., Hanotte, O., Groenen, M.A.M., 2009. BMC Genet 10, 86.
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Genetics 157, 1819–1829.
- Pérez O'Brien, A.M., Mészáros, G., Utsunomiya, Y.T., Sonstegard, T.S., Garcia, J.F., Van Tassell, C.P., Carvalho, R., da Silva, M.V.B., Sölkner, J., 2014. Livestock Science, Genomics Applied to Livestock Production 166, 121–132.
- Qanbari, S., Hansen, M., Weigend, S., Preisinger, R., Simianer, H., 2010. BMC Genet 11, 103.

Tableau 1. Description des marqueurs pour chaque lignée.

	Ponte		Chair
	L	RI	SYC
Nombre total de SNP	580 961	580 961	580 961
SNP Call rate < 95%	2 530	7 541	4 274
MAF ¹ < 0,05	327 452	258 772	181 584
HWE ² < 0,0001	3 857	12 538	6 664
SNP conservés	245 669 (42%)	300 362 (52%)	386 187 (66%)
Couverture du génome (Mb)	1 017	1 017	1 019
Distance moyenne entre SNP adjacent (kb)	4,1	3,4	2,6

¹ MAF: Nombre de SNP ayant une MAF <0,05² HWE: Nombre de SNP s'écartant de l'équilibre de Hardy-Weinberg P-value <0,0001.**Tableau 2.** Niveau du déséquilibre de liaison en fonction de la distance physique entre SNP.

	Ponte		Chair
	L	RI	SYC
Distance (Mb) ¹	Moyenne ± SD	Moyenne ± SD	Moyenne ± SD
0.0-0.025	0.62 ± 0.37	0.52 ± 0.38	0.34 ± 0.33
0.025-0.05	0.53 ± 0.37	0.46 ± 0.37	0.28 ± 0.29
0.05-0.075	0.49 ± 0.36	0.43 ± 0.36	0.25 ± 0.28
0.075-0.12	0.44 ± 0.35	0.39 ± 0.35	0.23 ± 0.26
0.12-0.2	0.39 ± 0.34	0.35 ± 0.33	0.2 ± 0.24
0.2-0.5	0.29 ± 0.3	0.27 ± 0.29	0.15 ± 0.2
0.5-1.5	0.16 ± 0.22	0.15 ± 0.21	0.08 ± 0.14
1.5-3.0	0.07 ± 0.14	0.07 ± 0.14	0.04 ± 0.09
3.0-5.0	0.04 ± 0.09	0.04 ± 0.09	0.02 ± 0.05
5.0-10.0	0.02 ± 0.05	0.02 ± 0.05	0.01 ± 0.02

¹ Distance (Mb): basé sur l'assemblage de référence Gallus gallus - 5.0.**Tableau 3.** Etendue du déséquilibre de liaison utile ($r^2 > 0.3$).

	Ponte		Chair
	L	RI	SYC
Génome	200-300 kb	200-300 kb	20-30 kb
Macro-chromosomes	300-400 kb	400-500 kb	30-40 kb
Chromosomes intermédiaires	200-300 kb	200-300 kb	20-30 kb
Micro-chromosomes	100-200 kb	90-100 kb	5-7.5 kb
Chromosome Z	500-1500 kb	400-500 kb	300-400 kb

Figure 1. Décroissance du déséquilibre de liaison en fonction de la distance entre SNP par catégorie de chromosome.