

ESTIMATION DES PARAMETRES GENETIQUES DANS UN SCHEMA OU LES PEDIGREES SONT CONNUS A POSTERIORI : ETUDE PAR SIMULATION

Chapuis Hervé¹, Typhanie Ruer², Magali Blanchet²

¹ GenPhySE, Université de Toulouse, INRA, ENVT, 31326 CASTANET TOLOSAN

² Grimaud Frères Sélection, La Corbière, Roussay 49450 SEVREMOINE

herve.chapuis@inra.fr

RESUME

La qualité des paramètres génétiques estimés dépend du nombre de familles mesurées et de leur taille dans la population étudiée. Classiquement, dans les populations avicoles, ces éléments sont pris en compte lors de l'éclosion afin de disposer d'un échantillon aussi représentatif que possible. Si les pedigrees ne sont pas parfaitement connus lors de l'éclosion mais a posteriori avec des marqueurs moléculaires (cas d'une reproduction au sol, sans cages individuelles, en monte naturelle avec des parquets de x mâles et y femelles), il faut s'assurer du dimensionnement du dispositif pour obtenir des paramètres exploitables. C'est particulièrement vrai dans un contexte expérimental où la taille des populations est limitée à la fois par le budget et par le respect du principe des 3 R (réduction). La question se pose dans le cas d'espèces ne se reproduisant pas en cages (oies) mais plus généralement en prévision d'un éventuel abandon des cages individuelles en sélection. Nous avons simulé une population expérimentale d'oies fonctionnant selon ce principe en faisant varier la taille des lots de $n=200$ à $n=1000$ pour étudier les propriétés statistiques (biais et dispersion) des paramètres génétiques estimés, qu'il s'agisse de caractères mesurés sur les candidats à la sélection ou sur des collatéraux abattus. Nous avons aussi envisagé les conséquences d'erreurs dans la généalogie. En appliquant quelques règles non contraignantes pour équilibrer les origines familiales, nous obtenons des propriétés très similaires pour les paramètres génétiques obtenus avec un pedigree connu à l'éclosion ou a posteriori. Un taux d'erreur de 5% dans l'assignation des parentés n'entraîne pas une nette dégradation de la propriété des paramètres génétiques. La taille des lots mis en place pour obtenir des paramètres exploitables devra cependant tenir compte du taux de succès de l'assignation à un couple parental unique. D'autres études seront nécessaires pour quantifier le gain génétique espéré dans de telles conditions.

ABSTRACT

Estimation of genetic parameters in a poultry design when pedigrees are not known at hatch. A simulation study.

Statistic properties of genetic parameters depend on the number and size of families sampled in the studied population. In a monitored poultry population, these factors are usually accounted for at hatch. In the case of a *a posteriori* DNA-pedigreed population, the design must be carefully setup, lest inaccurate genetic parameters would be estimated. Here we simulated a goose population featuring batches used for breeding or for slaughtering. Batch size varied from $n=200$ to $n=1000$ measured animals, and we also investigated the consequences of pedigree errors on the prediction error of genetic parameters. Using simple rules to limit disequilibrium in familial representation, we obtained very similar statistical properties between classical and *a posteriori* DNA-pedigreed designs. A 5% rate of false pedigree did not lead to an unbearable deterioration of the parameter accuracy. The computations were based on the number of animals uniquely assigned to a parental pair. The expected assignment success should therefore be accounted for at setup. Further study should be carried out to predict the expected genetic gain in case of a *a posteriori* DNA-pedigreed populations

INTRODUCTION

La connaissance des paramètres génétiques est indispensable pour la mise en place d'un schéma de sélection performant. Quelle est la part de la variance observée qui est imputable à la génétique et quelles sont les relations entre les caractères d'intérêt ? La réponse à ces questions dépend de l'héritabilité et des corrélations génétiques, qui doivent être estimées, à partir des performances collectées et des généalogies enregistrées depuis plusieurs générations. La qualité de l'estimation des paramètres génétiques dépend classiquement du nombre et de la taille des familles mesurées. Ces facteurs font généralement l'objet d'une planification et sont contrôlés dans le cadre d'un plan d'accouplement hiérarchique (1 ♂ croisé avec k ♀), qui suppose des cages individuelles et souvent le recours à l'insémination artificielle. Nous nous sommes intéressés au cas où les généalogies ne sont pas connues à l'éclosion (cas de la reproduction au sol, où les femelles peuvent pondre hors des nids) afin d'évaluer l'impact d'un tel schéma sur la précision et le biais des paramètres estimés. Dans ce cas les pedigrees sont obtenus a posteriori via des marqueurs moléculaires, et rien ne garantit une représentation équilibrée des différentes familles possibles. En utilisant une approche similaire à celle de Chapuis et David (2018), nous avons simulé un schéma de sélection d'oies comprenant des lots de renouvellement et des lots destinés au gavage afin d'évaluer la qualité des paramètres génétiques estimés.

1 MATERIELS ET METHODES

1.1 Population simulée

La population comporte 4 cheptels (Cf figure 1) mesurés (G_1 à G_4) et 2 générations d'ancêtres (G_0 et G_{-1}). Les cheptels G_1 et G_3 sont considérés comme des cheptels de renouvellement. G_2 et G_4 sont abattus. 4 caractères sont simulés de façon à estimer 2 héritabilités et une corrélation génétique dans chaque type de cheptel ainsi que la corrélation entre un caractère mesuré dans un lot abattu et dans un lot de renouvellement. Les caractères seront appelés car1 et car2 dans le lot de renouvellement, car3 et car4 dans le lot exclu du renouvellement. Nous avons testé deux valeurs différentes pour l'héritabilité : $h^2=0,3$ pour les caractères 1 et 3, $h^2=0,5$ pour les caractères 2 et 4. Les corrélations génétiques simulées sont ici de +0,5.

1.2 Principe de l'étude

Une population fictive organisée en 6 cheptels (figure 1) est simulée selon les règles classiques de la génétique quantitative. Les performances sont analysées pour estimer les paramètres génétiques, ce qui permet de comparer valeurs vraies et valeurs estimées. Les opérations étant répétées un grand nombre de fois ($n=400$), on peut calculer le biais et la précision des paramètres génétiques estimés, en fonction des facteurs de variation envisagés.

La qualité de l'estimation sera appréciée par la racine carrée de l'erreur quadratique moyenne ($\sqrt{MSE}$), qui combine biais et précision. Considérant que la vraie valeur d'un paramètre a une probabilité de 90% de se trouver à moins de deux écarts types de la moyenne, on considèrera qu'une valeur acceptable pour $\sqrt{MSE}$ doit être inférieure à la moitié de l'héritabilité, afin de la déclarer « significativement différente de 0 ». Pour une corrélation génétique non nulle, on cherche à s'assurer de son signe et là encore on considèrera satisfaisantes les valeurs de $\sqrt{MSE}$ inférieures, en valeur absolue, à la moitié de la corrélation simulée.

1.3 Facteurs de variation

1.3.1 Plans d'accouplement

Nous avons considéré 10 parquets de 5 ♂ accouplés à 20 ♀, soit un effectif de reproducteurs de 50 pères et 200 mères dans les cheptels G_{-1} , G_0 et G_1 . Les résultats obtenus sont comparés à plan de croisement hiérarchique de 50 mâles, chacun accouplé à 4 femelles. L'effectif de reproducteurs (qui est donc le nombre minimal d'animaux mesurés) est de 250 dans les deux schémas.

1.3.2 Taille des lots considérés

L'estimation des paramètres génétiques ne pose problème que quand on est en situation d'effectifs contraints. Nous avons considéré un schéma expérimental où les lots étaient de $n=600$ individus dans les cheptels de renouvellement et les lots abattus. Afin de tester la robustesse des paramètres en fonction des effectifs dans ces deux cheptels, nous avons fait varier la taille d'un type de lot (renouvellement ou gavage) de $n=200$ à $n=1000$, en fixant la taille des lots de l'autre type à $n=600$. Pour les lots de renouvellement, la taille minimale ne pouvant être inférieure à 250, les valeurs testées allaient de 300 à 1000.

1.3.3 Erreurs d'assignation

Afin de tenir compte d'une éventuelle erreur d'assignation consécutive à des erreurs de

génotypages, les paramètres génétiques ont aussi été estimés avec un pedigree erroné dans lequel, pour chacun des cheptels, 5% des animaux ont eu leurs parents aléatoirement échangés.

2 RESULTATS & DISCUSSION

2.1 Taille des lots

Les figures 2 et 3 représentent la qualité des estimations (exprimée au travers de $\sqrt{MSE}$) en fonction de la taille des lots de renouvellement (lots G1 et G3, figure 2) ou de celle des lots abattus (lots G2 et G4, figure 3). Comme attendu, la qualité des paramètres s'améliore avec la taille des lots ($\sqrt{MSE}$ diminue quand le nombre d'individus mesurés croît). Dans tous les cas on obtient des valeurs inférieures au seuil de tolérance fixé (0,15 pour une héritabilité supposée de 0,3 et 0,25 pour un paramètre ayant une valeur de 0,25). Comme on pouvait s'y attendre, dans la mesure où les animaux mesurés dans les lots abattus n'ont pas de descendants, la valeur de $\sqrt{MSE}$ est légèrement plus élevée sur la figure 3 signalant une estimation moins précise que pour les caractères mesurés dans les lots de renouvellement.

La valeur de $\sqrt{MSE}$ est meilleure (plus faible) pour les héritabilités, puis pour la corrélation génétique entre deux caractères mesurés sur les mêmes animaux. La valeur la plus élevée (qui reste inférieure à la borne de 0,25) est observée pour la corrélation génétique entre deux caractères mesurés chacun sur des animaux de lots distincts.

Il faut également noter que les effectifs considérés ici sont les animaux « efficaces », c'est-à-dire avec des mesures et un pedigree connu. Avec un succès de réassignation à un couple parental unique autour de 0,80 (donnée non publiée), il faudrait mettre en place des lots de 375 animaux pour obtenir les résultats lus pour 300 animaux.

2.2 Effet du plan d'accouplement

On ne constate aucune détérioration de la qualité des estimations qui serait due au passage d'un schéma classique (accouplements hiérarchiques en cages individuelles) à un schéma de reproduction au sol. Notons toutefois que nous avons ici considéré un nombre égal d'individus issus de chacun des 10 parquets de reproducteurs. C'est une règle simple à mettre en œuvre afin d'équilibrer le dispositif. Néanmoins, nos simulations envisagent, intra bloc, une grande diversité de situation, avec parfois des contributions très inégales des reproducteurs, ce qui nous permet de considérer qu'elles reflètent bien la situation envisagée en pratique. Dans un souci de préserver la diversité génétique, et en situation

d'effectifs limités, la panmixie totale (avec monte naturelle, ce qui suppose la prise en compte des comportements sexuels) n'est cependant pas souhaitée dans nos populations expérimentales.

2.3 Effet des erreurs d'assignation

L'introduction d'un taux de 5% de fausses généalogies dans le pedigree dégrade systématiquement la qualité des estimations. Cette dégradation reste néanmoins modérée et l'erreur sur les paramètres demeure acceptable. Nous avons considéré un taux d'erreur de 5% dans les pedigrees réassignés, à ne pas confondre avec le taux d'erreur des génotypages. Les algorithmes de réassignation utilisés (Barbotte et al, 2012 ; Allal et al, 2015) tiennent compte des éventuelles erreurs de génotypage et par conséquent, ce taux nous a semblé être une borne haute. Notons également que, hors des cas où la sélection génomique est mise en place, le taux d'erreur dans les pedigrees établis dans les schémas classiques n'est pas connu. Il est supposé nul, alors que des défaillances dans la traçabilité ne sont pas exclues.

CONCLUSION

Par simulation nous avons démontré qu'en appliquant des règles simples visant à limiter le déséquilibre des représentations familiales, l'utilisation de pedigrees connus a posteriori par génotypage (et donc non contrôlés lors de la mise en place du cheptel à l'éclosion) permettait d'estimer correctement des paramètres génétiques. Ces estimations sont, par ailleurs, robustes vis-à-vis d'un taux modéré d'erreurs de généalogies.

Comme toute modélisation, celle-ci a fait l'objet de simplifications et ne prétend donner qu'un ordre de grandeur de la précision attendue. Hors les situations extrêmes pour des effectifs mesurés très faibles, la marge de sécurité semble suffisante pour déceler une héritabilité non nulle, ce qui permet d'envisager des travaux exploratoires sur des nouveaux caractères, tout en limitant le nombre d'animaux utilisés, afin de respecter le principe de Réduction appliqué en expérimentation animale (Russel et Burch, 1958). La mise en œuvre dans des schémas de sélection commerciaux impliquant plusieurs caractères et donc l'estimation de corrélations génétiques requiert, quant à elle, des effectifs plus importants. Ces schémas sont concernés par un objectif de progrès génétique, dont la prédiction devra faire l'objet de travaux complémentaires en cas d'assignation de parenté. De même, si l'on envisageait une sélection en monte naturelle, des travaux complémentaires de modélisation seraient nécessaires, afin de prendre en compte le

déséquilibre de représentation familiale induit par les comportements sexuels et les compétitions entre reproducteurs.

REFERENCES BIBLIOGRAPHIQUES

- Allal, F., Doan, Q., Chatain, B., Vergnet, A., Vandeputte, M. (2015). Aquaculture 2015: Cutting edge science in Aquaculture, Montpellier, FRA (2015-08-23 - 2015-08-26).
- Barbotte, L., Boichard, D., Genestout, L., Haffray, P., Chapuis, H., Chantry-Darmon, C., Boscher, M. Y. (2012). Programme and Abstract Book 33rd Conference of the International Society for Animal Genetics (p. 100). Cairns, AUS (2012-07-15 - 2012-07-20).
- Chapuis, H., David, I. (2018). 11. World Congress on Genetics Applied to Livestock Production (WCGALP), Auckland, NZL. NZL : Massey University. 1130 p.
- Russell W., Burch R. (1959) The Principles of Humane Experimental Technique. Londres R.-U. : Methuen.

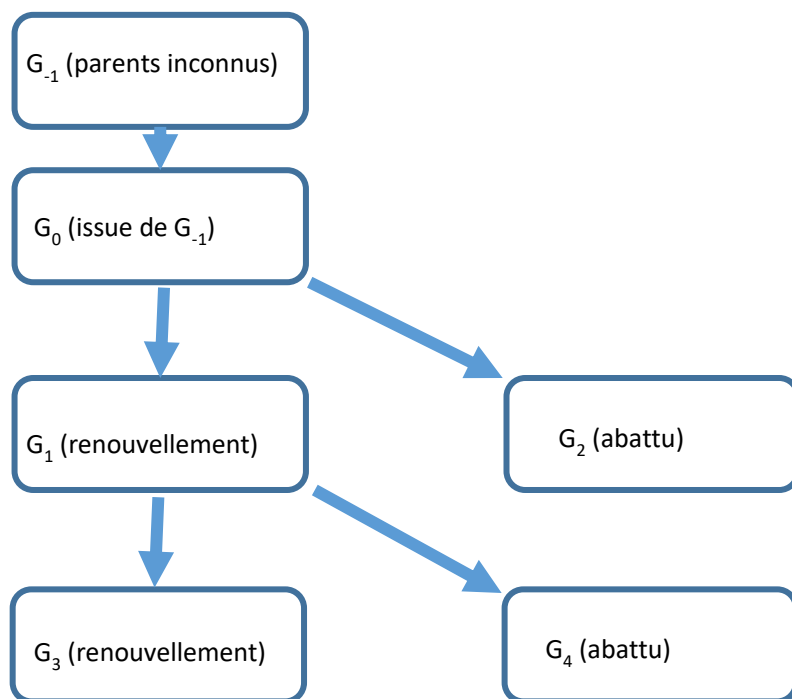


Figure 1: Organisation des différents cheptels simulés

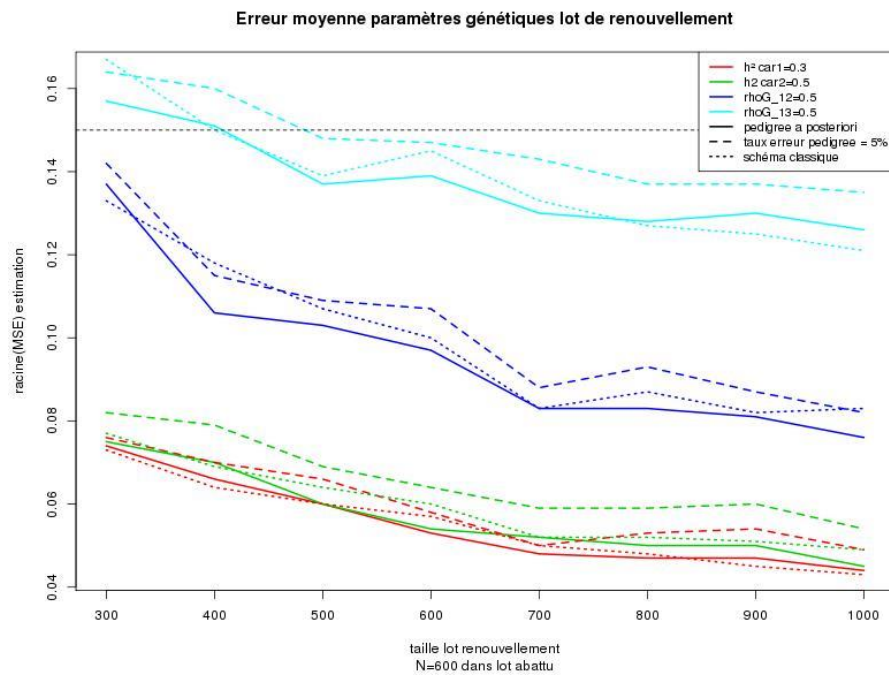


Figure 2: Erreur moyenne prédite pour les différents paramètres génétiques en fonction de la taille des lots de renouvellement (G_1 et G_3). Le schéma avec réassignation de parenté est comparé à un schéma classique (accouplements hiérarchiques et cages individuelles), et un taux d'erreur de 5% dans l'assignation est également envisagé.

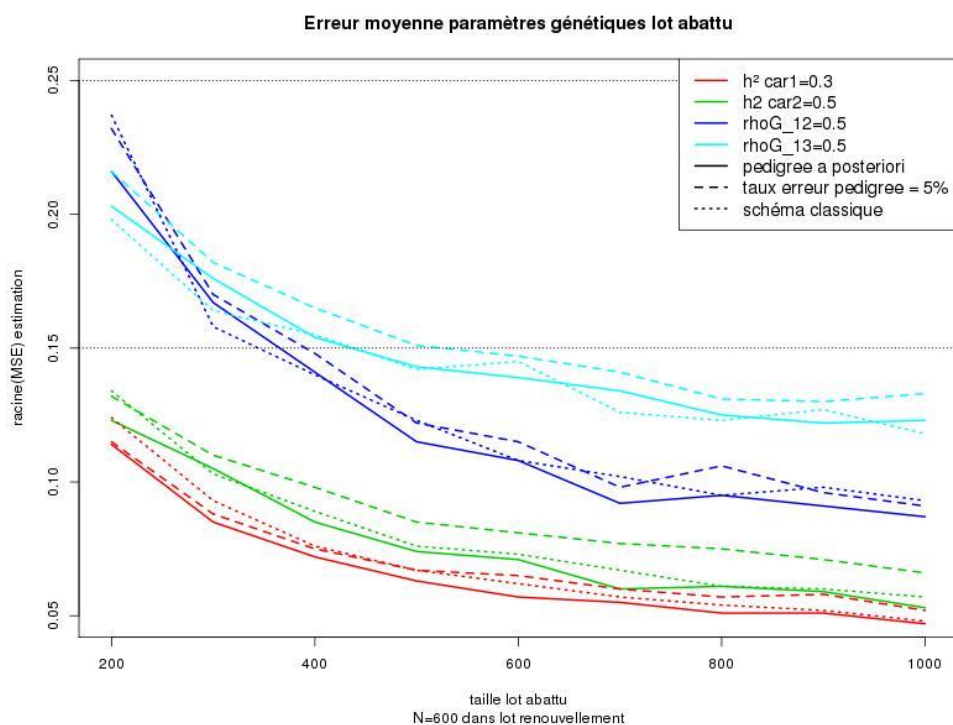


Figure 3: Erreur moyenne prédite pour les différents paramètres génétiques en fonction de la taille des lots abattus (G_2 et G_4). Le schéma avec réassignation de parenté est comparé à un schéma classique (accouplements hiérarchiques et cages individuelles), et un taux d'erreur de 5% dans l'assignation est également envisagé.