

INTÉRÊT DE L'UTILISATION DE L'IMPUTATION POUR L'ÉVALUATION GÉNOMIQUE EN POULE PONDEUSE

Herry Florian^{1,2}, Picard Druet David², Hérault Frédéric², Varenne Amandine¹, Burlot Thierry¹, Le Roy Pascale² et Allais Sophie²

¹NOVOGEN, 5 rue des Compagnons, Secteur du Vau Ballier, 22960 Plédran, ²PEGASE, INRA, Agrocampus Ouest, 16 Le Clos 35590 Saint-Gilles, France

florian.herry@inra.fr

RESUMÉ

La disponibilité en 2013 d'une puce à SNP commerciale haute densité (600 000 marqueurs) pour l'espèce poule a permis la mise en place de la sélection génomique dans les filières ponte et chair. Toutefois, le coût de cette puce reste élevé pour génotyper en routine l'ensemble des candidats à la sélection. L'un des enjeux est le développement de puces SNP basse densité (BD) de moindre coût. L'objectif est d'imputer l'ensemble des génotypes manquants pour revenir à une information 600K SNP avant de réaliser l'évaluation génomique. Toutefois, l'impact de l'utilisation directe des génotypages BD sur la précision des évaluations génomiques a été peu décrit.

L'objectif est donc de juger de l'intérêt de l'imputation des génotypages BD pour l'évaluation génomique dans une lignée de poule pondeuse. Les puces BD ont été construites selon une méthodologie équidistante. Les évaluations génomiques ont été réalisées sur le poids d'œuf et sur la force de fracture, avec une méthodologie ssGBLUP. L'impact des erreurs d'imputation, ou de l'absence d'imputation, sur le classement des candidats à la sélection lors d'une évaluation génomique sur ascendance a été évalué. Pour cela, les valeurs génomiques obtenues (GEBV), avec les génotypages imputés ou non depuis la puce BD, ont été comparées aux GEBV obtenues avec la puce haute densité (HD). La précision des GEBV a également été étudiée en prenant en référence les GEBV estimées sur descendance.

Le classement des candidats à la sélection avec les génotypages BD imputés est plus corrélé avec celui basé sur les génotypages HD que sans imputation. Toutefois, les corrélations sont toutes supérieures à 0.95 pour une densité supérieure à 5K SNP. L'impact de l'imputation sur le classement des individus est donc faible. Par ailleurs, la précision des GEBV estimées sur ascendance est équivalente entre les différentes stratégies (avec ou sans imputation) dans le cas du poids d'œuf pour une densité supérieure à 2K SNP, mais tend à se dégrader pour de très basses densités de SNP dans le cas de la force de fracture. L'utilisation de l'imputation est possible sans diminution significative de la précision des évaluations génomiques par ssGBLUP. L'absence d'imputation est également possible mais la dégradation observée sur la précision des GEBV semble dépendre des caractères.

ABSTRACT

Interest of using imputation for genomic evaluation in layer chicken

With the availability of the 600K Affymetrix® Axiom® high-density (HD) single nucleotide polymorphism (SNP) chip, genomic selection has been implemented in broiler and layer chicken. However, the cost of such HD SNP chip is still too high to genotype all selection candidates. A solution is to develop low density SNP chip, at a lower price, and to impute all missing markers before doing a genomic evaluation. However, the impact of the direct use of low density SNP chip on genomic evaluation accuracy is not well documented.

In this perspective, the interest of using or not imputation in genomic selection was studied in a pure layer line. An equidistant methodology was used to design low density SNP chips. Egg weight and egg shell strength were evaluated with ssGBLUP methodology. The impact of imputation errors or the absence of imputation on the ranking of the selection candidates was assessed with a genomic evaluation based on ancestry. To do so, genomic estimated breeding values (GEBV), with imputed or not imputed genotyping from the low density SNP chip, were compared to the GEBV obtained with the HD SNP chip. The accuracy of GEBV was also investigated by considering as reference the GEBV estimated on offspring.

With imputed HD genotyping, the ranking was slightly more correlated to that with HD genotyping than that with low density genotyping. However, Spearman correlations were above 0.95 with more than 5K SNPs. Thus the reordering of the breeders was limited. Finally, the GEBV accuracy estimated on ancestry was equivalent with or without imputation for a SNP density superior to 2K. But this accuracy tended to be decreased for some traits with less than 2K SNPs.

INTRODUCTION

Le développement d'une puce à SNP (Single Nucleotide Polymorphism) commerciale haute densité (HD) de 600 000 marqueurs en 2013 (Kranis et al., 2013) a permis le développement de la sélection génomique dans les filières poules de chair et ponte. Toutefois, les coûts d'une telle puce restent un problème. Des puces à SNP basse-densité (BD) à moindre coût peuvent être développées en sélectionnant un sous-ensemble de marqueurs de la puce HD puis en imputant les génotypes aux marqueurs manquants. Il existe trois principales méthodes pour développer une puce BD. La première consiste à sélectionner des SNP de façon équidistante en tenant compte ou non de la MAF (Fréquence Allélique Mineure) des SNP (Weigel et al., 2009 ; Wang et al., 2013 ; Herry et al., 2018). La deuxième consiste à sélectionner les SNP ayant des effets forts sur les caractères d'intérêt (Weigel et al., 2009 ; Zhang et al., 2011). Enfin la troisième consiste à sélectionner les SNP en tenant compte du déséquilibre de liaison (DL) entre SNP (Herry et al., 2018). Cette méthode a été étudiée compte tenu des particularités du génome aviaire et de la structure du DL (Robert et al., 2015). Les facteurs influençant la qualité de l'imputation ainsi que les relations entre qualité de l'imputation et évaluation génomique des candidats à la sélection sont bien documentés. À cause des erreurs d'imputation, il est attendu une diminution de la précision des évaluations génomiques des candidats à la sélection avec des génotypes imputés comparativement à celle obtenue avec des génotypes HD. La littérature confirme cela pour les puces BD allant jusqu'à 3K SNP avec une diminution, parfois limitée, de la précision des imputations (Weigel et al., 2009; Raoul et al., 2017). Enfin, à notre connaissance, peu d'études ont été menées en volailles (Wang et al., 2013). Peu d'études se sont penchées sur les conséquences d'une utilisation directe des génotypes BD sans imputation sur la précision des évaluations des candidats à la sélection.

L'objectif de cette étude est de juger de l'intérêt de l'imputation des génotypes BD pour l'évaluation génomique des candidats à la sélection.

1. MATÉRIELS ET MÉTHODES

1.1 Animaux

La population d'étude est constituée de quatre générations de coqs et de poules pondeuses Rhode Island issues d'une lignée pure commerciale créée et sélectionnée par Novogen (Plédran, France) dont les effectifs sont détaillés dans le Tableau 1.

1.2 Génotypes

Des prises de sang ont été réalisées au niveau de la veine brachiale des animaux et l'ADN a été extrait et hybridé sur la puce de génotype 600K Affymetrix® Axiom® HD. L'ensemble des animaux a été génotypé

pour 580 961 SNP. Les génotypes ont ensuite été filtrés par un contrôle qualité permettant de retenir 300 351 SNP informatifs pour les analyses, répartis sur les macro-chromosomes (1 à 5), les chromosomes intermédiaires (6 à 10), les micro-chromosomes (11 à 33) et le chromosome sexuel Z. Ces SNP sont désignés sous l'appellation 300K par la suite.

Parmi les individus génotypés de la génération G1, 67 individus mâles ont plus de 10 descendants avec des performances.

1.3 Phénotypes

Les mesures de performances sur le poids d'œuf (PO, en g) et sur la force de fracture (FF, en N) de la coquille ont été collectées à collectées entre 60 et 90 semaines et correspondent à des données individuelles. La force de fracture est mesurée grâce à une machine mesurant la force maximale (en N) appliquée à la coquille pour la fêler.

Tableau 1. Récapitulatif des données

	Rhode Island		
	Génotypes	Phénotypes	
Données	2362 individus	75 122 œufs	7983♀
Génération G0	447♂	29 318	2840♀
Génération G1	580♂	18 461	2028♀
Génération G2	132♂ ; 665♀	18 306	1885♀
Génération G3	55♂ ; 491♀	9037	1230♀

1.4 Simulation des puces basse-densité

À partir de la puce 300K, plusieurs puces BD ont été développées en sélectionnant un sous-ensemble de SNP de la puce HD. Une méthode « équidistante » (EQ), a été étudiée en sélectionnant des SNP à intervalles réguliers le long de chaque chromosome en maximisant la MAF des SNP retenus. 12 puces ont été créées avec des densités de SNP allant de 1K à 50K.

1.5 Qualité des imputations

Dans cette étude, les candidats à la sélection sont les 580 individus G1 avec des génotypes BD. Ceux-ci sont imputés à partir des génotypes HD des 447 mâles de la génération G0. Ces mâles G0 sont les pères ou les demi-frères des pères des candidats à la sélection (G1).

Pour chaque puce BD, la qualité de l'imputation des candidats à la sélection est mesurée par la corrélation de Pearson moyenne par SNP entre les vrais génotypes HD et les génotypes HD imputés.

1.6 Stratégies d'évaluations génomiques

Le poids d'œuf (PO) et la force de fracture de la coquille (FF) ont été évalués avec la méthodologie du Single Step GBLUP (Legarra et al., 2009). Les deux

caractères sont évalués simultanément selon un modèle animal multi-caractère.

Le premier objectif est d'étudier l'impact des erreurs d'imputation sur les évaluations génomiques sur ascendance. Une évaluation génomique sur ascendance « Asc_HD » est réalisée avec les génotypes HD des mâles G0 et des candidats, ainsi que les phénotypes des femelles G0. Une deuxième évaluation « Asc_Imp » est réalisée en remplaçant pour les candidats les génotypes HD par les génotypes BD imputés. Pour chaque puce BD, des corrélations de Spearman (corrélations de rang) sont calculées entre les vraies GEBV HD et les GEBV BD imputées pour les 67 mâles reproducteurs G1 avec des descendantes mesurées ainsi que pour les 150 meilleurs individus G1 selon leur GEBV HD, en fonction du caractère d'étude. Le deuxième objectif vise à étudier l'impact des imputations sur la précision des évaluations génomiques. Pour cela, une évaluation génomique sur descendance « Full_HD » des individus G1 avec descendance (donc les 67 reproducteurs G1) est réalisée en utilisant l'ensemble des informations disponibles (génotypes et performances des ascendantes, collatérales et descendantes). Pour chaque puce BD, des corrélations de Pearson sont calculées entre les GEBV « Full_HD » et les GEBV BD imputées estimées sur ascendance pour les 67 mâles reproducteurs G1.

Enfin, il est également intéressant d'étudier les possibles conséquences d'une utilisation directe des génotypes BD sans imputation sur les évaluations génomiques. Pour cela, les deux objectifs précédents sont repris en utilisant directement les génotypes BD de l'ensemble des animaux.

1.7 Logiciels utilisés

Les génotypes des candidats à la sélection sont imputés avec le logiciel FImpute (Sargolzaei et al., 2014).

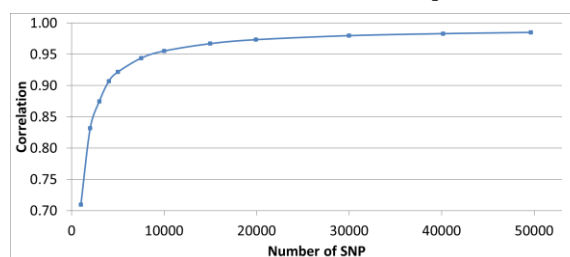
Le scénario « Full_HD » avec l'ensemble des informations disponibles est utilisé pour estimer les paramètres génétiques du modèle. Le programme remlf90 (Misztal et al., 2002) est utilisé pour estimer les composantes de la variance, génétique et résiduelle. Les différentes évaluations sont ensuite réalisées avec blupf90, ces composantes étant fixées.

2. RÉSULTATS ET DISCUSSION

2.1 Qualité des imputations

Les corrélations moyennes entre les vrais génotypes HD et les génotypes BD imputés augmentent avec le nombre de SNP sur les puces BD (Figure 1). Pour des densités de 3K et 20K SNP, les corrélations sont respectivement de 0.8747 et 0.9374. Un point d'inflexion est noté entre 5K et 10K SNP. Enfin, pour une densité de 1K la corrélation moyenne diminue à 0.7098 témoignant d'une qualité d'imputation dégradée.

Figure 1. Évolution des corrélations de Pearson entre génotypes HD et génotypes BD imputés en fonction du nombre de SNP sur les puces BD.



2.2 Impact des erreurs d'imputations

Les résultats sont similaires pour les deux caractères et ne sont développés que pour le PO.

2.1.1 Sur les 150 meilleurs individus G1

La figure 2 montre pour le PO une augmentation des corrélations de Spearman entre les GEBV « Asc_HD » et les GEBV « Asc_Imp » avec une augmentation du nombre de SNP sur les puces BD. En effet, pour les puces 3K et 20K, les corrélations sont respectivement de 0.9045 et 0.9885. Ceci est cohérent avec la littérature. Wang et al. (2013) ont montré en poulet de chair qu'utiliser des génotypes 20K imputés à partir de 766 SNP permettait d'augmenter par deux les GEBV pour le poids des individus ou l'intensité de ponte par rapport aux GEBV obtenus avec des génotypes 20K imputés à partir de 377 SNP.

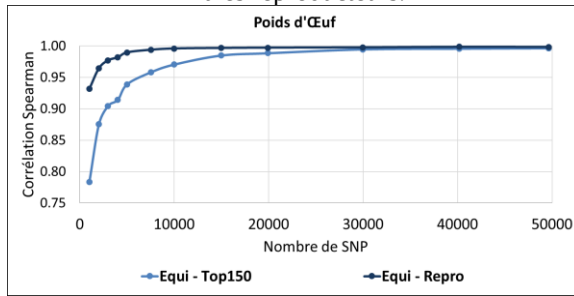
Enfin, avec plus de 5K SNP, les corrélations moyennes sont supérieures à 0.95, ce qui indique un reclassement plutôt faible des meilleurs individus.

2.1.2 Sur les individus reproducteurs G1

Les corrélations de Spearman ont été calculées pour les 67 individus reproducteurs G1. La figure 2 montre pour le PO une augmentation des corrélations de Spearman avec le nombre de SNP sur les puces BD. En effet, pour les puces 3K et 20K, les corrélations sont respectivement de 0.9771 et 0.9972. Les résultats sont donc supérieurs à ceux obtenus pour les 150 meilleurs individus G1. Ceci s'explique par le fait que les 67 individus reproducteurs ne sont pas les meilleurs individus pour le PO, mais les meilleurs individus pour un index multi-caractère. Ces 67 individus sont donc répartis dans l'ensemble du classement des 580 individus G1 pour le PO.

Enfin, pour une densité de SNP supérieure à 2K, des corrélations supérieures à 0.95 sont obtenues, indiquant un reclassement plutôt faible des 67 individus reproducteurs. En revanche, pour la puce 1K, la corrélation est de 0.9316, signe qu'une diminution importante du nombre de SNP peut avoir des conséquences plus fortes sur les évaluations et le reclassement des individus.

Figure 2. Évolution des corrélations de Spearman entre les GEBV « Asc_HD » et les GEBV « Asc_Imp » pour les 150 meilleurs mâles ou les 67 mâles reproducteurs.



2.3 Impact sur la précision des évaluations génomiques

Les corrélations de Pearson sont calculées pour le PO et FF entre les GEBV « Full_HD » et les GEBV BD imputées estimées sur ascendance pour les 67 individus reproducteurs G1. Les résultats sont similaires pour les deux caractères et ne sont développés que pour le PO. Les corrélations varient jusqu'à une densité de 20K SNP puis deviennent stables pour des densités supérieures. Les corrélations pour les puces 1K et 20K sont respectivement de 0.4472 et 0.4854. Toutefois, la corrélation entre les GEBV « Full_HD » et les GEBV « Asc_HD » est de 0.4848. Cette corrélation représente la valeur théorique maximale atteignable en réalisant une évaluation sur ascendance avec la puce HD (au lieu des puces BD). Les corrélations associées aux différentes puces ne sont significativement pas différentes de celle obtenue avec la puce HD. En effet l'erreur standard pour chaque puce est de ± 0.11 . Ceci est en accord avec les résultats précédents montrant un impact faible des erreurs d'imputation sur le reclassement des reproducteurs. En effet, pour une densité supérieure à 3K SNP, les corrélations entre GEBV HD estimées sur ascendance et GEBV BD imputées estimées sur ascendance sont supérieures à 0.99. Ces résultats sont cohérents avec la littérature. VanRaden et al. (2012) ont montré qu'en moyenne pour 28 caractères en bovin, la fiabilité des résultats diminuait de 61.1% avec 300K SNP à 60.7% avec 45K SNP (et imputés à 300K). En revanche, Chen et al. (2014) ont montré chez les bovins que 50K SNP imputés à partir de 384 SNP entraînait une diminution de la précision des évaluations de 0.61 à 0.49 pour le rendement laitier, et de 0.62 à 0.53 pour le taux de cellules.

Enfin, il est important de noter que la différence entre les densités des puces BD et la densité de la puce HD (300K) est beaucoup plus élevée que dans d'autres études.

2.4 Impact de l'absence d'imputation

Les résultats sont similaires pour les deux caractères et ne sont développés que pour le PO.

2.4.1 Sur les 150 meilleurs individus G1

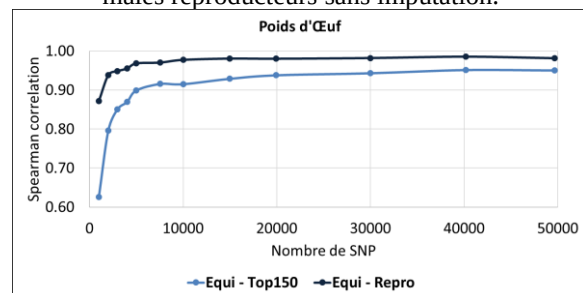
La figure 3 montre une augmentation des corrélations de Spearman entre les GEBV « Asc_HD » et les GEBV « Asc_BD » avec le nombre de SNP sur les puces BD. En effet, pour les puces 3K et 20K, les corrélations sont respectivement de 0.8507 et 0.9379. Comparativement aux résultats obtenus avec imputation (0.9045 et 0.9885 pour les puces 3K et 20K respectivement), il est noté une diminution des résultats des évaluations sans imputation avec une baisse des corrélations de 0.0538 et 0.0506 pour les puces 3K et 20K respectivement. De plus, les résultats sont significativement plus faibles pour les puces à très basse densité. En effet, pour la puce 1K, la corrélation moyenne est de 0.6261 sans imputation quand elle est de 0.7833 avec imputation. L'impact sur le reclassement des individus est donc conséquent pour de très faibles densités de SNP.

Enfin, à partir de 20K SNP, les corrélations semblent atteindre un seuil pour une corrélation moyenne de 0.95 alors qu'avec les imputations elles étaient supérieures à 0.99.

2.4.2 Sur les individus reproducteurs

Les corrélations de Spearman entre les GEBV « Asc_HD » et les GEBV « Asc_BD » ont été calculées pour les 67 individus reproducteurs G1. La figure 3 montre pour le PO une augmentation des corrélations de Spearman avec le nombre de SNP sur les puces BD. En effet, pour les puces 3K et 20K, les corrélations sont respectivement de 0.9484 et 0.9802. Les corrélations sont donc supérieures à celles obtenues pour les 150 meilleurs individus G1 du fait d'une meilleure distribution des 67 individus reproducteurs dans le classement des 580 individus G1. En revanche, en comparant avec les résultats obtenus avec imputation, les résultats sont plus faibles mais ne sont significativement pas différents. Toutefois, les corrélations pour la puce 1K sont de 0.8718, indiquant un reclassement des individus plus élevé que précédemment (0.9316) et pouvant impacter la précision des évaluations génomiques.

Figure 3. Évolution des corrélations de Spearman entre les GEBV « Asc_HD » et les GEBV « Asc_BD » pour les 150 meilleurs mâles ou les 67 mâles reproducteurs sans imputation.



2.5 Impact de l'utilisation directe des génotypages BD sur la précision des évaluations génomiques

Les corrélations de Pearson sont calculées pour le PO et FF entre les GEBV « Full_HD » et les GEBV HD

non imputés estimées sur ascendance pour les 67 individus reproducteurs G1.

Pour des densités de SNP supérieures à 2K, les résultats sont plutôt stables. En effet, pour les puces 3K et 20K, les corrélations sont respectivement de 0.4471 et 0.4675 pour le PO. Toutefois les erreurs standards associées à ces corrélations sont de ± 0.11 , ce qui indique qu'il n'y a pas de différence significative entre les différents résultats. Ceci est cohérent avec les résultats précédents montrant un impact léger de l'absence d'imputation sur le reclassement des reproducteurs. De plus, la corrélation entre les GEBV « Full_HD » et les GEBV « Asc_HD » est de 0.4848. Pour une densité de SNP supérieure à 2K, une utilisation directe des puces BD sans imputations pourrait donc permettre d'obtenir des résultats aussi bons que ceux obtenus avec imputation ou qu'avec la puce HD elle-même.

En revanche, la corrélation moyenne pour la puce 1K est de 0.4018 (± 0.11). De même, pour la FF, la corrélation moyenne pour la puce 1K est de 0.2980 (± 0.12) alors que la corrélation entre les GEBV « Full_HD » et les GEBV « Asc_HD » est de 0.4027. Bien que non significatives, ces baisses de corrélations montrent que l'utilisation directe de puces à très faible densité sans imputation pourrait avoir des impacts sur la précision des évaluations.

3. CONCLUSION

L'objectif des sélectionneurs est d'obtenir de bons résultats d'évaluations génomiques des candidats à la sélection. En mettant en lien précision des imputations et conséquences sur les évaluations génomiques, il a été montré que la densité de SNP n'avait finalement qu'un impact très faible sur le reclassement des meilleurs individus et des reproducteurs, ainsi que sur la précision des évaluations génomiques. Sans imputation, l'utilisation directe des génotypes BD des candidats à la sélection a montré que des résultats moins corrélés étaient obtenus comparés à ceux obtenus avec imputation. Pour des densités de SNP supérieures à 2K, les corrélations entre GEBV HD et GEBV BD restent toutefois supérieures à 0.90. En revanche, le reclassement des meilleurs individus pour un caractère est plus impacté, notamment pour de faibles densités de SNP. Ainsi, peu de marqueurs peuvent suffire pour faire le tri entre bonnes et mauvaises familles mais peuvent ne pas suffire pour trier intra-famille le meilleur candidat.

Enfin, il convient de noter qu'en fonction des caractères étudiés, une précision plus faible des évaluations génomiques est obtenue pour de très faibles densités de génotypes (1K).

BIBLIOGRAPHIE

- Chen L., Li C., Sargolzaei M. et al., 2014. PLoS One. 9:e101544.
 Hérault F., Yon J., Herry F. et al., 2016. <https://prodinra.inra.fr/record/375447>
 Herry F., Hérault F., Picard Druet D. et al., 2018. BMC Genetics. 19:108-121.
 Kranis A., Gheyas A.A., Boschiero C. et al., 2013. BMC Genomics. 14:59-71.
 Legarra A., Aguilar I. et Misztal I., 2009. J. Dairy Sci. 92:4656-4663.
 Misztal I., Tsuruta S., Strabel T. et al., 2002. Proc. 7th. World Congr. Genet. Appl. Livest. Prod.
 Raoul J., Swan A.A., et Elsen J.M., 2017. Genet. Sel. Evol. 49:76-87.
 Robert R., Hérault F., Romé H. et al., 2015. Proc. 9th. Eur. Symp. Poult. Genet.
 Sargolzaei M., Chesnais J.P. et Schenkel F.S., 2014. BMC Genomics. 15:478-489.
 VanRaden P.M., Null J.D., Sargolzaei M. et al., 2012. J. Dairy Sci. 95:1-11.
 Wang C., Habier D., Peiris B.L. et al., 2013. Poult. Sci. 92:1712-1723.
 Weigel K.A., de los Campos G., González-Recio O. et al., 2009. J. Dairy Sci. 92:5248-5257.
 Zhang Z., Ding X., Liu J. et al., 2011. J. Dairy Sci. 94:3642-3650.