



HAL
open science

Génotypage en haute-résolution d'une région très polymorphe impliquée dans la réponse aux pathogènes : le CMH (Complexe Majeur d'Histocompatibilité)

Bertrand Bed'Hom, Olympe Chazara

► **To cite this version:**

Bertrand Bed'Hom, Olympe Chazara. Génotypage en haute-résolution d'une région très polymorphe impliquée dans la réponse aux pathogènes : le CMH (Complexe Majeur d'Histocompatibilité). 9. Journées de la Recherche Avicole, Mar 2011, Tours, France. pp.779-783. hal-01019047

HAL Id: hal-01019047

<https://hal.science/hal-01019047>

Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

GENOTYPAGE EN HAUTE-RESOLUTION D'UNE REGION TRES POLYMORPHE IMPLIQUEE DANS LA REponse AUX PATHOGENES : LE CMH (COMPLEXE MAJEUR D'HISTOCOMPATIBILITE)

Bed'Hom Bertrand¹, Chazara Olympe²

¹*INRA Génétique Animale et Biologie Intégrative - Domaine de Vilvert - 78352 JOUY EN
JOSAS*

²*Cambridge University, Department of Pathology - Tennis court road - CB2 1QP
CAMBRIDGE (UK)
bertrand.bedhom@jouy.inra.fr*

RESUME

La région génomique du CMH (Complexe Majeur d'Histocompatibilité) est un locus déterminant pour le contrôle des maladies infectieuses chez les Vertébrés. C'est une région très dense en gènes, codant notamment pour des molécules qui présentent les antigènes aux lymphocytes T. Certains de ces gènes montrent des niveaux de polymorphisme très importants. La caractérisation des allèles du CMH est cruciale pour les études de variabilité génétique de réponse aux pathogènes. Historiquement, de nombreux allèles du CMH ont été identifiés chez la Poule par sérologie, puis avec un marqueur ADN très polymorphe. Afin de couvrir l'ensemble du CMH, d'améliorer la précision de génotypage et de comprendre les relations évolutives entre les allèles, un jeu de 96 marqueurs de type SNP (Single Nucleotide Polymorphism) a été développé. Le choix des marqueurs a été réalisé à partir d'une liste de SNPs identifiés par comparaison des séquences déjà disponibles et par une approche de reséquençage spécifique. Environ 1900 échantillons ont été génotypés avec ce panel et les résultats ont confirmé la grande diversité du CMH, avec un très grand nombre d'allèles différents déterminés. La corrélation entre tous les SNPs est très faible, ce qui valide l'utilisation d'un jeu dense de SNPs pour caractériser le CMH. Les différents allèles permettent d'établir des relations claires avec les données de sérologie, en améliorant la précision de génotypage. Les données obtenues permettent également d'établir des relations évolutives entre les différents allèles, y compris par des événements de recombinaison. Le polymorphisme du CMH est bien évalué par ce jeu de 96 SNPs : la dissection fine de la grande variabilité génétique de cette région sera très importante dans les projets d'étude de la réponse aux pathogènes chez la Poule.

ABSTRACT

High resolution genotyping of a very polymorphic region involved in response to pathogens : MHC (Major Histocompatibility Complex)

The genomic region of MHC (Major Histocompatibility Complex) is a locus decisive for the control of infectious diseases in Vertebrates. This is a very gene-rich region, coding especially for antigen-presenting molecules to T-cells. Some genes display very high levels of polymorphism. The characterization of MHC alleles is crucial for studies of genetic variability of response to pathogens. Historically, many MHC alleles have been identified in Chicken by serology, using a highly polymorphic DNA marker. In order to cover entirely the MHC, to improve genotyping accuracy and to understand evolutive relationships between alleles, a set of 96 SNP (Single Nucleotide Polymorphism) markers has been developed. The choice of markers has been performed from a list of SNPs identified by comparison of sequences yet available and by a targeted resequencing approach. Around 1900 samples have been genotyped with this set and the results are confirming the high MHC diversity, with many different alleles determined. The correlation between all the SNPs is very low, which confirms the use of dense SNP panel to characterize MHC. The different alleles allow the establishment of clear relationships with serological data, and even improve genotyping accuracy. The data obtained allow as well establishing evolutive relationships between different alleles, including by recombination events. The MHC polymorphism is correctly evaluated by this set of 96 SNPs : the fine dissection of the genetic variability of this region will be very important for the projects studying the response to pathogens in Chicken.

INTRODUCTION

Les SNPs (Single Nucleotide Polymorphism), des variations d'une seule base dans la séquence du génome, sont les polymorphismes décrits les plus fréquents. Ils permettent d'obtenir une meilleure résolution et une meilleure couverture du génome que les marqueurs microsatellites pour les analyses génétiques. Cela est particulièrement vrai pour la région B du MHC (Major Histocompatibility Complex) de la Poule, où une seule séquence répétée atypique est utilisée comme marqueur pour une région de 240 kb (Fulton et al., 2006). Cette faible densité de séquences répétées utilisables comme marqueurs est probablement lié aux caractéristiques particulières du MHC chez la Poule, compact, avec des introns et distances intergéniques courts, et considéré comme « minimal et essentiel » par Kaufman et al. (1999). Cette région du MHC contient, comme chez les autres Vertébrés, quelques-uns des gènes les plus polymorphes du génome, comme les gènes de classe I et de classe II, codant pour des protéines responsables de la présentation des antigènes aux lymphocytes T.

Le grand nombre de SNPs présents dans la région du MHC pourrait donc être une solution pour son génotypage. C'est particulièrement intéressant chez la Poule, dont la région B a été associée avec résistance ou susceptibilité à des pathogènes et des réponses différentielles à des vaccins (Pinard-van der Laan et al., 1998 ; Lee et al., 2004). Ces résultats ont été démontrés avec un génotypage par sérologie, une technique lourde, difficilement standardisable et qui ne permet pas de comparaison inter-races.

Des SNPs ont été décrits dans le génome de la Poule depuis le reséquençage partiel par le BGI (Beijing Genome Institute) du génome de plusieurs races (ICPMC, 2004). Le chromosome 16, dont l'assemblage génomique ne comprend que la région du MHC, comporte la plus grande densité de SNPs du génome, ce qui permet donc de disposer de nombreux marqueurs potentiels. Cependant, cette région a également des particularités défavorables, comme les duplications de gènes et les familles multigéniques, ainsi que de très faibles distances entre SNPs, particulièrement pour les gènes présentant un très fort taux de polymorphisme. C'est un problème pour certaines technologies de génotypage de SNPs à haut débit, comme Illumina GoldenGate, qui requiert qu'aucun polymorphisme ne soit présent dans les 60 pb flanquant le SNP à génotyper.

Nous décrivons ici la méthode utilisée pour sélectionner les 96 SNPs composant le panel de génotypage de la région B du MHC, et les premiers résultats du génotypage de ce panel sur un système BeadXpress d'Illumina.

1. MATERIELS ET METHODES

1.1. Caractérisation de SNPs par reséquençage ciblé

Des zones conservées dans la région génomique du MHC ont été identifiées, afin de déterminer des amorces permettant d'amplifier des 24 fragments ADN correspondant à 11 gènes cibles (*TRIM27.1*, *TRIM41*, *B-BTN2*, *Blec1*, *Tapasin*, *B-DMA*, *B-DMB2*, *B-DMB1*, *TAP1*, *TAP2* et *CDIA1*). Les cibles ont été choisies sur la base des données déjà disponibles concernant leur polymorphisme (Shiina et al., 2007), mais aussi en fonction de la faisabilité technique.

Après amplification à partir des ADN de 96 individus appartenant à 47 populations, les fragments de PCR (de 306 pb à 1179 pb) ont été séquencés au CNS (Centre National de Séquençage). Les populations utilisées comprenaient 6 lignées commerciales (4 lignées chair et 2 lignées pondeuses), 16 populations White Leghorn (15 lignées expérimentales INRA et une population commerciale) et 25 races (12 races européennes, 7 races asiatiques et 6 populations locales africaines), chaque population étant représentée par 2 animaux.

Pour chacune des 24 PCR (plusieurs PCR par gène), l'analyse de la qualité des séquences et la détermination des variations a été conduite pour les 96 individus avec NovoSNP (Weckx et al., 2005). Seuls les SNPs ayant une forte probabilité d'être vrais (score NovoSNP supérieur ou égal à 12) ont été conservés pour les analyses ultérieures. Les SNPs ont été ensuite confirmés manuellement par contrôle des alignements d'électrophorogrammes.

1.2. Autres sources de SNPs

Tous les SNPs de la région B du MHC de la Poule, selon la séquence du MHC de Shiina et al. (2007, AB268588, 241 kb) ont été extraits de la base de données GenBank, et les caractéristiques de leur polymorphisme compilés à partir du navigateur de génome UCSC (genome.ucsc.edu), quand elles étaient disponibles. C'était le cas de la majorité des SNPs du BGI, obtenus par reséquençage partiel de 3 races (une poule pondeuse, un poulet de chair et un poulet Silkie). Par ailleurs, les 14 séquences homozygotes d'une partie de la région B de différents haplotypes MHC décrits par Hosomichi et al. (2008) ont été alignés sur la séquence AB268588, et les positions polymorphes déduites à l'aide du logiciel BioEdit (Hall, 1999).

1.3. Présélection des SNPs

Seuls les SNPs bi-alléliques ont été conservés (quelques SNPs présentent plus de 2 allèles). Puis les séquences flanquantes (100 pb en amont et en aval) de chaque SNP ont été utilisées pour une recherche de séquence par BLAST (Altschul et al., 1990) sur la séquence du génome et la séquence de référence AB268588. Seuls les SNPs dont les séquences flanquantes ont une localisation unique dans le

génomique ont été conservés. Ensuite les SNPs sans autre polymorphisme (incluant les indels et les SNPs) dans les 40 pb autour de la position du SNP ont été sélectionnés. Comme certaines sous-régions étaient mal couvertes (particulièrement les régions comportant peu de polymorphismes utilisables ou des répétitions), ces régions ont été utilisées pour cribler les bases dbEST de GenBank pour obtenir et aligner des séquences de transcrits de ces sous-régions. Quelques SNPs additionnels ont été identifiés à partir de ces alignements.

1.4. Soumission technique et choix final

Les SNPs retenus ont été soumis pour une validation technique de la possibilité de génotypage avec la technologie GoldenGate auprès d'Illumina. Le score recommandé (note technique Illumina pour le génotypage GoldenGate) pour les SNPs est d'au moins 0,4 : des SNPs ayant un score supérieur à 0,4 ont été choisis sur la base de leur distribution sur la séquence AB268588 et de leur polymorphisme connu pour composer un jeu de 96 SNPs.

1.5. Génotypage des SNPs

Le panel de 96 SNPs a été génotypé sur un peu plus de 1900 échantillons d'ADN à la plateforme génomique PlaGe (Toulouse) sur un appareil Illumina BeadXPress, avec la technologie GoldenGate. Les échantillons proviennent surtout de Poule homozygotes pour des sérotypes déterminés. Les fichiers d'intensité obtenus ont été analysés avec le logiciel GenomeStudio (Illumina), pour déterminer les allèles des SNPs. La visualisation des SNPs a été faite avec le logiciel Flapjack (Plant Bioinformatics Group, SCRI), et l'analyse de corrélation entre SNPs avec le logiciel Haploview (Barrett et al., 2005).

2. RESULTATS ET DISCUSSION

2.1. Jeu de 96 SNPs

A partir de plus de 20 kb de séquences analysées par réséquençage ciblé de 47 populations, 594 SNPs ont été identifiés, distribués de manière hétérogène sur plus de 147 kb. La comparaison des 14 séquences de différents sérotypes avec la séquence de référence AB268588 a permis d'identifier 4201 SNPs pour 57 kb de séquences. Enfin, 24 SNPs additionnels proviennent des alignements de séquences ESTs et de la base de donnée dbSNP. A partir de ces 4819 SNPs, l'utilisation des critères de sélection décrits a réduit à 179 le nombre de SNPs soumis pour validation technique à Illumina. En fonction des notes obtenues pour ces SNPs, ils ont été classés en différentes catégories : 5 SNPs avec une mauvaise note technique (entre 0,245 et 0,372), 28 SNPs avec une note

moyenne (entre 0,4 et 0,599) et 146 SNPs avec une bonne note technique (entre 0,6 et 0,993).

Nous avons conservé 96 SNPs (Figure 1), avec des notes techniques comprises entre 0,45 et 0,993, répartis sur 241312 pb, ce qui représente en moyenne 1 SNP pour 2256 pb. La distance minimale entre 2 SNPs est de 79 pb, et la distance maximale de 11763 pb (Figure 2B). Les grandes distances sans SNP sont observées dans les régions où les critères utilisés ne permettaient de conserver aucun SNP : cela correspond à des régions de gènes de familles multigéniques (B-G1, B-G2 et B-G3 ; B-LB1 et B-LB2 ; B-F1 et B-F2), mais aussi à des gènes uniques comportant des structures internes répétées, comme TNXB ou C4.

2.2. Génotypage des SNPs

Le génotypage du jeu de 96 SNPs a donné de bons résultats : le taux données manquantes est de 3,5%. Plus de la moitié d'entre elles correspondent à seulement 4 marqueurs dont les taux de données manquantes sont supérieurs à 10%. Les quelques échecs correspondent à des problèmes techniques lors du génotypage, où à sa mauvaise résolution, ne permettant pas une détermination univoque des allèles. L'hétérozygotie moyenne sur l'ensemble des marqueurs et des échantillons est inférieure à 15%. Ce résultat est la conséquence du génotypage d'échantillons pour la plupart homozygotes. La distribution de la fréquence de l'allèle mineur (MAF) est relativement linéaire, entre 0 et 50% : l'informativité est donc assez différente d'un SNP à l'autre (Figure 2A). L'étude de la corrélation (r^2) à l'aide du logiciel Haploview révèle que le niveau de corrélation entre SNPs consécutifs est particulièrement bas (Figure 3) : seules 5 paires de SNPs consécutifs ont une corrélation supérieure à 0,5, ce qui est très inhabituel au regard de la faible distance (de l'ordre de 2 kb) entre SNPs consécutifs. On peut noter également quelques corrélations à plus longue distance pour quelques paires de SNPs, ce qui peut suggérer des co-évolutions entre gènes distants, ou des conversions géniques.

2.1. Discussion

Ce jeu de 96 SNPs permet de capturer de manière beaucoup plus dense la variabilité génétique de la région du MHC. La densité importante de marqueur se révèle nécessaire pour le génotypage, et on peut noter que le niveau de corrélation entre SNPs est très faible au regard de la densité de marqueurs, conséquence du grand nombre d'allèles existant dans les populations pour le locus MHC. L'analyse des premiers résultats permet de montrer que les haplotypes obtenus ont une très bonne corrélation avec les données historiques de la sérologie et du génotypage avec un des seuls marqueurs polymorphe disponible jusqu'ici. De plus, le niveau de résolution

obtenu est meilleur, car certains échantillons non différenciables par la sérologie ou le marqueur polymorphe ont des haplotypes distincts avec le panel SNP. De même, on peut prédire l'appartenance d'échantillons sans information préalable à des groupes sérologiques connus, y compris dans des populations sans donnée sérologique préliminaire, ce qui est le cas de nombreuses races ou populations locales dans le monde. Enfin, ce jeu de SNP permet de décrire des relations évolutives entre allèles, en mettant en évidence des événements historiques de recombinaison ou de conversion génique durant l'évolution du locus MHC. Par exemple, la comparaison des haplotypes B12 et B19 montrent qu'ils ne diffèrent que par l'extrémité du MHC correspondant à la région de classe III. De manière contrastée, l'haplotype B13 présente un profil SNP très différent des deux précédents (Figure 4).

CONCLUSION

La région du MHC est complexe, avec un haut niveau de diversité génétique, en relation avec les modalités évolutives qui gouvernent cette région impliquée dans

la réponse aux pathogènes. Cette diversité peut être explorée efficacement avec un panel de 96 SNPs développé dans ce but. Les génotypes obtenus permettent d'obtenir des corrélations avec la sérologie, d'améliorer les méthodes précédentes de génotypage moléculaire et de déterminer les relations évolutives entre les allèles. Les résultats montrent que le nombre d'haplotypes différents est considérable, et que l'évolution à ce locus est la conséquence d'événements de recombinaison et probablement également de conversion génique. Ce panel devrait permettre de disséquer plus finement les variations génétiques dans la région du MHC pour des projets en rapport avec la résistance aux pathogènes.

REMERCIEMENTS

Merci à Nicolas Bruneau pour la préparation des échantillons, à Diane Esquerré pour en avoir assuré le génotypage à PlaGe, et aux fournisseurs d'échantillons génotypés (Janet Fulton, Helle Juul-Madsen, Chi-Sheng Chang, Pascale Quéré, Fanny Calenge, Pete Kaiser, Olivier Hanotte, Anna Slawinska).

REFERENCES BIBLIOGRAPHIQUES

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J., 1990. *J. Mol. Biol.*, (215), 403-410.
 Barrett J.C., Fry B., Maller J., Daly M.J., 2005. *Bioinformatics*, (21), 263-265.
 Fulton J.E., Juul-Madsen H.R., Ashwell C.M., McCarron A.M., Arthur J.A., O'Sullivan N.P., Taylor Jr R.P., 2006. *Immunogenetics*, (58), 407-421.
 Hall T.A., 1999. *Nucl. Acids. Symp. Ser.*, (41), 95-98.
 Hosomichi K., Miller M.M, Goto R.M., Wang Y., Suzuki S., Kulski J.K., Nishibori M., Inoko H., Hanzawa K., Shiina T., 2008. *J. Immunol.*, (181), 3393-3399.
 ICPMC (International Chicken Polymorphism Map Consortium), 2004. *Nature*, (432), 717-722.
 Kaufman J., Milne S., Göbel T.W., Walker B.A., Jacob J.P., Auffray C., Zoorob R., Beck S., 1999. *Nature*, (401), 923-925.
 Shiina T., Briles W.E., Goto R.M., Hosomichi K., Yanagiya K., Shimizu S., Inoko H., Miller M.M., 2007. *J. Immunol.*, (178), 7162-7172.
 Weckx S., Del-Favero J., Rademakers R., Claes L., Cruts M., De Jonghe P., Van Broeckhoven C., De Rijk P., 2005. *Genome Res.*, (15), 436-442.

Figure 1. Distribution des SNPs (flèches) du jeu de 96 sur la région B du MHC

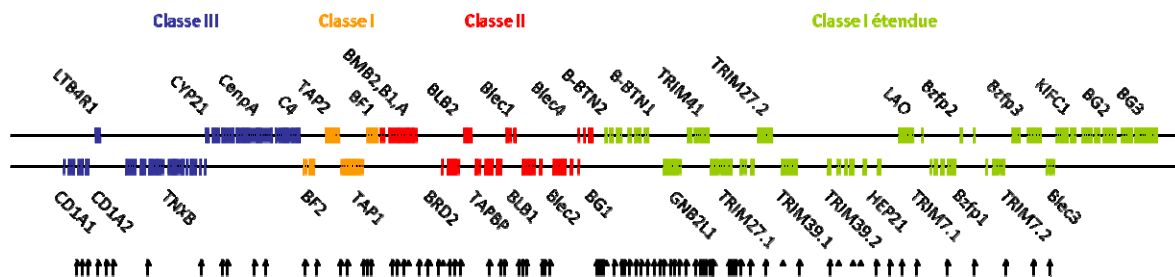


Figure 2. (A) Distribution des MAF (Minor Allelic Frequencies)
(B) Distribution des distances entre 2 SNPs consécutifs

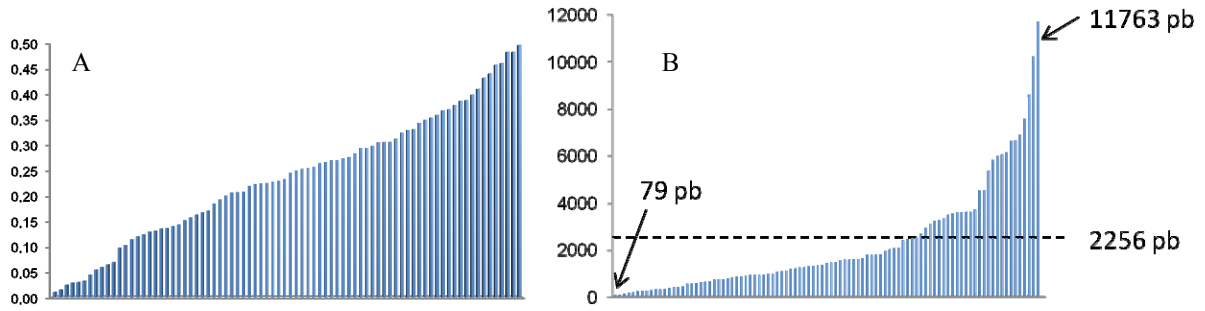


Figure 3. Représentation des corrélations entre SNPs, de blanc (faible) à noir (forte)

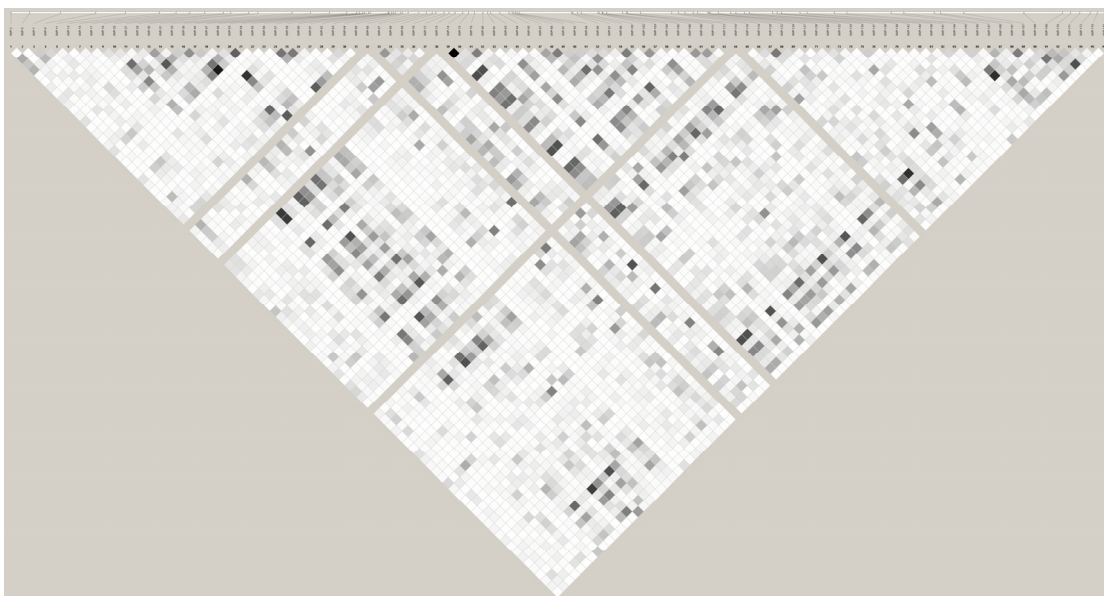


Figure 4. Représentation des génotypes SNPs pour des animaux de sérotypes B12, B19 et B13
(rouge = G, vert = A, orange = C, bleu = T)

