



HAL
open science

Intérêt de la sélection génomique dans les programmes de sélection porcins : cas d'une lignée mâle de grande taille

Thierry Tribout

► **To cite this version:**

Thierry Tribout. Intérêt de la sélection génomique dans les programmes de sélection porcins : cas d'une lignée mâle de grande taille. Sciences agricoles. AgroParisTech, 2013. Français. NNT : 2013AGPT0054 . tel-01135169

HAL Id: tel-01135169

<https://pastel.hal.science/tel-01135169>

Submitted on 24 Mar 2015

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.



Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech)

Spécialité : Génétique Animale

présentée et soutenue publiquement par

Thierry TRIBOUT

le 1^{er} octobre 2013

Intérêt de la sélection génomique dans les programmes de sélection porcins : cas d'une lignée mâle de grande taille

Directeur de thèse : **Florence PHOCAS**
Co-encadrement de la thèse : **Catherine LARZUL**

Jury

M. Etienne VERRIER, Professeur, AgroParisTech, France
M. Luc JANSS, Senior Scientist, Aarhus University, Denmark
M. Andrés LEGARRA, Chargé de Recherches, INRA SAGA, France
M. Alain CHARCOSSET, Directeur de Recherches, INRA UMR de Génétique Végétale, France
M. Michel SOURDIOUX, Responsable Programme Génétique, Gene+, France
Mme Florence PHOCAS, Chargée de Recherches, INRA GABI, France

Président
Rapporteur
Rapporteur
Examineur
Examineur
Examinatrice

Remerciements

Ce travail a été réalisé dans l'unité GABI de l'INRA de Jouy-en-Josas. Je tiens tout d'abord à remercier Didier Boichard et Denis Milan, qui se sont succédé à la direction du Département de Génétique Animale de l'INRA, pour m'avoir autorisé à réaliser cette thèse conjointement à mon activité d'Ingénieur de Recherche dans l'équipe Génétique du Porc de GABI, et pour les moyens mis à ma disposition pour mener à bien ce travail. Je remercie également l'ANR qui a apporté un financement partiel à cette thèse dans le cadre du projet UTOPIGE.

Je remercie les membres du jury : Etienne Verrier, qui en a assuré la présidence et s'est chargé des démarches administratives, Luc Janss et Andrés Legarra qui m'ont fait l'honneur d'être rapporteurs de ce travail, ainsi que Michel Sourdioux et Alain Charcosset qui m'ont fait le grand plaisir d'en être examinateurs.

Je remercie également tous les membres du comité de suivi de cette thèse :

- Pascal Croiseau et François Guillaume, dont les compétences méthodologiques et informatiques ont orienté mes choix techniques quant à la réalisation des simulations ;
- Jean-Pierre Bidanel, qui par sa grande connaissance des programmes de sélection porcins a contribué au choix des scénarios à simuler ;
- Stéphane Lemarié, responsable de l'Unité d'Economie Appliquée de l'INRA de Grenoble, que je remercie particulièrement pour ses explications et précieux conseils quant à la manière d'évaluer l'intérêt économique de la sélection génomique, et pour m'avoir reçu à Grenoble afin de discuter des résultats de mes simulations.

Je tiens bien évidemment à exprimer ma gratitude à mes directrices de thèse, Catherine Larzul et Florence Phocas, qui m'ont encadré au cours de ces quatre années et demie, m'ont permis de prendre du recul par rapport à mon travail et mes résultats, et ont relu de manière critique ce mémoire. Mention particulière à Florence, qui bien qu'évoluant habituellement dans le monde bovin, a fait l'effort de s'intéresser à l'univers porcin. En dehors de ses qualités scientifiques et de son encadrement, je lui suis également très reconnaissant pour sa grande disponibilité malgré son emploi du temps chargé, pour son soutien dans les moments de doute (ils ont été nombreux !) et pour la gentillesse et l'humanité dont elle a fait preuve au cours des dernières années.

Je remercie également mes collègues de l'IFIP (en particulier Christophe Bazin, Sandrine Schwob, Isabelle Mérour, Claire Hassenfratz, Joël Bidanel, Alban Bouquet et Isabelle

Delaunay) et des Entreprises de Sélection Porcine (Thierry Bellec, Bruno Ligonesche, Michel Sourdioux et Cédric Pincent), pour leur expertise et pour les informations qu'ils m'ont fournies sur l'organisation de la sélection et de la filière porcine. Je vais à présent travailler sur les bovins, mais qu'ils sachent que mes collaborations passées avec eux resteront de bons souvenirs.

Un grand merci à Jack Dekkers, qui a pris le temps de discuter avec moi de mon projet de thèse lors d'une de ses venues en France et dont les conseils ont été extrêmement précieux, ainsi qu'à Didier Boichard, qui en sacrifiant un week-end pour relire ce manuscrit, a su me remotiver dans la phase finale de la rédaction.

Mes remerciements vont également aux membres du « café cochon » du bâtiment 211, trop nombreux pour être cités individuellement, pour les moments quotidiens de détente et de convivialité passés auprès d'eux dans la bonne humeur.

Si un diplôme de la « meilleure voisine de bureau d'un thésard » existait, ma collègue Marie-Pierre le mériterait sans conteste. Je la remercie sincèrement pour toute l'aide qu'elle m'a apportée (utilisation d'EndNote, finalisation et mise en page du manuscrit, ...), ainsi que pour la patience exemplaire dont elle a fait preuve au cours de ces dernières années, y compris dans mes grands moments de stress. Je la soupçonne cependant d'être à présent encore plus soulagée que moi par l'achèvement de cette thèse ...

Enfin, je ne saurais terminer sans adresser mes remerciements (accompagnés d'excuses) à ma compagne Laurence que je tenais régulièrement (et parfois contre son gré) informée des avancées de ce travail et à qui j'ai infligé la relecture de ce manuscrit, ainsi qu'à mes deux filles Elora et Anaïs qui ont désormais par la force des choses, à 9 et 11 ans, une vision assez précise de l'organisation des programmes de sélection porcins ... Je les remercie pour leur soutien infailible, et pour m'avoir rappelé, lorsque c'était nécessaire, qu'il y a d'autres choses importantes dans la vie.

Résumé

L'objectif de ce travail de thèse était d'évaluer l'intérêt de mettre en place des évaluations génomiques dans les programmes de sélection porcins. Des simulations stochastiques ont été réalisées dans le cas d'un programme de sélection d'une lignée mâle de grande taille contenant 1 050 femelles reproductrices et 50 verrats, sélectionnée pendant 10 ans pour améliorer un objectif de sélection combinant 2 caractères, respectivement mesurés sur 13 770 candidats par an (**Car1**) et sur 270 collatéraux par an (**Car2**) issus de 10% des portées. Dans la situation de référence, les valeurs génétiques étaient estimées selon la méthodologie du BLUP-Modèle Animal (**BLUPMA**). Dans une première étude, nous avons comparé le scénario BLUPMA à un scénario génomique dans lequel tous les candidats étaient génotypés. Les évaluations génomiques s'appuyaient sur deux populations de référence (**PR**) initialement constituées de 13 770 candidats pour Car1 et de 1 000 collatéraux pour Car2, et dont les tailles respectives augmentaient annuellement, en considérant les mêmes capacités de phénotypage que dans le scénario BLUPMA. Les résultats montrent que des évaluations génomiques améliorent nettement la précision d'estimation des valeurs génétiques des candidats pour les deux caractères et le progrès génétique réalisé annuellement sur l'objectif global de sélection (+27% à +33% selon les héritabilités considérées), tout en réduisant significativement l'augmentation de la consanguinité dans la population. Un second scénario génomique a été simulé, dans lequel les candidats n'étaient plus phénotypés et les évaluations génomiques s'appuyaient sur une PR uniquement constituée de collatéraux phénotypés pour Car1 et Car2. Dans ce cas, la précision des valeurs génomiques estimées et la réponse à la sélection pour Car1 sont nettement plus faibles que dans le scénario BLUPMA, montrant que la sélection génomique ne permet pas de mettre fin au phénotypage des animaux. La mise en place d'évaluations génomiques nécessitant de génotyper un grand nombre d'individus, elle entraîne un surcoût important par rapport au scénario BLUPMA. Dans une seconde étude, nous avons montré que ce surcoût peut être largement réduit en présélectionnant les candidats à génotyper sur la base de leur valeur génomique estimée sur ascendance. Il est ainsi possible de réduire de manière significative le nombre de candidats à génotyper tout en préservant une grande partie de l'avantage de la sélection génomique par rapport à la sélection conventionnelle BLUPMA. Ainsi, une diminution de 40% du nombre de candidats génotypés ne réduit que de 3 à 4% le progrès génétique annuel sur l'objectif global. Nous avons également montré qu'au-delà d'un certain seuil d'investissement, une dépense supplémentaire pour améliorer l'efficacité du programme de sélection est plus efficacement investie dans la mise en place d'évaluations génomiques que dans l'augmentation de la capacité de phénotypage des collatéraux dans le dispositif conventionnel. Ce seuil d'intérêt de mise en place d'un programme génomique est d'autant plus bas que le coût du génotypage est faible et que le coût de phénotypage des collatéraux est élevé.

L'ensemble de nos résultats suggère qu'il serait intéressant de mettre en place des évaluations génomiques dans un programme de sélection d'une lignée porcine mâle de grande taille, notamment dans la population Piétrain collective française, dont la structure est proche de celle de la population simulée dans nos études.

Abstract

Interest of genomic selection in a pig sire line breeding scheme.

The aim of this work was to evaluate the interest of implementing genomic evaluations in pig breeding schemes. Stochastic simulation was used. The simulated population was a pig sire line containing 1,050 breeding females and 50 boars. The line was selected for 10 years for a breeding goal including two uncorrelated traits, recorded on, respectively, 13,770 candidates per year (**trait1**) and 270 relatives per year born in 10% of the litters (**trait2**). In the reference breeding scheme (**BLUPAM**), the selection was based on pedigree-based BLUP estimated breeding values (**EBVs**). In a first study, we compared the BLUPAM scenario to an alternative genomic breeding scheme with the same phenotyping capacities, where all candidates for selection were genotyped. The genomic breeding values for **trait1** and **trait2** were estimated using two training populations (**TP**). The first one (**TP1**) was made up of selection candidates (phenotyped for **trait1**) and the second one (**TP2**) of relatives phenotyped for **trait2**. The size of **TP1** and **TP2** increased, respectively, from 13,770 to 55,080 and from 1,000 to 3,430 over time. Our results show that genomic evaluations significantly improve the accuracy of the EBVs of the candidates for both traits and therefore the annual genetic trends for the global breeding goal (+27% to +33% depending on trait heritability), while significantly reducing the inbreeding rate. A second genomic scenario was simulated, in which the candidates were no longer phenotyped for **trait1**, and the genomic breeding values were estimated with one single **TP** made up of relatives phenotyped for both traits. In that case, the accuracy of EBVs and the annual genetic trends for **trait1** are significantly lower than in the reference (**BLUPAM**) scenario. This shows that a large **TP** is required to outperform the current schemes for traits recorded on the candidates.

The implementation of genomic evaluations requires the genotyping of a large number of animals, and therefore generates additional costs compared to **BLUPAM** breeding schemes. In a second study, we showed that genotyping a subset of candidates that have been pre-selected according to their parental EBV allows to significantly reduce the extra costs of a genomic breeding scheme while preserving most of its superiority in terms of genetic trends and inbreeding over the **BLUPAM** breeding scheme. For instance, reducing the number of genotyped candidates by 40% only reduced by 3 to 4% the global annual genetic trend. We also showed that even a very marked increase in the number of relatives phenotyped for **trait2** in a **BLUPAM** scenario does not allow to be as efficient as a genomic scenario when the number of genotyped candidates is large. Finally, we showed that the economic interest of genetic selection can be characterized by an additional cost threshold; below this threshold, it is preferable to maintain pedigree-based BLUP evaluations and increase the number of relatives, while implementing genomic evaluation is more efficient above this threshold. The value of this threshold depends on the cost of phenotyping additional relatives and on genotyping costs.

Our results suggest that implementing genomic evaluations in a large size pig sire line can be a valuable strategy. This strategy could for instance easily be applied to the French Piétrain population, which resembles the nucleus population simulated in this study.

SOMMAIRE

INTRODUCTION	7
PARTIE 1 : SYNTHÈSE BIBLIOGRAPHIQUE	15
1. L'ÉVALUATION GÉNOMIQUE : PRINCIPE ET MÉTHODES.....	17
1.1. Principe général	17
1.2. Constitution de la population de référence	20
1.3. Estimation des effets des segments chromosomiques	21
1.3.1. Méthode des moindres carrés.....	22
1.3.2. Méthodes de régression pénalisée.....	22
1.3.2.1. Le BLUP génomique	23
1.3.2.2. La Ridge Régression.....	23
1.3.2.3. Le LASSO.....	24
1.3.2.4. L'Elastic Net	24
1.3.3. Les méthodes bayésiennes	25
1.4. Estimation des valeurs génomiques.....	27
2. FACTEURS INFLUENÇANT LA PRÉCISION DE L'ÉVALUATION GÉNOMIQUE	29
2.1. Effet de la taille efficace N_e de la population considérée.....	31
2.2. Effet de la taille N_p de la population de référence.....	31
2.3. Effet de l'héritabilité du caractère considéré.....	34
2.4. Effet de la structure de la population de référence et de la distance génétique avec la population évaluée	36
2.5. Effet de la densité de marqueurs.....	39
2.6. Effet de la nature des segments chromosomiques considérés	43
2.7. Effet de la méthode d'estimation des effets des segments chromosomiques	44
3. MÉTHODE RETENUE POUR ÉVALUER L'INTÉRÊT DE METTRE EN PLACE DES ÉVALUATIONS GÉNOMIQUES DANS UN PROGRAMME DE SÉLECTION PORCIN.....	47
3.1. Modèles déterministes	47
3.2. Modèles stochastiques	48
3.3. Modèles pseudo-stochastiques.....	50
PARTIE 2. Article 1 : Perspectives d'application de la sélection génomique dans les schémas d'amélioration génétique porcins.....	51

PARTIE 3. Article 2 : Efficiency of genomic selection in a purebred pig male line.....	61
PARTIE 4. Article 3 : Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme.	77
DISCUSSION GENERALE ET PERSPECTIVES.....	95
1. RESUME DES PRINCIPAUX RESULTATS	97
2. LIMITES DE NOTRE MODELE STOCHASTIQUE	98
2.1. Longueur du génome simulé.	98
2.2. Génération du DL dans la population de base.	99
2.3. Diminution de la variabilité génétique dans la population.	100
2.4. Objectif de sélection et caractères considérés.	101
2.5. Modèle d'évaluation génétique.	102
3. RENTABILITE ECONOMIQUE DE LA MISE EN PLACE D'EVALUATIONS GENOMIQUES DANS LA POPULATION PIETRAIN COLLECTIVE FRANCAISE	104
4. INTERET DE LA SELECTION GENOMIQUE DANS LES POPULATIONS PORCINES MATERNELLES	110
5. INTERET DE LA SELECTION GENOMIQUE DANS LES POPULATIONS PORCINES DE FAIBLE EFFECTIF	112
6. SELECTION GENOMIQUE POUR LA VALEUR EN CROISEMENT	113
ANNEXES	119
ANNEXE 1	121
REFERENCES BIBLIOGRAPHIQUES	125

INTRODUCTION

Jusqu'à récemment, la grande majorité des programmes d'amélioration génétique des populations animales d'élevage s'appuyaient sur le « modèle génétique infinitésimal », qui considère le génome comme une « boîte noire » contenant un très grand nombre de gènes non identifiés ayant chacun un effet très faible sur les caractères. Le principe sous-jacent à ce modèle est qu'un individu porteur d'allèles favorables aux gènes impliqués dans le déterminisme d'un caractère aura de bonnes performances, et qu'il va transmettre à ses descendants une partie de ces allèles et donc de sa supériorité. En s'appuyant sur une méthode statistique adaptée comme le BLUP (Best Linear Unbiased Prediction ; Henderson, 1975) appliquée à un modèle individuel (Quaas et Pollak, 1980), on peut ainsi estimer la valeur génétique des individus d'une population à partir de leurs performances propres et/ou de performances d'individus apparentés et d'informations généalogiques permettant de tenir compte de la proportion d'allèles communs attendus en moyenne compte tenu de leur degré de parenté. La sélection des allèles favorables aux gènes dans la population se fait ainsi implicitement, sans avoir besoin de s'intéresser aux gènes eux-mêmes.

Chez le porc, comme dans les autres espèces, les programmes d'amélioration génétique fondés sur le modèle génétique infinitésimal sont très efficaces, comme en témoignent les évolutions génétiques réalisées au cours des précédentes décennies (voir par exemple Tribout et al. (2003 ; 2004). Cependant, en dépit de sa simplicité et de son efficacité, ce modèle présente plusieurs limites. Tout d'abord, l'information des apparentés d'un jeune individu sans descendance phénotypée permet d'estimer les composantes paternelles et maternelles de sa valeur génétique, mais pas son aléa de méiose. La précision d'estimation de sa valeur génétique est donc limitée. Par ailleurs, l'hypothèse d'un très grand nombre de gènes ayant chacun un effet très faible ne reflète qu'imparfaitement la réalité. On connaît en effet depuis longtemps des gènes ayant un effet majeur sur les performances de certains caractères, et on peut supposer qu'il existe un nombre non négligeable de gènes d'effet important. Il est alors tentant de sélectionner directement les individus de la population pour leur génotype à ces gènes, afin de fixer les allèles favorables plus rapidement qu'avec les méthodes classiques et de mieux valoriser la variabilité génétique intra famille.

Le génotype des individus à un gène d'intérêt n'est cependant en général pas accessible. En effet, le gène lui-même n'est la plupart du temps même pas identifié, tout au plus a-t-il été localisé dans une région plus ou moins précise d'un chromosome ; on parle alors de locus à effet quantitatif ou QTL (pour Quantitative Trait Locus). On est donc contraint de s'appuyer sur le phénotype des individus à un indicateur indirect du génotype à

ce gène (ou à ce QTL) pour distinguer les individus porteurs d'allèles favorables ou défavorables. Ces « marqueurs » peuvent être de nature diverse : visuels (coloration de la robe de l'animal), physiologiques (réaction de l'animal à un stimulus externe), biochimiques (protéines sanguines), ou encore des marqueurs génétiques (fragments d'ADN), et doivent présenter différentes formes dans la population, associées le plus étroitement possible aux différents allèles du gène ou du QTL. La connaissance du variant d'un marqueur porté par un individu permet ainsi de savoir indirectement, avec toutefois un risque d'erreur, si l'individu est génétiquement intéressant à sélectionner car porteur d'un allèle favorable au gène ou au QTL. Cette pratique est connue sous le nom de « sélection assistée par marqueurs » (SAM), dont l'une des toutes premières applications en sélection animale fut l'éradication de l'allèle de sensibilité à l'halothane dans la population porcine Landrace Français dans les années 1980. Ce programme est néanmoins longtemps resté la seule application de la SAM chez le porc, faute de marqueurs disponibles permettant de mettre en évidence des gènes d'effets moins marqués.

Dans les années 1990, la découverte des marqueurs microsatellites et la production dans les différentes espèces animales d'élevage de cartes génétiques (Archibald et al., 1995) constituées grâce à ces marqueurs ont attisé l'enthousiasme de la communauté scientifique pour la SAM. De nombreux programmes de recherche ont été réalisés, conduisant à la mise en évidence d'un grand nombre de QTL dans l'ensemble des espèces animales de rente (Animal QTLdb, Hu et al., 2013). Des programmes de SAM ont été mis en place avec succès dans plusieurs espèces, en France par exemple chez les bovins laitiers (Fritz et al., 2003) ou encore chez les ovins pour la résistance à la tremblante (Sidani et al., 2010). Dans l'espèce porcine en revanche, force est de constater que jusqu'à la fin des années 2000, les applications de la SAM ont été très peu nombreuses (Van der Steen et al., 2005 ; Luther et al., 2009 ; Schwob et al., 2009) en dépit des centaines de QTL localisés. Cette situation s'explique principalement par le fait que les marqueurs microsatellites, seuls marqueurs disponibles jusqu'à récemment, sont généralement éloignés des mutations causales, ce qui entraîne une diminution rapide de l'efficacité de la sélection en raison des recombinaisons entre marqueurs et QTL.

Au cours des dernières années, les progrès du séquençage des génomes humains et des principales espèces domestiques ont permis de mettre en évidence un nouveau type de marqueurs répartis sur l'ensemble du génome et en nombre extrêmement élevé (plusieurs centaines de milliers à plusieurs millions de marqueurs sur l'ensemble d'un génome) : les marqueurs SNP, dont le polymorphisme consiste en la variation d'une seule base

nucléotidique (« Single Nucleotide Polymorphism »). Grâce au développement de puces à SNP, il est aujourd'hui possible de génotyper très rapidement et pour un coût modéré (de l'ordre de 100€) un individu pour plusieurs dizaines de milliers de SNP. Avec une telle densité de marqueurs, il est possible de suivre la transmission entre générations de l'ensemble du génome, et donc *a fortiori* de tous les gènes intervenant sur un phénotype.

En 2001, Meuwissen et al. (2001) ont montré qu'en subdivisant le génome en un très grand nombre de segments déterminés par un ou plusieurs marqueurs, puis en estimant l'effet de chaque segment grâce à un échantillon d'individus phénotypés et génotypés pour ces marqueurs, on peut, en sommant les effets associés aux génotypes d'un individu pour tous ces segments, disposer d'une estimation très précise de sa valeur génétique, et ceci dès son plus jeune âge. C'est ce qu'on appelle l'évaluation génomique.

L'évaluation génomique présente potentiellement de nombreux avantages par rapport aux méthodes d'évaluation classiques :

- elle permet de disposer d'une estimation de la valeur génétique de tout individu de la population pour lequel on dispose d'un échantillon d'ADN, sans qu'il soit nécessaire de le phénotyper, ou de phénotyper un individu apparenté (parent, collatéral ou descendant) ;
- cette déconnection entre mesure des phénotypes et sélection des candidats permet plus de flexibilité dans l'organisation de la sélection, et permet en théorie la sélection sur des objectifs plus complexes, incluant par exemple des caractères non mesurables en routine ;
- sous certaines conditions, la précision d'estimation des valeurs génomiques peut être très élevée, supérieure à celle des valeurs génétiques estimées par les méthodes traditionnelles ;
- la valeur génomique d'un individu peut être estimée dès sa naissance avec une précision supérieure à celle d'une évaluation traditionnelle sur ascendance, ce qui permet de réduire fortement l'intervalle entre générations dans la population par rapport à un dispositif s'appuyant sur du testage sur descendance qui reporte nécessairement la mise en service d'un reproducteur ;
- dans le cas où le coût de génotypage d'un candidat est inférieur à celui de son testage (performances propres, contrôle de collatéraux ou testage sur descendance), l'évaluation génomique permet d'augmenter le nombre de candidats à la sélection et l'intensité de sélection des reproducteurs de renouvellement, sur la voie mâle comme sur la voie femelle.

Le progrès génétique réalisé annuellement dans une population en sélection étant proportionnel à la précision d'estimation des valeurs génétiques des candidats et à l'intensité de sélection pratiquée, et inversement proportionnel à l'intervalle entre générations (Falconer,

1981), la sélection génomique est donc potentiellement source d'un gain d'efficacité important dans les programmes de sélection.

Cette nouvelle méthodologie est utilisée dans les évaluations de populations bovines laitières dans plusieurs pays depuis 2009 (Hayes et al., 2009a). Dans cette espèce, l'évaluation génomique permet désormais de disposer d'une estimation précise de la valeur génétique des jeunes candidats mâles sans plus avoir à mettre en œuvre un testage sur descendance long et coûteux. Les sélectionneurs peuvent dorénavant sélectionner plus précocement les reproducteurs parmi un nombre plus grand de candidats à la sélection, tout en conservant un niveau élevé de précision des valeurs génétiques estimées. L'augmentation des intensités de sélection et la forte diminution de l'intervalle entre générations permises par le passage du testage sur descendance à l'évaluation génomique, ainsi que l'augmentation de la précision d'évaluation des femelles, ont considérablement accru le progrès génétique réalisé annuellement. De plus, les économies réalisées grâce à l'arrêt du testage sur descendance ont permis de financer le génotypage des animaux, si bien que cette amélioration d'efficacité s'est faite à coût approximativement constant.

Un tel succès est tentant pour les acteurs des programmes de sélection des autres espèces animales de rente. Chez le porc, la disponibilité depuis 2009 d'une puce de 60 000 marqueurs SNP (*porcineSNP60* Beadchip de la société Illumina) permet d'envisager la mise en œuvre d'évaluations génomiques, et des premières études évaluant la précision de prédictions génomiques dans certaines populations porcines étrangères ont été publiées (Ostersen et al., 2011 ; Cleveland et al., 2012). Une sélection génomique nécessite toutefois de génotyper un grand nombre d'individus, ce qui génère un surcoût important sans qu'aucune source importante d'économie sur le programme de sélection ne soit identifiée par ailleurs. Avant de s'engager dans une telle évolution dans les populations porcines françaises, il est donc important de s'interroger sur l'intérêt de l'évaluation génomique et de quantifier le gain d'efficacité attendu de cette nouvelle méthodologie. Ces interrogations sont à l'origine de ce travail de thèse.

La première partie de ce mémoire sera consacrée à la présentation du principe général de l'évaluation génomique, des principales méthodes permettant d'estimer la valeur génomique des individus, ainsi qu'à l'inventaire des facteurs influençant la précision d'une telle sélection.

Dans la deuxième partie, nous présenterons l'organisation actuelle des programmes d'amélioration génétique porcins, en mettant en évidence leurs forces et leurs faiblesses. Nous

inventorierons les aspects des programmes porcins susceptibles d'être améliorés par la mise en place d'évaluations génomiques.

Les troisième et quatrième parties de ce mémoire présenteront les deux études réalisées durant cette thèse, visant à comparer par simulation les efficacités technique et économique d'un programme de sélection d'une lignée porcine de type mâle s'appuyant sur des évaluations génétiques BLUP-modèle animal à celles d'un programme de sélection s'appuyant sur des évaluations génomiques.

PARTIE 1 :
SYNTHESE BIBLIOGRAPHIQUE

1. L'EVALUATION GENOMIQUE : PRINCIPE ET METHODES

1.1. Principe général

La variabilité génétique d'un caractère est déterminée par un nombre inconnu de gènes existant sous plusieurs formes (appelées « allèles »), répartis sur le génome des individus. Certains de ces allèles ont un effet favorable sur le phénotype du caractère, d'autres ont un effet défavorable. En supposant que les allèles d'un gène ou de plusieurs gènes n'interagissent pas entre eux (absence de dominance et d'épistasie, respectivement), la valeur génétique d'un individu pour un caractère donné est donc égale à la somme des effets sur ce caractère de chacun des allèles portés par son génome. Si tous les gènes affectant un caractère étaient identifiés, il suffirait de génotyper pour l'ensemble des mutations causales un échantillon d'individus phénotypés, et d'estimer l'effet des différents génotypes à ces mutations pour pouvoir ensuite calculer la valeur génétique de n'importe quel animal de la population connaissant son génotype. En pratique cependant, les plupart des gènes responsables de la variabilité génétique d'un caractère restent aujourd'hui encore inconnus. Plusieurs auteurs (Lande et Thompson, 1990 ; Haley et Visscher, 1998) ont proposé de contourner ce problème en s'appuyant sur l'association entre le génotype (connu) des individus à des marqueurs moléculaires et leur génotype (inconnu) aux mutations causales pour les caractères d'intérêt.

Pour expliquer l'origine de cette association, considérons un individu fondateur mis à la reproduction il y a plusieurs générations qui a eu de nombreux descendants dans une population de taille finie (Figure 1).

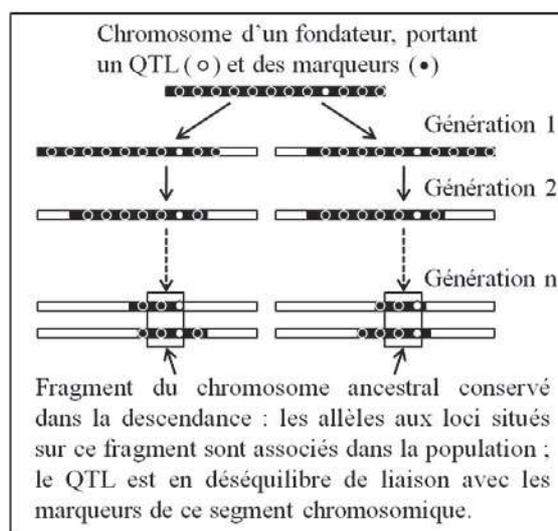


Figure 1 : Transmission de segments chromosomiques entre un fondateur et ses descendants (d'après Hayes (2011)).

Cet individu a transmis à ses produits des gamètes issus de ses chromosomes et modélés par des événements de recombinaison *a priori* aléatoires. Au fur et à mesure des générations et au gré des événements de recombinaison, des segments de taille décroissante du chromosome ancestral ont été conservés et sont actuellement portés par un grand nombre d'individus de la population. En conséquence, les allèles aux loci présents dans cette région du génome hérités de cet individu ancestral sont statistiquement associés dans la population ; autrement dit, le génotype à un locus de la région est corrélé au génotype à un autre locus de la région. Cette association préférentielle entre allèles à différents loci est appelée « déséquilibre de liaison » (DL). La force de cette association peut être quantifiée par divers indicateurs, dont l'un des plus couramment utilisés est le paramètre r^2 (Hill et Robertson, 1968) qui varie entre 0 (indépendance entre les allèles à deux loci) et 1 (les allèles aux deux loci sont totalement liés). Si on considère dans une population deux loci bialléliques A et B, le paramètre r^2 entre ces deux loci se calcule selon la formule :

$$r^2 = \frac{(freq(A_1B_1)*freq(A_2B_2) - freq(A_1B_2)*freq(A_2B_1))^2}{freq(A_1)*freq(A_2)*freq(B_1)*freq(B_2)} \text{ (Hill et Robertson, 1968)}$$

où $freq(A_1)$, $freq(A_2)$, $freq(B_1)$ et $freq(B_2)$ sont, respectivement, les fréquences dans la population considérée des allèles 1 et 2 aux loci A et B, et $freq(A_1B_1)$, $freq(A_1B_2)$, $freq(A_2B_1)$ et $freq(A_2B_2)$ sont, respectivement, les fréquences dans la population considérée des haplotypes A_1B_1 , A_1B_2 , A_2B_1 et A_2B_2 .

Un exemple de calcul du DL (r^2) est présenté sur la figure 2.

En moyenne, le DL sera élevé pour deux loci très proches, puisque la probabilité qu'une recombinaison génétique intervienne sur un intervalle très petit est faible. A l'inverse, plus les loci sont distants, plus la probabilité que des événements de recombinaison se produisent au fur et à mesure des générations sur cet intervalle est élevée, et plus le DL est faible dans une population panmictique de grande taille. Toutefois, un DL élevé peut exister entre loci distants ou même non liés, dans le cas par exemple de l'existence d'une structuration de la population.

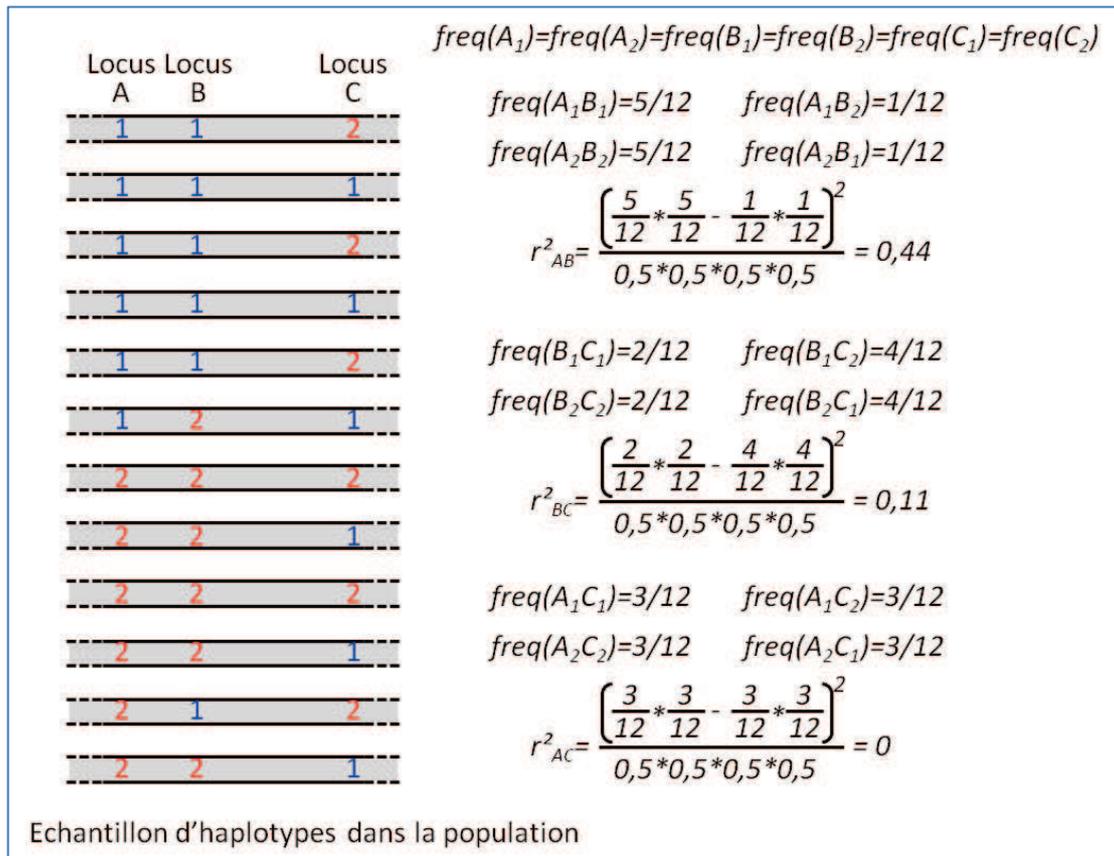


Figure 2 : Calcul du r^2 entre loci.

A, B et C sont 3 loci bi-alléliques situés sur un même chromosome. Les loci A et B sont en déséquilibre de liaison moyen. Les loci B et C présentent un faible déséquilibre de liaison. Les loci A et C sont en équilibre de liaison.

L'étendue du DL dans une population fermée, autrement dit la force de l'association entre les génotypes à des loci éloignés d'une certaine distance, dépend de l'effectif génétique (N_e) actuel et passé de la population considérée. Cet effectif théorique N_e , défini par Wright (1931), correspond au nombre efficace de reproducteurs d'une population fictive idéale (en situation d'accouplements panmictiques et de distribution équilibrée des effectifs de descendants) dans laquelle l'augmentation de consanguinité (ou le taux d'évolution des fréquences des gènes) se fait au même rythme que dans la population réelle considérée.

Une relation entre le N_e d'une population et l'espérance du r^2 entre 2 loci éloignés d'une distance de c (exprimée en Morgan) a été établie par (Sved, 1971) : $E(r^2) = \frac{1}{4N_e c}$.

Dans une population de faible N_e , un reproducteur va transmettre en espérance des fragments de son génome à une proportion importante de la population, résultant en une faible variété d'haplotypes. Les recombinaisons entre segments identiques seront sans effet sur le DL. On observera donc un DL de niveau plus élevé à longue distance que dans une population de plus

grand effectif génétique. Etant donné que le DL entre loci distants tend à décroître rapidement au cours des générations, le DL entre 2 loci éloignés rend compte de l'effectif génétique de la population dans un passé proche. A l'inverse, le DL entre 2 loci proches rend compte du N_e de la population il y a un grand nombre de générations. Hayes et al. (2003) ont estimé que la valeur de N_e estimée à l'aide de la formule de Sved (1971) pour deux loci éloignés d'une distance de c Morgan correspondait à l'effectif génétique de la population il y a approximativement $1/2c$ générations.

Le génome des individus d'une population peut donc être considéré comme un ensemble de segments chromosomiques issus de reproducteurs ancestraux et conservés au cours des générations car transmis sans recombinaison. Si un de ces segments porte un gène et un ou plusieurs marqueurs moléculaires, la connaissance du génotype des individus à ce(s) marqueur(s) renseigne donc de façon indirecte sur leur génotype au gène d'intérêt, la qualité de cette information déduite étant d'autant meilleure que le DL entre marqueur(s) et gène est élevé.

C'est sur cette association entre marqueurs moléculaires (connus) et gènes (inconnus) que s'appuie la sélection assistée par marqueurs et l'évaluation génomique, dont l'objectif est d'utiliser les marqueurs pour prédire la valeur génétique des candidats à la sélection. Le principe de base - établi par Lande et Thompson (1990) pour les régions QTL, puis étendu par Haley et Visscher (1998) à l'ensemble du génome - est de « découper » le génome en un grand nombre de segments chromosomiques déterminés par un ou plusieurs marqueurs moléculaires en déséquilibre de liaison avec les gènes impactant le phénotype au caractère considéré, d'estimer l'effet de chacun de ces segments dans une population de référence, puis de prédire, à partir des effets précédemment estimés, la valeur génétique des animaux qui ne font pas partie de cette population de référence, tels que de jeunes candidats à la sélection.

1.2. Constitution de la population de référence

Concrètement, l'étape d'estimation des effets chromosomiques nécessite la constitution d'un groupe d'individus - la « population de référence » (PR) - pour lesquels on dispose à la fois du génotype pour un très grand nombre de marqueurs et du phénotype pour les caractères pour lesquels on souhaite réaliser une sélection. La densité de marqueurs doit être suffisamment grande pour que tous les loci contribuant à la variabilité génétique d'un caractère présentent un déséquilibre de liaison suffisant avec les marqueurs qui leur sont proches (ou avec une combinaison de ces marqueurs). En pratique, dans les populations

animales, plusieurs dizaines de milliers de marqueurs répartis sur l'ensemble du génome semblent nécessaires. Cette PR doit être représentative de la population que l'on souhaite sélectionner (dans l'idéal, l'ensemble des allèles des différents loci ségrégeant dans la population doivent y être représentés), et d'effectif suffisamment grand pour permettre une estimation précise de l'effet des segments chromosomiques, car de celle-ci dépendra la précision de la prédiction des valeurs génomiques des candidats.

Les segments chromosomiques dont on estime l'effet peuvent être déterminés chacun par le génotype à un marqueur unique ou par un haplotype à plusieurs marqueurs adjacents. Pour des raisons de simplicité de mise en œuvre, la plupart des applications d'évaluation génomiques actuelles s'appuient sur des marqueurs individuels. Toutefois, afin de conserver un caractère général, on parlera dans les paragraphes suivants de « segments chromosomiques » pouvant correspondre à l'un ou l'autre cas.

1.3. Estimation des effets des segments chromosomiques

Une fois la PR constituée, la principale difficulté consiste à estimer un très grand nombre d'effets de segments chromosomiques ($M =$ plusieurs dizaines de milliers) à partir d'un nombre en général beaucoup plus faible ($n =$ quelques centaines à quelques milliers) de phénotypes, ce qui revient à vouloir résoudre le système ci-dessous contenant plus d'inconnues que d'équations:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{X}\mathbf{g} + \mathbf{e}$$

où \mathbf{y} est le vecteur contenant le phénotype y_i de chacun des n individus de la PR, $\mathbf{1}_n$ est un vecteur unité de n éléments, μ est la moyenne générale, \mathbf{X} est la matrice contenant les vecteurs \mathbf{X}_i des génotypes de chaque individu i pour chacun des M segments chromosomiques, \mathbf{g} est le vecteur des effets des génotypes de chacun des M segments chromosomiques sur le caractère, et \mathbf{e} est le vecteur des résiduelles non expliquées.

Pour cela, plusieurs méthodologies statistiques ont été proposées, présentant des coûts de calcul et des efficacités très variables. Ces méthodes peuvent s'appliquer à l'estimation des effets de marqueurs individuels ou d'haplotypes, selon que chaque segment de chromosome est déterminé par un seul marqueur ou par plusieurs marqueurs contigus. Dans le premier cas, on estimera un seul effet par marqueur biallélique, alors que dans le second cas, plusieurs effets devront être estimés pour chaque segment.

Des synthèses détaillées comparant les différentes méthodes d'évaluation génomique disponibles ont été réalisés par de los Campos et al. (2013) et Robert-Granié et al. (2011).

Nous nous contenterons ici de présenter brièvement le principe des plus communes d'entre elles.

1.3.1. Méthode des moindres carrés

La méthode des moindres carrés est fréquemment utilisée en statistique pour trouver les valeurs des paramètres décrivant le mieux les données (minimisant l'écart entre performances observées et performances prédites à l'aide de ces paramètres). Cependant, elle ne peut s'appliquer que lorsque le nombre de paramètres à estimer est inférieur au nombre d'observations. Dans le cas de l'évaluation génomique, on est donc amené dans un premier temps à sélectionner un sous ensemble de M' segments chromosomiques parmi M (avec $M' < n$) en estimant l'effet de chacun des M segments par régression dans un modèle univarié, et en ne conservant que les M' segments dont l'effet est le plus significatif. Dans un second temps, les effets \hat{g}_m de ces M' segments sont ré-estimés dans un modèle de régression multiple comme ci-dessous, les segments non conservés étant supposés avoir un effet nul :

$$\begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix},$$

où \mathbf{X} est la matrice d'incidence reliant les M' génotypes (ou haplotypes) aux n phénotypes et $\mathbf{1}_n$ est un vecteur unité de n éléments.

Cette méthode, très simple d'un point de vue calculatoire, pose cependant les problèmes du choix du seuil de significativité qui ne doit pas être trop laxiste afin de ne retenir dans le modèle que les marqueurs liés aux QTL mais permettre néanmoins de capter une part importante de la variation. Au-delà de ce problème de puissance, l'utilisation d'un seuil conduit à une forte surestimation des effets estimés de certains segments du fait du biais de sélection (« effet Beavis » ; Xu, 2003). Pour ces raisons, cette méthode, préconisée par Lande et Thompson (1990), ne permet pas de prendre en compte dans les modèles de prédiction la variation expliquée par l'ensemble des QTL présents sur le génome, s'avère beaucoup moins performante que les méthodes décrites ci-dessous et n'est en pratique pas utilisée comme méthode d'évaluation génomique.

1.3.2. Méthodes de régression pénalisée

Afin de contourner ce problème de surestimation lié à la pré-sélection des segments à conserver, il est possible d'estimer les effets de tous les segments chromosomiques simultanément (même lorsque leur nombre est supérieur à celui des observations), en ajoutant à la méthode des moindres carrés une contrainte sur les effets du modèle. Cette contrainte a

pour conséquence de régresser vers 0 (ou de « pénaliser ») les estimations des effets. La pénalisation sur l'effet d'un segment chromosomique est d'autant plus forte que la quantité d'information disponible pour estimer cet effet est faible.

Diverses méthodes de pénalisation ont été proposées, dont les plus courantes sont décrites ci-dessous.

1.3.2.1. Le BLUP génomique

Dans la méthode du BLUP génomique, ou GBLUP (Meuwissen et al., 2001), on suppose que les effets de tous les segments chromosomiques qu'on cherche à estimer sont des effets aléatoires tirés dans une même loi normale centrée sur 0 et de variance σ_g^2 . On estime ces effets par la méthode du BLUP en résolvant le système suivant :

$$\begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \sigma_e^2 / \sigma_g^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}$$

où σ_e^2 est la variance résiduelle du modèle. Les variances σ_g^2 et σ_e^2 peuvent être directement estimées avec ce modèle sur les données de la population de référence. Une autre approche peut être d'estimer dans un premier temps la variance génétique totale σ_G^2 et la variance résiduelle σ_e^2 par REML (Restricted Maximum Likelihood) en s'appuyant sur les pedigrees, puis de déduire une estimation de la variance σ_g^2 des effets des SNP en divisant σ_G^2 par le nombre n de SNP, ou selon la formule $\sigma_g^2 = \frac{\sigma_G^2}{2 \sum_{i=1}^n p_i q_i}$ (Habier et al., 2007a), où p_i et q_i sont les fréquences alléliques du $i^{\text{ème}}$ SNP.

1.3.2.2. La Ridge Régression

Les effets des segments chromosomiques \hat{g}_m estimés par la méthode classique des moindres carrés minimisent la somme des carrés des résidus du modèle décrivant les données $\{\sum_i (y_i - \mu - \sum_m x_{im} g_m)^2\}$, où x_{im} est le génotype de l'individu i au $m^{\text{ème}}$ segment chromosomique, et g_m est l'effet du $m^{\text{ème}}$ segment chromosomique. La « Ridge Regression » (Hoerl et Kennard, 1970) consiste à ajouter à la méthode des moindres carrés une contrainte sur la somme des carrés des effets estimés. Ceux-ci minimisent alors la somme $\{\sum_i (y_i - \mu - \sum_m x_{im} g_m)^2 + L \sum_m g_m^2\}$, et sont également les solutions du système :

$$\begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + L \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}.$$

On constate que la méthode GBLUP représente un cas particulier de la Ridge Regression où $L = \sigma_e^2 / \sigma_g^2$. Cependant, contrairement au cas du GBLUP dans lequel les variances σ_g^2 et σ_e^2 ont une signification génétique, le paramètre L de la Ridge Regression est déterminée plus ou moins arbitrairement, par exemple par validation croisée sur les données.

1.3.2.3. Le LASSO

Le LASSO (Least Absolute Shrinkage and Selection Operator ; Tibshirani, 1996) est une méthode de régression qui ajoute à la méthode des moindres carrés une contrainte sur la somme de la valeur absolue des effets estimés des segments chromosomiques. Les paramètres estimés minimisent donc l'expression $\{\sum_i (y_i - \mu - \sum_m x_{im} g_m)^2 + L \sum_m |g_m|\}$, où la valeur de L peut être déterminée par validation croisée sur les données.

Cette contrainte a pour conséquence, non seulement de régresser les estimations vers 0 en fonction de la quantité d'information disponible comme avec la Ridge Regression, mais également de mettre à 0 certains effets de segments chromosomiques. Le nombre de segments ayant un effet nul sera d'autant plus élevé que la contrainte mise sur les paramètres sera forte. Le LASSO est donc une méthode qui sélectionne les segments ayant un effet sur la valeur génomique estimée des individus évalués, contrairement à la Ridge Regression et au GBLUP qui conduisent à l'estimation d'un effet pour chacun des segments initialement considérés. Pour un nombre d'observations donné, la variance des estimations obtenues par LASSO est donc plus faible que celle des estimations obtenues par Ridge Regression.

La méthode du LASSO ne permet cependant de conserver qu'un nombre maximal de paramètres égal au nombre d'observations dans la PR, ce qui peut ne pas être pertinent dans le cas de caractères à déterminisme génétique complexe. De plus, lorsque plusieurs segments chromosomiques sont très corrélés entre eux, ce qui est le cas en situation de déséquilibre de liaison important à grande distance dans la population, le LASSO ne retient qu'un seul segment choisi arbitrairement dans le groupe, à l'inverse de la Ridge Regression avec laquelle l'effet global de la région se distribue sur ses différents segments.

1.3.2.4. L'Elastic Net

L'Elastic Net (Zhou et Hastie, 2005) est une méthode de régression qui ajoute à la méthode des moindres carrés une double contrainte sur la somme des carrés (comme la Ridge Regression) et sur la somme des valeurs absolues (comme le LASSO) des effets estimés des

segments chromosomiques. Les paramètres estimés par l'Elastic Net minimisent donc l'expression :

$$\{\sum_i (y_i - \mu - \sum_m x_{im} g_m)^2 + L(\alpha \sum_m |g_m| + (1 - \alpha) \sum_m g_m^2)\},$$

où L et α peuvent être choisis par validation croisée sur les données.

Comme le LASSO, l'Elastic Net conduit à une sélection de segments chromosomiques ayant un effet non nul, tout en conservant les avantages de la Ridge Regression.

1.3.3. Les méthodes bayésiennes

Les méthodes GBLUP et Ridge Regression décrites précédemment font l'hypothèse que tous les segments chromosomiques ont un effet sur le caractère considéré, ces effets étant tous issus d'une même loi normale de très faible variance, approximativement égale à la variance génétique totale divisée par le nombre de segments. Cette hypothèse est cohérente avec le modèle infinitésimal supposant un déterminisme génétique des caractères régi par une "infinité" de loci de très faible effet. En conséquence, ces méthodes régressent très fortement les effets estimés des segments chromosomiques vers la moyenne, et ne permettent pas à certains segments d'avoir un effet important s'écartant d'une distribution normale.

Cependant, de nombreux programmes expérimentaux de détection de QTL ou de validation de gènes candidats conduits dans les différentes espèces animales ou végétales ont permis de mettre en évidence pour la plupart des caractères un nombre plus ou moins grand de loci ayant un effet important. De plus, on connaît pour certains caractères quelques gènes à effet majeur, comme le gène de sensibilité à l'Halothane pour la quantité de muscle dans la carcasse et la qualité de la viande chez le porc, ou le gène *DGATI* pour la composition en matière grasse du lait chez les bovins. A l'inverse, il est probable qu'un grand nombre de régions du génome n'a aucun effet sur le déterminisme d'un caractère donné. Il semble donc en réalité que la variabilité génétique de la plupart des caractères quantitatifs résulte en fait d'un petit nombre de loci d'effet important, de quelques loci d'effet moyen à faible, et d'un grand nombre de loci ayant un effet très faible ou nuls (Hayes et Goddard, 2001). Dans ce contexte, les méthodes GBLUP et Ridge Régression ne semblent donc pas optimales.

Des méthodes bayésiennes ont été proposées (Meuwissen et al., 2001) afin de permettre d'estimer les effets des segments chromosomiques en respectant une distribution *a priori* plus réaliste de ces effets. Concrètement, on décrit, comme pour le GBLUP, la distribution de l'effet de chaque segment chromosomique par le biais d'une loi dépendant de

paramètres (en général une loi normale); cependant, contrairement au GBLUP, la variance de la distribution de l'effet des segments n'est pas fixe et identique pour tous les segments, mais peut varier d'un segment à l'autre, selon une distribution justement choisie pour refléter la connaissance *a priori* de l'utilisateur. Cette information *a priori* sur la distribution des effets des loci est ensuite combinée avec les données réellement observées sur la PR (phénotypes et géotypes / haplotypes aux segments chromosomiques), afin d'obtenir de meilleures estimations (dites estimations *a posteriori*) des effets des segments chromosomiques que celles qui pourraient être obtenues en considérant uniquement les observations.

Cependant, l'obtention de la distribution conjointe *a posteriori* de tous les paramètres est généralement trop complexe pour être obtenue directement, et il est nécessaire de recourir à des techniques d'échantillonnages conditionnels successifs et répétés de chaque paramètre, supposant connus les observations et l'état des autres paramètres du modèle à cet instant. En pratique, plusieurs dizaines de milliers de cycles de tirages sont nécessaires. Au final, la distribution des tirages successifs d'un même paramètre représente sa distribution *a posteriori*, dont on peut déduire, par exemple, sa valeur la plus probable. Le gain de précision dans l'estimation des effets des segments chromosomiques par rapport aux méthodes de régression pénalisée est donc obtenu au prix d'un accroissement très important des temps de calcul.

Plusieurs méthodes bayésiennes d'estimation des effets des segments chromosomiques ont été développées, variant entre elles au niveau de l'information considérée *a priori*. Chronologiquement, la première méthode développée, appelée « Bayes A », fait l'hypothèse que chaque QTL a sa variance propre, qu'un grand nombre de QTL ont un effet faible et que quelques QTL ont un effet important (Meuwissen et al., 2001). Ceci se traduit par l'utilisation d'une distribution de Chi^2 inversée comme distribution *a priori* des variances des effets des segments chromosomiques. En réalité cependant, beaucoup des segments chromosomiques ne contiennent aucun QTL et n'ont donc strictement aucun effet sur le caractère. La distribution *a priori* de la méthode « Bayes A », qui ne permet pas à un nombre suffisamment grand de segments d'avoir un effet nul, n'est donc pas réellement adéquate. (Meuwissen et al., 2001) ont alors proposé une autre méthode, nommée « Bayes B », avec laquelle une proportion π de segments n'a aucun effet (on leur impose une variance nulle), une distribution de Chi^2 inversée étant à nouveau utilisée comme distribution *a priori* des variances des effets des $(1-\pi)$ autres segments chromosomiques.

De nombreuses variantes à ces méthodes ont été ensuite proposées (Gianola, 2013). Une des plus populaires actuellement est la méthode « Bayes C π » (Habier et al., 2011), qui fait l'hypothèse que seule une proportion (1- π) de segments a un effet comme dans le « Bayes B », mais que la variance de la loi de distribution des effets de ces segments est fixe et identique pour tous ces segments. Les méthodes BayesC et BayesC π diffèrent dans le fait que le paramètre π est fixé ou estimé. On peut également citer la méthode « Bayes R » (Erbe et al., 2012), qui est une généralisation de la méthode précédente, dans laquelle on distingue n+1 catégories de segments :

- une proportion π de segments n'ayant aucun effet,
 - une proportion π_i de segments ($i \in [1, n]$) ayant un effet tiré dans une loi de variance σ_{gi}^2
- avec $\pi + \sum_i \pi_i = 1$.

1.4. Estimation des valeurs génomiques

Une fois les effets des segments chromosomiques estimés, on peut calculer la valeur génomique de n'importe quel individu de la population pour lequel on dispose du génotype aux marqueurs considérés dans la PR, en sommant simplement les effets estimés des génotypes (ou haplotypes) portés par l'individu aux différents segments :

$$\widehat{VG}_i = \sum_m X_{i,m} \hat{g}_m$$

où $X_{i,m}$ indique le génotype (ou l'haplotype) de l'individu i au segment chromosomique m et \hat{g}_m est l'effet estimé du segment chromosomique m .

Estimation directe de la valeur génomique de candidats par BLUP génomique :

Dans la méthode GBLUP décrite ci-dessus, la valeur génomique estimée d'un animal (appartenant à la PR ou pas) était calculée comme la somme des effets estimés des segments chromosomiques de son génome :

$$\widehat{VG}_i = \sum_m X_{i,m} \hat{g}_m,$$

eux-mêmes préalablement estimés sur la PR en résolvant le système :

$$\begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{X} \\ \mathbf{X}' \mathbf{1}_n & \mathbf{X}' \mathbf{X} + \sigma_e^2 / \sigma_g^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{X}' \mathbf{y} \end{bmatrix}.$$

Une autre approche (Habier et al., 2007a ; VanRaden, 2008) conduisant aux mêmes valeurs génomiques estimées pour les individus consiste à estimer directement la valeur génomique d'un individu génotypé (qu'il appartienne ou pas à la PR) sans estimer les effets des segments chromosomiques, mais en résolvant un système analogue au système des équations du modèle mixte obtenu lorsque les performances sont décrites avec un modèle animal individuel. Dans ce système, la matrice de parenté entre individus (A) classiquement utilisée est remplacée par une « matrice de parenté génomique » (G) décrivant la parenté « réelle » entre les individus observée à travers leurs génotypes aux marqueurs :

$$\begin{bmatrix} \mathbf{1}_n' \mathbf{1}_n & \mathbf{1}_n' \mathbf{Z} \\ \mathbf{Z}' \mathbf{1}_n & \mathbf{Z}' \mathbf{Z} + \alpha \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n' \mathbf{y} \\ \mathbf{Z}' \mathbf{y} \end{bmatrix}$$

où $\hat{\mathbf{a}}$ est le vecteur des valeurs génomiques estimées des individus génotypés (identiques aux \widehat{VG}_i ci-dessus), \mathbf{y} est le vecteur des phénotypes des individus de la PR, \mathbf{Z} est la matrice d'incidence reliant les performances aux valeurs génomiques, et α est égal au rapport de la variance résiduelle sur la variance génétique de la population.

Alors que la matrice de parenté A habituellement utilisée pour estimer les valeurs génétiques contient pour chaque paire d'individus évalués leur parenté espérée compte tenu de leur pedigree, les coefficients de parenté entre les paires d'individus contenus dans la matrice de parenté génomique G sont calculés en fonction du nombre d'allèles communs entre individus sur l'ensemble des marqueurs. Ainsi, si on considère 3 germains I, J et K (issus des mêmes père et mère), les 3 coefficients de parenté entre I et J, entre I et K et entre J et K seront identiques dans la matrice A (par exemple 0,5), mais pourront différer dans la matrice G , par exemple si les individus I et J ont reçu plus d'allèles identiques de leurs parents que l'individu K, comme illustré sur la figure 3. Dans ce cas, I et J seront considérés plus apparentés entre eux que I et K et que J et K. Cette meilleure estimation de la parenté réelle entre individus permet donc une estimation plus précise de la valeur génétique des individus que lorsque l'évaluation génétique s'appuie sur des parentés moyennes déduites du pedigree comme dans une évaluation BLUP-modèle animal (Habier et al., 2007a).

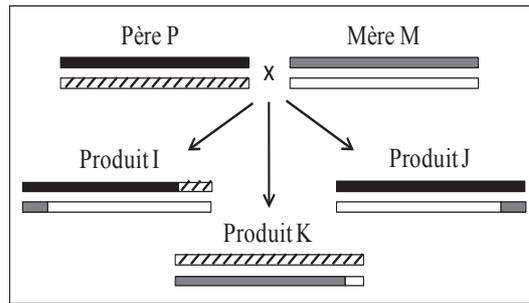


Figure 3 : Génotype à un chromosome de 3 individus I, J et K issus des mêmes parents P et M. I et J partagent une plus grande quantité de matériel génétique entre eux qu'avec K.

Les valeurs génomiques estimées par GBLUP selon ces deux approches étant au final identiques, l'utilisation du modèle estimant les effets des marqueurs sera préférable dans le cas où le nombre de SNP est inférieur au nombre d'individus génotypés. A l'inverse, lorsque le nombre de marqueurs excède l'effectif d'individus génotypés, l'estimation directe des valeurs génomiques sera plus rapide.

Un grand nombre de méthodes sont donc disponibles pour estimer la valeur génomique des animaux, s'appuyant sur des hypothèses différentes quant à la distribution des effets des segments chromosomiques considérés. Des éléments de comparaison de l'efficacité des principales méthodes seront donnés dans la section suivante.

2. FACTEURS INFLUENÇANT LA PRECISION DE L'ÉVALUATION GÉNOMIQUE

La précision d'estimation des valeurs génomiques des individus évalués est un facteur déterminant de l'efficacité d'un programme de sélection s'appuyant sur des évaluations génomiques. Daetwyler et al. (2008) et Goddard (2009) ont proposé des formules déterministes pour prédire la corrélation entre la valeur génomique estimée et la valeur génétique vraie d'un individu (cette corrélation, où « accuracy », sera notée « Acc » par la suite). Dans ces formules, la précision des valeurs génomiques estimées dépend de la taille de la population de référence (N_p), de l'héritabilité du caractère considéré (h^2), du nombre de segments chromosomiques indépendants dans le génome de la population (M_e), et de la taille efficace de la population considérée (N_e) pour la formule de Goddard :

$$Acc = \sqrt{\frac{N_p h^2}{N_p h^2 + M_e}} \text{ (Daetwyler et al., 2008) ;}$$

$$Acc = \sqrt{1 - \frac{\lambda}{2N_p \sqrt{a}} \ln\left(\frac{1+a+2\sqrt{a}}{1+a-2\sqrt{a}}\right)}, \text{ où } a = 1 + 2\frac{\lambda}{N_p}, \lambda = \frac{M_e k}{h^2}, \text{ et } k = \frac{1}{\ln(2N_e)} \text{ (Goddard, 2009).}$$

Ces deux formules produisent des résultats proches, bien que les précisions prédites en s'appuyant sur la formule de Daetwyler soient plus faibles pour des caractères d'héritabilités faibles à moyennes (Hayes et al., 2009b).

Evaluation de M_e . Le nombre de segments indépendants M_e est une entité théorique qui représente le nombre d'effets d'haplotypes indépendants que l'on doit estimer à l'aide de la population de référence dont on dispose. M_e dépend de la structure génétique de la population considérée, et notamment de l'étendue du déséquilibre de liaison dans la population. En effet, si un déséquilibre de liaison important existe entre des loci éloignés, les individus de la population partagent des segments génomiques de grande taille, et le nombre de segments indépendants dont il faut estimer l'effet est réduit. A l'inverse, dans une population dans laquelle le déséquilibre de liaison disparaît rapidement lorsque la distance entre loci augmente, les génotypes à des loci peu éloignés sont moins corrélés entre eux, et le nombre de segments indépendants est plus grand.

Comme énoncé précédemment, l'étendue du déséquilibre de liaison, et par conséquent la valeur de M_e , sont liés à la taille efficace de la population considérée. Plusieurs formules de prédiction de M_e ont été proposées, dépendant de la longueur du génome (L , en Morgan) et de N_e :

$$(a) M_e = \sum_i \frac{2N_e L_i}{\ln(4N_e L_i)} \text{ (Goddard, 2009), où } L_i \text{ est la longueur du chromosome } i \text{ (avec}$$

$$L = \sum_i L_i) ;$$

$$(b) M_e = 4N_e L \text{ (Goddard, 2009 ; Hayes et al., 2009c)}$$

$$(c) M_e = 2N_e L \text{ (Hayes et al., 2009c).}$$

Ces trois formules donnent cependant des estimations très différentes pour M_e . Ainsi, en considérant une population de taille efficace $N_e=100$ et un génome de longueur $L=30M$ constitué de 18 paires de chromosomes de tailles identiques, on obtient respectivement les valeurs de 923, 12 000 et 6 000 pour M_e . En utilisant ces trois valeurs de M_e dans la formule de Daetwyler et al. (2008) pour un caractère d'héritabilité 0,3 et une population de référence de taille $N_p=1\ 000$, on obtient des estimations de précision de valeurs génomiques estimées

sensiblement différentes (0,50 , 0,16 et 0,22 , respectivement). Hayes et al. (2009b) ont cependant observé une bonne adéquation entre la précision de valeurs génomiques estimées sur données réelles dans des populations bovines laitières Holstein américaines et australiennes et la précision théorique prédite en considérant que $M_e = 2N_eL$, suggérant d'utiliser cette expression de M_e .

La figure 4, établie à l'aide de la formule de Daetwyler et al. (2008) en considérant que $M_e = 2N_eL$, présente la précision des valeurs génomiques estimées en considérant différentes tailles de populations de référence, de valeurs d'héritabilité et de tailles efficaces. On constate d'après cette figure que les trois paramètres N_p , h^2 et N_e impactent fortement la précision d'estimation des valeurs génomiques.

2.1. Effet de la taille efficace N_e de la population considérée

Pour une taille de population de référence et une héritabilité données, la taille efficace de la population a un effet négatif sur la précision. Ainsi, pour $h^2=0,3$ et $N_p=5\ 000$, la précision prédite des valeurs génomiques estimées est égale à 0,58 et 0,33 pour des populations de tailles efficaces $N_e=50$ et 200, respectivement. En effet, une population de grande taille efficace présente un déséquilibre de liaison qui décroît rapidement en fonction de la distance entre loci et donc une variabilité haplotypique importante, ce qui nécessite d'estimer un plus grand nombre d'effets de segments chromosomiques indépendants à partir d'une même quantité d'information. A taille de population de référence donnée, la variance d'estimation de chaque effet estimé est donc plus grande que dans le cas d'une population de faible taille efficace.

2.2. Effet de la taille N_p de la population de référence

Pour une taille efficace de population (et donc un nombre de segments indépendants) et une héritabilité données, la précision augmente logiquement avec la taille de la population de référence. En effet, l'augmentation de la quantité d'information disponible permet de réduire la variance d'estimation de l'effet de chaque segment chromosomique. Par exemple, si $N_e=100$ et $h^2=0,3$, la précision des valeurs génomiques estimées augmente de 30% (0,58 vs 0,45) si la taille de la population de référence est augmentée de 5000 à 10000 individus. On constate sur la figure 4 que l'effet marginal d'une augmentation de N_p diminue à mesure que N_p augmente. Cependant, une augmentation du nombre d'individus phénotypés et génotypés continue d'améliorer la précision dans la plupart des situations, y compris pour de grandes

valeurs de N_p ($>20\,000$ individus). Le gain de précision est d'autant plus grand que l'héritabilité du caractère considéré est faible et que l'effectif génétique de la population est élevé.

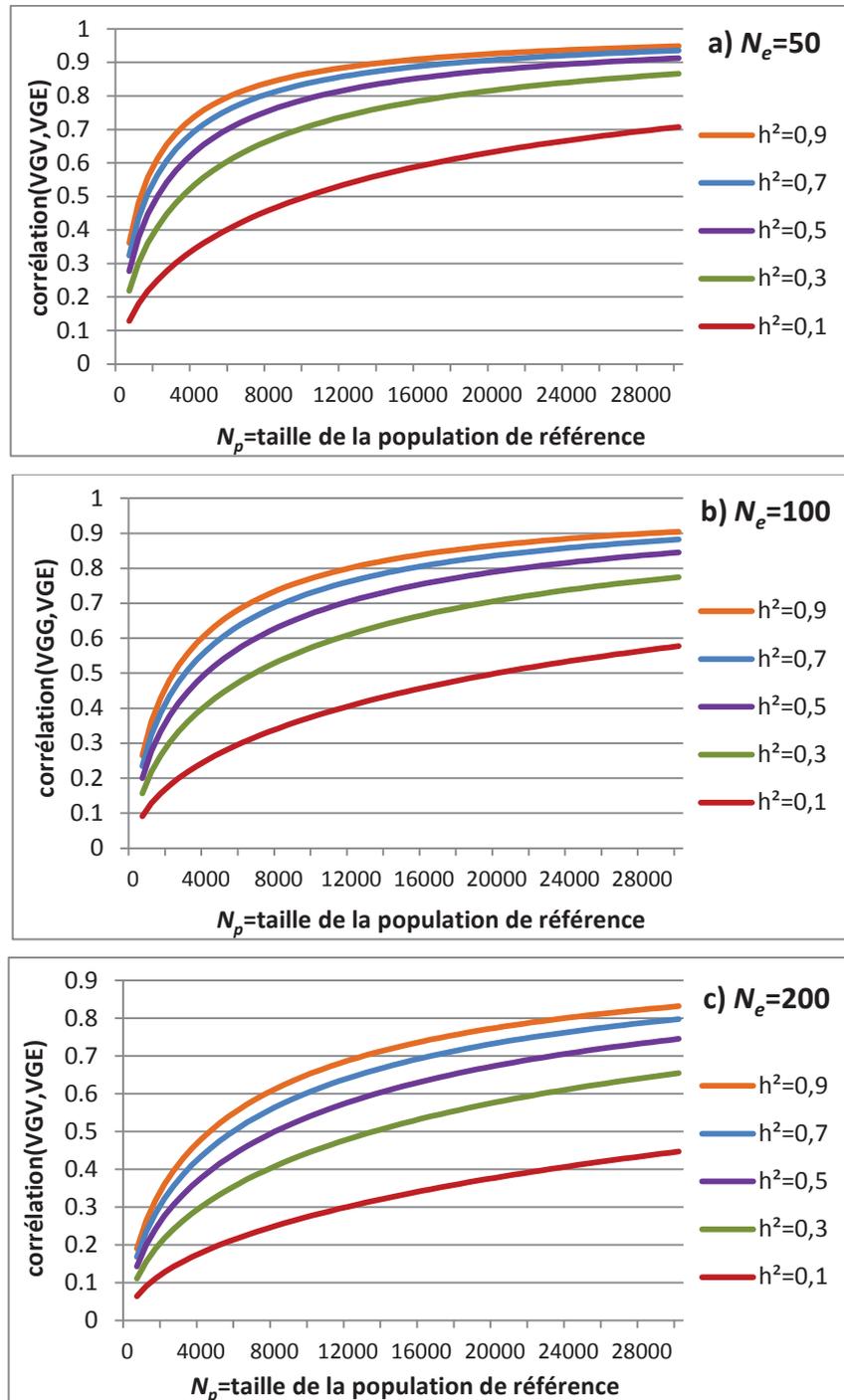


Figure 4 : Précision des valeurs génomiques estimées en fonction de la taille de la population de référence (N_p) et de l'héritabilité du caractère considéré (h^2), pour une population de taille efficace $N_e=50$ (figure 4a), 100 (figure 4b) ou 200 (figure 4c). D'après Daetwyler et al. (2008).

Plusieurs études s'appuyant sur des simulations stochastiques ont effectivement montré un effet positif de la taille de la population de référence sur la précision des évaluations génomiques. Ainsi, Brito et al. (2011) ont simulé une population présentant un déséquilibre de liaison comparable à celui observé dans une population bovine allaitante et ont estimé la précision des valeurs génomiques par la méthode du GBLUP en s'appuyant sur des populations de référence constituées d'un nombre croissant de taureaux évalués sur descendance (en moyenne, 73 produits par père). Comme présenté sur la figure 5, une augmentation de la taille de la population de référence de 480 à 960 ou 1 920 individus se traduit par un accroissement moyen de 28% et 59% de la précision des évaluations, respectivement.

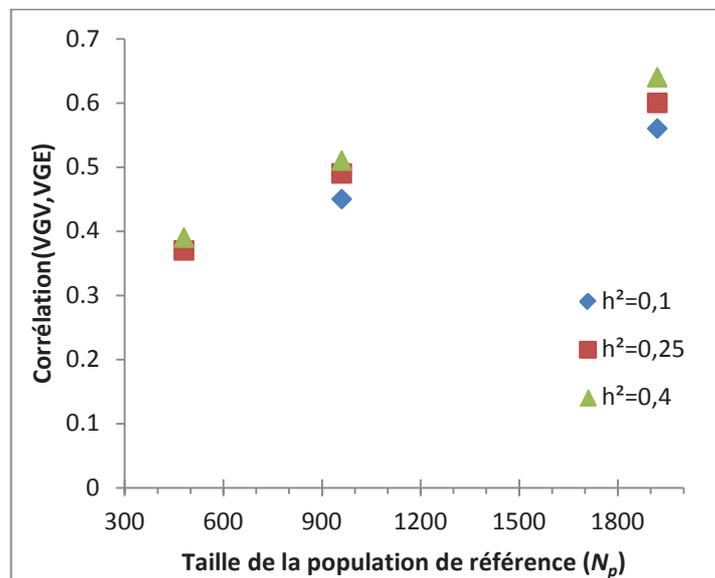


Figure 5 : Précision des valeurs génomiques estimées par GBLUP dans une population bovine, en s'appuyant sur des populations de référence de tailles (N_p) variables constituées de taureaux évalués sur descendance (73 produits en moyenne par père), pour un caractère d'héritabilité 0,1 ou 0,25 ou 0,4. D'après Brito et al. (2011).

Le bénéfice d'un accroissement de la population de référence a également pu être observé sur données réelles dans le cadre des évaluations génomiques européennes de la population bovine Holstein. Dans le cadre du projet EuroGenomics (Lund et al., 2010), les taureaux génotypés et testés sur descendance dans chacun des programmes de sélection français, allemand, scandinave et hollandais (environ 4 000 mâles par programme) ont été regroupés afin de constituer une population de référence commune de plus grande taille. En moyenne, sur l'ensemble des caractères évalués dans chaque programme de sélection, la précision des

index génomiques estimés en s'appuyant sur la population de référence EuroGenomics était 10% supérieure à celle des index estimés à l'aide de la population de référence nationale (voir figure 6).

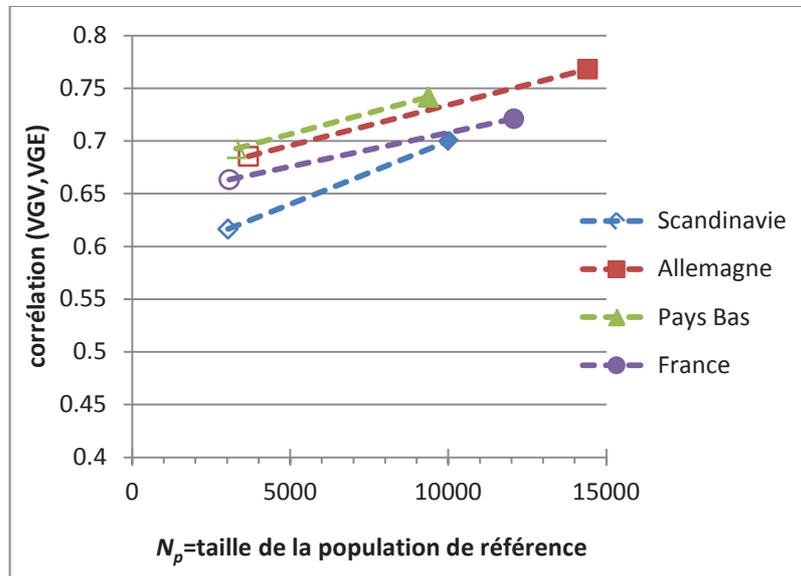


Figure 6 : Précision des valeurs génomiques estimées (moyenne sur l'ensemble des caractères évalués) en population bovine Holstein dans les programmes de sélection scandinave, allemand, hollandais et français en s'appuyant sur une population de référence nationale de petite taille (marques vides) ou sur la population internationale EuroGenomics (marques pleines). D'après Lund et al. (2010).

2.3. Effet de l'héritabilité du caractère considéré

Les formules déterministes de Daetwyler et al. (2008) et de Goddard (2009) indiquent que la précision des valeurs génomiques estimées augmente avec l'héritabilité, à taille de population de référence et taille efficace fixées. On constate par exemple sur la figure 4b que, pour $N_e=100$ et $N_p=10\,000$, la précision des valeurs génomiques estimées se limiterait à 0,38 pour un caractère d'héritabilité égale à 0,10, mais s'élèverait à 0,73 pour un caractère d'héritabilité égale à 0,70.

Cet effet positif de l'héritabilité sur la précision a été observé par de nombreux auteurs à partir de données simulées de manière stochastique, par exemple Nielsen et al. (2011) et de Roos et al. (2009), comme illustré sur la figure 7.

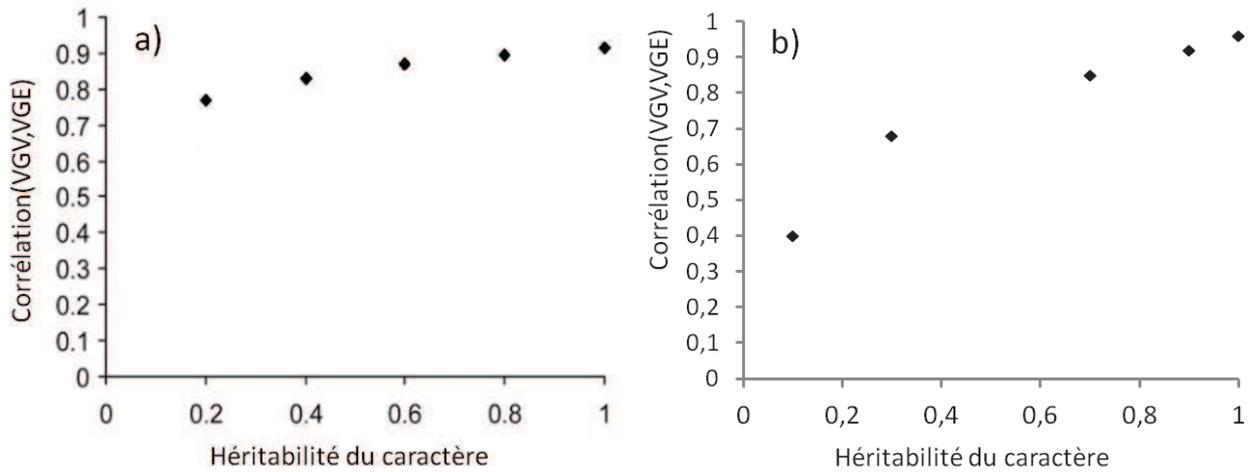


Figure 7 : Précision des valeurs génomiques estimées en fonction de l'héritabilité du caractère considéré. (a) d'après Nielsen et al. (2011) : population de référence constituée de 4 000 individus germains des 1 000 candidats évalués ; (b) d'après de Roos et al. (2009) : population de référence constituée de 1 000 individus des 2 générations précédant la génération des 1 000 individus évalués.

Cette relation peut s'expliquer par le fait que la corrélation entre le phénotype d'un individu et sa valeur génétique vraie est égale à $\sqrt{h^2}$. Autrement dit, l'informativité du phénotype d'un individu de la population de référence pour prédire l'effet génétique de chaque segment chromosomique (dont la somme est égale à la valeur génétique de l'individu) est positivement liée à l'héritabilité du caractère considéré. En conséquence, le nombre d'observations nécessaires pour estimer l'effet des segments chromosomiques à un niveau de précision voulu sera plus important pour un caractère peu héritable que pour un caractère très héritable.

Compte tenu de cette relation, il est avantageux, lorsque c'est possible, de constituer une population de référence composée d'individus génotypés et évalués sur descendance, en considérant pour leur « phénotype » la performance moyenne de leurs produits. En effet, la corrélation entre la moyenne des performances de n produits d'un individu et sa valeur génétique vraie est égale à $\sqrt{\frac{nh^2}{4+(n-1)h^2}}$ (Minvielle, 1990), qui est généralement supérieur à $\sqrt{h^2}$ même pour de faibles valeurs de n . Tout se passe alors comme si l'héritabilité du caractère était égale au carré de la précision de la valeur de l'animal évalué sur descendance. Ainsi, d'après la formule de Daetwyler et al. (2008) appliquée à une population telle que $N_e=100$, prédire des valeurs génomiques avec une précision de 0,5 pour un caractère d'héritabilité 0,1 requiert une population de référence d'environ 20 000 individus disposant de leur seul phénotype propre, alors que le même niveau de précision pourrait être atteint avec

une population de référence de seulement 3 500 individus ayant chacun 50 produits phénotypés. Les espèces animales dont les programmes de sélection sont organisés autour d'un testage sur descendance (ruminants laitiers, ruminants allaitants pour certains caractères) sont donc dans une position favorable pour mettre en place des évaluations génomiques efficaces. A titre d'exemple, on peut citer le cas des caractères de fertilité chez les bovins laitiers Holstein, très peu héritables ($h^2 < 0.1$) mais pour lesquels les taureaux constituant la population de référence utilisée pour l'évaluation génomique disposent d'un phénotype sur descendance précis ($CD \sim 0,60$), permettant l'estimation de valeurs génomiques pour les jeunes candidats avec une précision de l'ordre de 0,70.

Chez les espèces végétales pour lesquelles il est possible d'obtenir des lignées pures ou bien des clones, un moyen similaire d'augmenter le pouvoir prédictif d'une population de référence de taille donnée consisterait à répliquer les individus de la population de référence de manière à disposer pour chacun d'une moyenne de performances plus précise qu'une performance unique. De plus, en phénotypant un même individu dans différents milieux, on dispose ainsi d'une estimation plus stable de sa valeur, moins soumise à l'environnement, ce qui permet d'augmenter l'héritabilité des performances moyenne (Gallais, 1990), donc le pouvoir prédictif de la population de référence.

2.4. Effet de la structure de la population de référence et de la distance génétique avec la population évaluée

La précision des valeurs génomiques estimées prédite à l'aide de la formule déterministe de Daetwyler et al. (2008) ou de celle de Goddard (2009) dépend uniquement, pour une héritabilité, une taille efficace et une longueur de génome fixées, de la taille de la population de référence utilisée pour établir les équations de prédiction de la valeur génomique. Autrement dit, on s'attendrait, d'après ces formules, à ce que la précision des valeurs génomiques estimées soit identique pour tous les candidats évalués à partir d'une même population de référence. Plusieurs auteurs ont cependant mis en évidence que la composition de la population de référence impacte fortement la précision des évaluations génomiques.

Ainsi, Pszczola et al. (2012) ont montré que, à taille de population de référence fixée, la précision des valeurs génomiques estimées augmente lorsque l'apparentement entre les individus constituant la population de référence diminue. En effet, minimiser la parenté au sein de la population de référence permet de maximiser le nombre d'haplotypes de la

population qui y sont représentés, d'équilibrer la représentation de ces haplotypes, et donc d'augmenter la précision d'estimation des valeurs génomiques pour l'ensemble des candidats.

Par ailleurs, de nombreuses études sur données simulées (Nielsen et al., 2011 ; Clark et al., 2012 ; Pszczola et al., 2012), confirmées par des observations sur données réelles (Habier et al., 2010 ; Clark et al., 2012 ; Cleveland et al., 2012) ont montré que la précision de la valeur génomique estimée d'un candidat est positivement liée à son apparentement à la population de référence. En fait, le déséquilibre de liaison entre marqueurs et mutations causales utilisé pour estimer la valeur génomique d'un candidat peut être scindé en deux parties : un DL (plutôt sur de courtes distances) valable pour l'ensemble de la population, et un DL (concernant des segments chromosomiques plus longs transmis par un ascendant proche) qui n'existe qu'intra famille. On peut donc concevoir la précision de la valeur génomique d'un candidat comme la somme d'une précision de base résultant de l'exploitation du DL populationnel et dépendant de N_p , N_e et h^2 comme décrit dans les formules de Daetwyler et al. (2008) et de Goddard (2009) à laquelle vient s'ajouter une précision complémentaire tirant profit d'un DL familial et des relations de parenté (Habier et al., 2013). Ce supplément de précision sera d'autant plus grand que le candidat aura des apparentés proches dans la population de référence (Habier et al., 2007a ; Habier et al., 2010 ; Clark et al., 2012 ; Pszczola et al., 2012).

Il ressort de ces différents résultats qu'une population de référence devrait idéalement être constituée d'individus représentant au mieux la variabilité de la population évaluée, donc peu apparentés entre eux, et contenir des apparentés proches de chaque candidat (en particulier leurs parents) afin de maximiser la part de précision s'appuyant sur le DL familial. Ces recommandations empiriques sont confirmées par les résultats de Rincent et al. (2012), qui ont comparé dans deux populations de maïs la capacité prédictive de populations de référence constituées d'individus échantillonnés selon différents critères, et étudié la structure génétique de la meilleure population de référence de leur étude (figure 8).

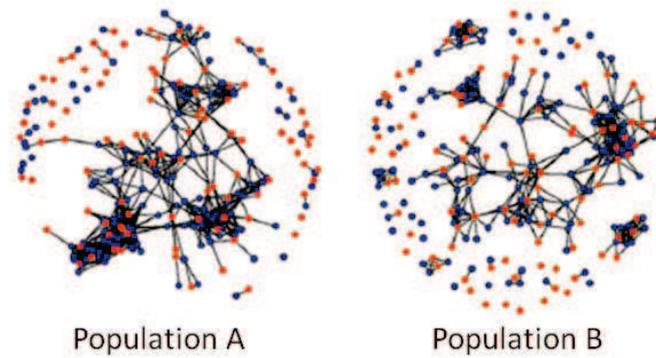


Figure 8 : Représentation en réseau des coefficients de parenté génomique des individus de deux populations de maïs, d'après Rincent et al. (2012). Les points rouges représentent les individus constituant la population de référence optimale, les points bleus représentent les candidats évalués. Deux individus dont le coefficient de parenté génomique est supérieur à 0,2 sont reliés. Les individus de la population de référence optimale sont répartis sur l'ensemble du réseau, et sont peu apparentés au sein des clusters.

Les associations entre marqueurs et mutations causales tendent naturellement à disparaître au cours des générations, car rompues par les recombinaisons génétiques. En conséquence, si la population de référence n'est pas régulièrement complétée ou renouvelée avec de nouveaux individus récents, la précision de l'évaluation génomique va progressivement diminuer, comme montré par exemple par Bastiaansen et al. (2012) et Solberg et al. (2009) par simulations (figure 9), ou par Wolc et al. (2011) sur données réelles dans un programme de sélection de poules pondeuses. Le DL populationnel et le DL familial tendent tous deux à disparaître au cours des générations. Cependant, les segments chromosomiques partagés par les individus fortement apparentés sont plus longs que les segments conservés au niveau de la population, et ont donc une forte probabilité d'être rompus rapidement. En pratique, ceci se traduit par une perte importante de précision au cours des premières générations qui reflète la disparition rapide du DL familial, suivie d'une diminution de précision plus lente et plus régulière correspondant à l'érosion continue du DL populationnel, comme observé sur la figure 9. Afin de conserver son pouvoir prédictif et donc maintenir l'efficacité de la sélection, une population de référence devrait donc continuellement intégrer des individus récents, ce qui implique de ré-estimer régulièrement les effets des segments chromosomiques. Un moyen de s'affranchir du renouvellement continu de la population de référence consisterait à disposer de marqueurs en très fort DL avec la plupart des mutations causales déterminant les caractères d'intérêt.

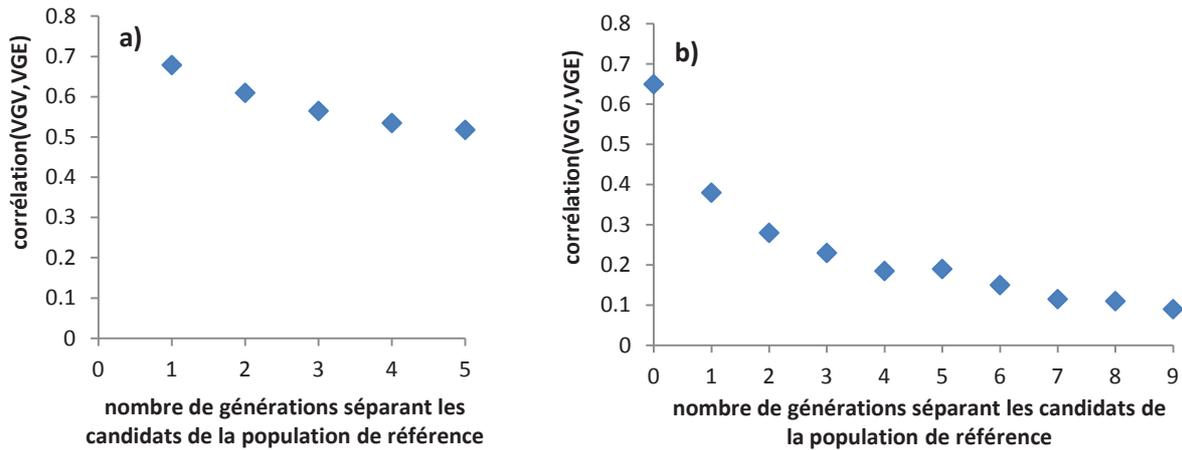


Figure 9 : Evolution de la précision des valeurs génomiques estimées en fonction du nombre de générations séparant les candidats évalués de la population de référence. (a) d'après Solberg et al. (2009) : population de référence=1 000 individus, $h^2=0,5$, 100 SNP/M ; (b) (Bastiaansen et al., 2012) : population de référence=500 individus, $h^2=0,25$, 350 SNP/M.

Les parts relatives de la précision provenant de ces deux sources d'information varient selon la méthode d'évaluation génomique considérée (Habier et al., 2007a ; 2010). Certaines, comme le GBLUP tirent plus profit des relations de parenté entre individus, alors que d'autres comme les méthodes bayésiennes, s'appuient plus largement sur le DL populationnel. On peut donc penser que ces dernières pourraient préserver plus longtemps la capacité prédictive d'une population de référence non renouvelée, comme rapporté par Habier et al. (2007a) à partir de données simulées. Cependant, Wolc et al. (2011) n'ont pas observé de différence significative dans l'évolution de la précision des valeurs génomiques estimées par GBLUP, BayesA ou BayesC π dans une population de poules pondeuses au cours des cinq générations suivant la constitution de la population de référence, ce qui suggère que les méthodes bayésiennes sont en pratique elles aussi sensibles au niveau d'apparentement avec la population de référence.

2.5. Effet de la densité de marqueurs

Plusieurs auteurs ont mis en évidence une relation positive entre la précision des valeurs génomiques estimées et la densité des marqueurs pour lesquels les individus de la population de référence et les candidats évalués sont génotypés. Par exemple, Calus et al. (2008) ont comparé par simulation la précision des valeurs génomiques estimées en s'appuyant sur une population de référence de 1 100 individus pour plusieurs densités de marqueurs variant de 40 SNP/M à 780 SNP/M. Comme illustré sur la figure 10, une

augmentation de la densité des marqueurs s'est traduite par une amélioration significative de la précision des valeurs génomiques estimées.

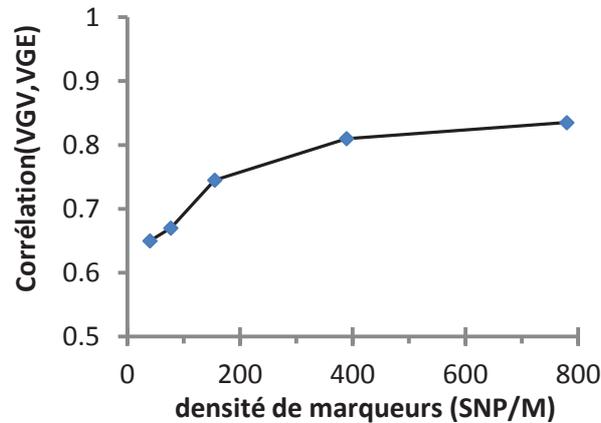


Figure 10 : Précision des valeurs génomiques estimées en fonction de la densité de marqueurs SNP. D'après Calus et al. (2008) ; $N_p=1\ 100$ individus, $N_e=100$, $h^2=0,5$.

VanRaden et al. (2009) ont également observé un gain de précision dans l'évaluation génomique de taureaux Holstein nord-américains lorsque le nombre de SNP pour lequel les animaux étaient génotypés augmentait de 10 000 à 20 000 marqueurs, puis de 20 000 à 40 000 marqueurs.

En fait, l'évaluation génomique, pour être précise, nécessite que le DL entre les marqueurs et les QTL soit suffisamment fort, afin que les effets estimés des segments chromosomiques déterminés par les marqueurs reflètent bien les effets des mutations causales. Etant donné qu'en moyenne le DL entre deux loci augmente lorsque la distance les séparant diminue, accroître le nombre de marqueurs sur le génome permet d'augmenter la probabilité que les QTL soient en fort déséquilibre de liaison avec les marqueurs les entourant.

Les formules déterministes de Daetwyler et al. (2008) et de Goddard (2009) font l'hypothèse que les marqueurs pour lesquels les individus sont génotypés expliquent la totalité de la variance génétique des caractères considérés, autrement dit que le déséquilibre de liaison entre marqueurs et mutations causales est total, ce qui en pratique n'est pas le cas. Un paramètre d'ajustement q a donc été proposé par Daetwyler (cité par Calus et al., 2013), qui correspond à la racine carrée de la variance génétique expliquée par les marqueurs :

$$Acc = q \sqrt{\frac{N_p h^2}{N_p h^2 + M_e}}$$

Par exemple, dans le cas des populations bovines laitières génotypées avec une puce de 54K SNP, ces auteurs ont estimé que $q = \sqrt{0.8}$. On peut toutefois penser que la part de variance expliquée par les marqueurs, et par conséquent la valeur de ce terme correctif, varie selon les caractères, en fonction de la complexité de leur déterminisme génétique.

Le nombre et la position des mutations causales étant inconnus, le niveau de DL attendu entre mutations causales et marqueurs peut être évalué au travers du DL entre marqueurs adjacents, lui-même dépendant du nombre de marqueurs pour lequel les individus sont génotypés et de l'étendue du DL dans la population considérée. La figure 11 présente le DL (r^2) moyen entre marqueurs dans la population bovine Holstein européenne et dans les 4 principales populations porcines françaises sélectionnées. On constate que le DL est plus élevé dans les populations porcines qu'en Holstein pour une distance entre loci donnée. La densité de marqueurs permettant de disposer d'un niveau moyen donné de DL entre marqueurs adjacents serait donc *a priori* plus faible chez le porc qu'en Holstein. Par exemple, environ 100 000 SNP seraient nécessaires dans la population bovine pour un r^2 moyen de 0,25, alors qu'environ 20 000 SNP suffiraient chez le porc. A l'inverse, à une distance de 60 kb, qui correspond approximativement à la distance moyenne approximative entre marqueurs adjacents sur la puce bovine 54K et sur la puce porcine 60K, le r^2 est de l'ordre de 0,2 en Holstein et de 0,3 chez le porc.

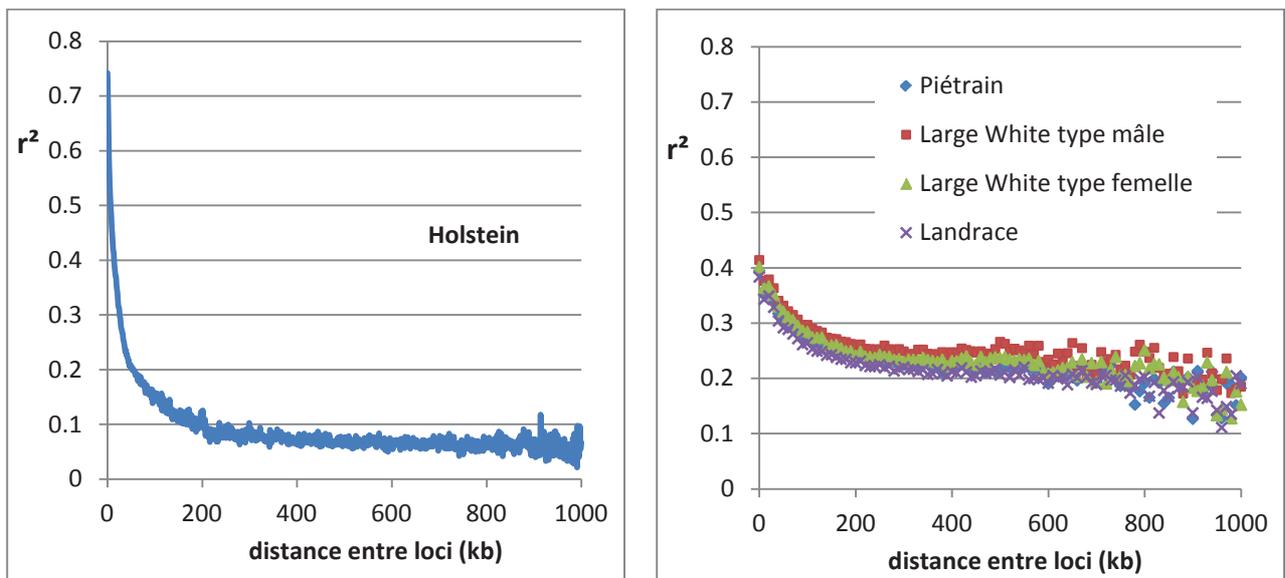


Figure 11 : DL moyen (r^2) en fonction de la distance entre loci dans la population bovine laitière Holstein européenne (Hoze, communication personnelle) et dans les 4 populations porcines collectives françaises (Boitard, communication personnelle).

Un accroissement de la densité de marqueurs pourrait également permettre de préserver plus longtemps la capacité prédictive d'une population de référence. Solberg et al. (2009) ont en effet observé par simulations qu'augmenter la densité de marqueurs avait pour conséquence de ralentir la chute de précision des évaluations génomiques lorsque le nombre de générations séparant les candidats évalués de la population de référence augmentait (figure 12). Ce phénomène peut s'expliquer par la plus faible probabilité qu'une recombinaison génétique se produise entre une mutation causale et des marqueurs plus proches.

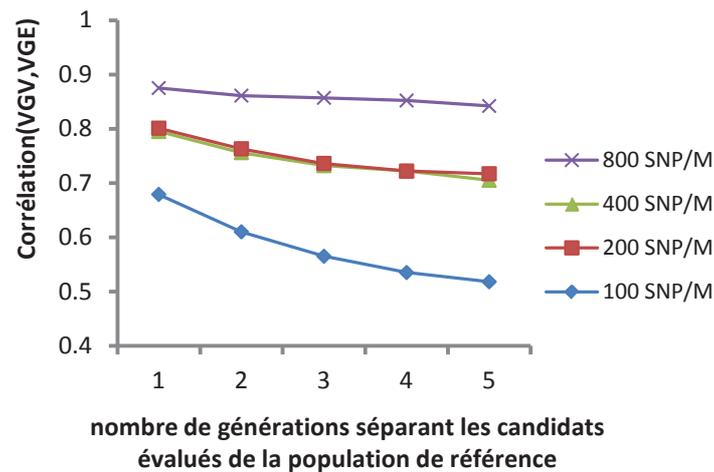


Figure 12 : Evolution de la précision des valeurs génomiques estimées en fonction du nombre de générations séparant les candidats évalués de la population de référence, pour 4 densités de marqueurs SNP ; d'après Solberg et al. (2009). $N_e=100$, $N_p=1\ 000$, $h^2=0,5$.

On pourrait donc *a priori* envisager de génotyper les individus pour des puces extrêmement denses, afin de maximiser le DL entre marqueurs et mutations causales et la précision des valeurs génomiques estimées. Cependant, une comparaison des valeurs génomiques estimées dans les populations Holstein française (Boichard, communication personnelle), américaine (VanRaden et al., 2013) et scandinave (Su et al., 2013) en s'appuyant sur des puces de 54 000 SNP ou de 777 000 SNP a montré un gain de précision très faible, inférieur à 1% en moyenne sur les caractères considérés, et légèrement plus élevé avec des méthodes bayésiennes qu'avec un GBLUP selon VanRaden et al. (2013) et Su et al. (2013). Dans le cas du GBLUP, ces résultats peuvent s'expliquer par le fait que l'estimation de la parenté entre individus sur laquelle s'appuie la méthode est déjà très bonne avec 54 000 marqueurs et n'est que très peu améliorée par un nombre plus élevé de marqueurs. Les méthodes bayésiennes tirent plus profit du DL entre marqueurs et QTL, ce qui explique le gain de précision plus élevé que pour le GBLUP. Cependant, on peut penser que le plus grand

nombre d'effets à estimer sous les modèles bayésiens utilisés par ces auteurs avec une plus forte densité de marqueurs entraîne une perte de puissance, pouvant expliquer l'augmentation modérée de précision.

2.6. Effet de la nature des segments chromosomiques considérés

La nature des segments chromosomiques élémentaires pour lesquels on estime un effet à l'aide de la population de référence peut également avoir un impact sur la précision des valeurs génomiques estimées. En effet, l'évaluation génomique repose sur le principe que les marqueurs permettent de suivre la transmission des segments chromosomiques ancestraux dans la population, comme illustré sur la figure 1. On suppose donc que deux individus ayant le même génotype à un endroit donné du génome ont reçu un même segment chromosomique ancestral, c'est-à-dire qu'ils sont porteurs d'un segment IBD (« identical by descent ») dans cette région.

Pour des raisons de simplicité de mise en œuvre, la plupart des applications d'évaluations génomiques actuelles considèrent des segments chromosomiques élémentaires chacun déterminé par un seul marqueur. Cependant, on peut penser que considérer des haplotypes de plusieurs marqueurs contigus permet un meilleur suivi de la transmission des segments IBD dans la population, car la probabilité qu'un même haplotype de plusieurs marqueurs proches soit porté par plusieurs individus sans être un segment IBD est faible. Considérer des haplotypes peut donc permettre d'augmenter le DL entre les segments définis par les marqueurs et les QTL.

Calus et al. (2008) ont comparé par simulations stochastiques la précision des valeurs génomiques estimées à l'aide d'une même population de référence, en considérant des marqueurs SNP individuels ou des haplotypes de 2 ou 10 marqueurs SNP contigus, pour 5 densités de marqueurs (exprimées en r^2 entre marqueurs adjacents) et 2 valeurs d'héritabilité (0,10 ou 0,50). Leurs résultats (figure 13) montrent que l'utilisation d'haplotypes permet d'améliorer la précision de l'évaluation génomique par rapport à une situation s'appuyant sur des marqueurs individuels, en particulier lorsque le r^2 entre marqueurs adjacents est faible (inférieur à 0,19 dans leur étude). Cependant, pour de plus fortes densités de marqueurs, cet avantage se réduit, voire disparaît.

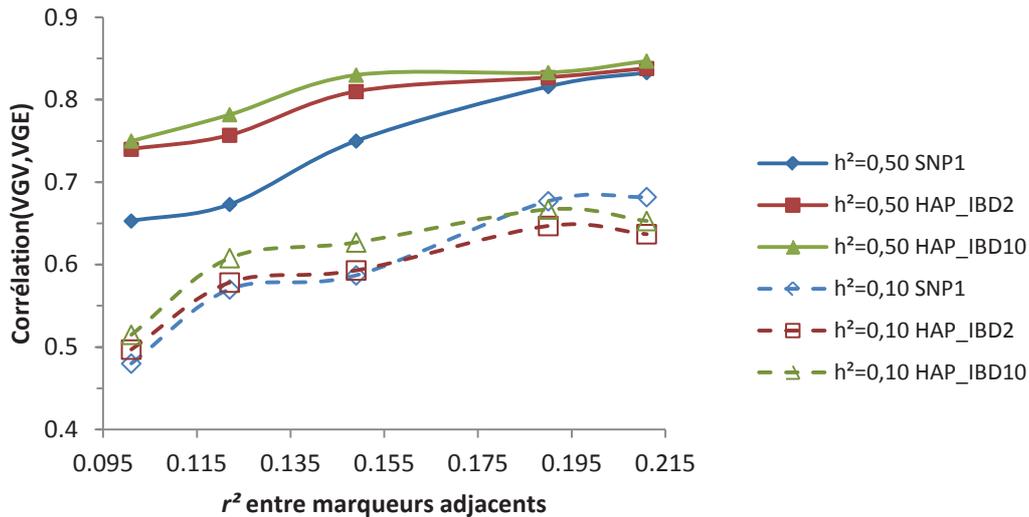


Figure 13 : Précision des valeurs génomiques estimées en fonction du r^2 entre marqueurs adjacents, pour 3 modèles d'évaluation génomique : SNP1 = estimation d'effets de SNP individuels ; HAP_IBD2 = estimation d'effets d'haplotypes de 2 marqueurs adjacents ; HAP_IBD10 = estimation d'effets d'haplotypes de 10 marqueurs adjacents. D'après Calus et al. (2008). $N_p=1\ 100$ individus, $N_e=100$, $h^2=0,5$ ou $0,1$.

Utiliser des haplotypes permet de disposer d'un DL plus fort entre les segments chromosomiques élémentaires et les mutations causales, ce qui est avantageux en termes de précision d'estimation des effets des segments chromosomiques élémentaires. Le gain de précision par rapport à un modèle s'appuyant sur des marqueurs individuels diminue cependant lorsque la densité de marqueurs augmente. Par ailleurs, le nombre d'effets d'haplotypes à estimer à partir de la population de référence est supérieur au nombre d'effets de marqueurs individuels, en particulier si les haplotypes sont définis sur un grand nombre de marqueurs, ce qui tend à réduire la précision d'estimation de l'effet des haplotypes. On peut donc penser que pour les plus fortes densités de SNP, l'avantage d'un DL plus fort est compensé par le handicap du nombre d'effets à estimer, rendant au final le modèle s'appuyant sur des marqueurs individuels aussi efficace que celui s'appuyant sur des haplotypes. Cependant, une optimisation de la longueur des haplotypes et un regroupement des haplotypes ayant des probabilités IBD élevées pourrait permettre de limiter le nombre d'effets à estimer tout en préservant un avantage en termes de DL.

2.7. Effet de la méthode d'estimation des effets des segments chromosomiques

Les différentes méthodes d'évaluation génomique reposent sur des hypothèses diverses quant au déterminisme génétique des caractères évalués. On peut donc s'attendre à

des écarts de performances entre elles, selon que ces hypothèses sont proches ou non du déterminisme réel des caractères.

Meuwissen et al. (2001) ont comparé par simulation la précision des valeurs génomiques estimées avec 4 méthodes à l'aide d'une population de référence de 2 000 individus : méthode des moindres carrés, GBLUP, BayesA et BayesB. Pour cela, ils ont simulé un génome de 10M portant des QTL dont l'effet était tiré dans une loi Gamma, générant ainsi quelques QTL ayant un effet très fort, un nombre limité de QTL ayant un effet moyen, et un grand nombre de QTL ayant un effet très faible. Dans ces conditions, les auteurs ont mis en évidence une nette infériorité de la méthode des moindres carrés par rapport aux trois autres méthodes (-60% de précision en moyenne), en raison d'une surestimation des effets des segments chromosomiques estimés. Parmi les trois autres méthodes, le GBLUP s'est révélé 8% à 14% moins précis que les méthodes bayésiennes, la précision la plus élevée des valeurs génomiques estimées ayant été obtenue avec la méthode BayesB.

Daetwyler et al. (2010) ont comparé par simulations la précision des valeurs génomiques estimées par GBLUP et BayesB sur un génome de 10M. Les auteurs ont fait varier la taille efficace de la population ($N_e=200, 1\ 000$ ou $2\ 000$), l'héritabilité du caractère considéré ($h^2=0,1$ ou $0,3$ ou $0,5$), la taille de la population de référence ($N_p=200, 500, 1\ 000$ ou $2\ 000$), le nombre de QTL (de environ $0,05 * N_e$ à environ $1,8 * N_e$) et la distribution dans laquelle l'effet des QTL était échantillonné (gaussienne ou double exponentielle). Leurs résultats, présentés sur la figure 14 dans le cas d'une distribution normale de l'effet des QTL et pour $N_e = N_p = 1\ 000$, montrent que la précision du GBLUP est très peu sensible au nombre de QTL déterminant le phénotype. A l'inverse, la méthode bayésienne, 50% à 60% plus précise que le GBLUP lorsque le nombre de QTL est très faible, voit son efficacité diminuer lorsque celui-ci augmente, devenant même moins précise que le GBLUP lorsque le nombre de QTL est grand. De plus, la précision du GBLUP serait très peu dépendante de la forme de la distribution des effets des QTL, alors que la méthode bayésienne serait un peu plus efficace dans le cas d'une distribution à « queue épaisse » que dans celui d'une distribution normale des effets des QTL.

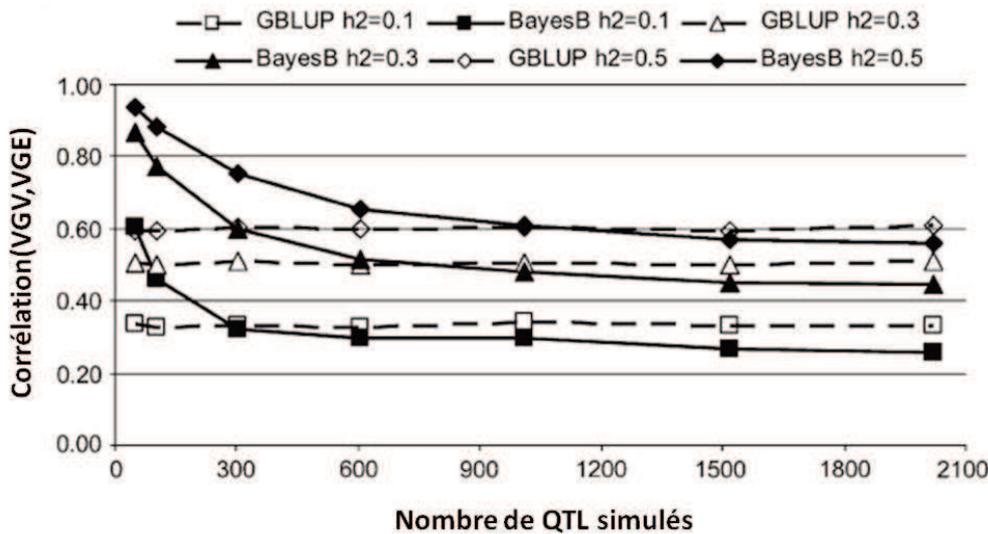


Figure 14 : Précision des valeurs génomiques estimées par les méthodes GBLUP et BayesB en fonction du nombre de QTL simulés, pour $N_e=1\ 000$ et $N_p=1\ 000$. D’après Daetwyler et al. (2010).

Coster et al. (2010) ont comparé par simulation la précision des valeurs génomiques estimées selon 3 méthodes en fonction du nombre de QTL et de la distribution de leurs effets : une méthode bayésienne proche du BayesB (BM), une méthode de régression pénalisée proche du LASSO (LARS), et la méthode de régression Partial Least Squares (PLS). Ces auteurs ont également observé des différences d’efficacité entre méthodes, et montré que le déterminisme génétique du caractère avait un impact sur la précision de la méthode bayésienne et du LARS alors que la méthode PLS semblait robuste à ces hypothèses.

Malgré les écarts de précision parfois importants entre méthodes observés par simulations, les résultats obtenus sur données réelles révèlent en fait dans la plupart des cas des performances assez semblables pour les méthodes les plus couramment utilisées. Ainsi, Ostersen et al. (2011) ont réalisé des évaluations génomiques pour un caractère de vitesse de croissance et un caractère d’efficacité alimentaire dans la population porcine Duroc danoise, à l’aide d’une population de référence de 1 735 verrats génotypés pour la puce porcine 60K. Les précisions moyennes des valeurs génomiques estimées avec les trois méthodes comparées – GBLUP, LASSO bayésien et une méthode bayésienne proche du BayesB – se sont révélées très proches (0,57 à 0,58 pour la vitesse de croissance, et 0,44 à 0,45 pour l’efficacité alimentaire). Hayes et al. (2009a) et VanRaden et al. (2009) ont également rapporté dans les populations Holstein nord-américaine, australienne et néo-zélandaises des écarts de précision très faibles pour la plupart des caractères évalués entre les valeurs génomiques estimées par GBLUP ou à l’aide de méthodes bayésiennes. Pour quelques caractères soumis à l’effet d’un

gène majeur, comme le taux butyreux ou le taux protéique du lait pour lesquels le gène *DGATI* explique environ 20 à 30% de la variabilité génétique, les méthodes bayésiennes autorisant un écart à la normalité des effets des marqueurs se sont toutefois révélés 7 à 10% plus précises que le GBLUP (Moser et al., 2009 ; VanRaden et al., 2009). Peu de gènes majeurs ont cependant été mis en évidence à ce jour, et la plupart des caractères semblent être déterminés par un grand nombre de gènes. On peut donc penser qu'en règle générale, le choix de la méthode d'évaluation génomique parmi les plus fréquemment employées – GBLUP, méthodes bayésiennes, Elastic Net, PLS – devrait être assez neutre en terme de précision des valeurs génomiques estimées.

3. METHODE RETENUE POUR EVALUER L'INTERET DE METTRE EN PLACE DES EVALUATIONS GENOMIQUES DANS UN PROGRAMME DE SELECTION PORCIN

Dans le cadre de cette thèse, nous avons choisi d'évaluer l'intérêt de mettre en place des évaluations génomiques chez le porc en nous appuyant sur des simulations. Ceci nécessite de modéliser le fonctionnement de programmes de sélection porcins organisés autour d'évaluations génétiques de type BLUP-modèle animal (**BLUP-MA**) ou d'évaluations génomiques, afin de comparer l'efficacité de ces différents modèles.

Trois types de modèles sont envisageables : les modèles déterministes, les modèles stochastiques, et les modèles pseudo-stochastiques.

3.1. Modèles déterministes

Un modèle déterministe est un modèle dans lequel les critères d'efficacité que l'on souhaite évaluer sont complètement décrits et calculés à l'aide d'équations mathématiques les reliant aux variables élémentaires caractérisant la population considérée d'un point de vue démographique (taille de la population, sex-ratio, ...), génétique (paramètres génétiques des caractères considérés, ...), zootechnique (fraction d'animaux phénotypés, ...) ou économique (valeur économique des caractères, coût du phénotypage, ...). Les modèles déterministes s'appuient sur la théorie classique des indices de sélection et reposent sur les hypothèses du modèle génétique infinitésimal.

Cette approche a été utilisée par plusieurs auteurs pour simuler des programmes de sélection animaux fondés sur des évaluations génomiques et comparer leur efficacité à celle de programmes traditionnels, par exemple König et al. (2009) et Pryce et al. (2010) en bovins laitiers, Pimentel et König (2012) en bovins allaitants, ou Dekkers (2007b) chez le porc (il s'agissait en fait dans ce dernier cas de sélection assistée par marqueurs et non pas de sélection génomique).

Ces modèles ont l'avantage d'être peu coûteux en ressources informatiques, et l'obtention des résultats pour un jeu de paramètres d'entrée est très rapide. Ce type de modèle est donc bien adapté à l'optimisation, puisqu'il est possible d'obtenir la réponse pour les critères d'efficacité étudiés sur l'ensemble de la plage de variation des paramètres d'entrée, et de déterminer le jeu de paramètres permettant la meilleure efficacité du programme de sélection étudié.

Cependant, dans ces modèles, la précision des évaluations génomiques, paramètre clé de la réponse à la sélection, est soit un paramètre d'entrée du modèle fixé *a priori*, soit estimée à l'aide d'une des formules déterministes décrites précédemment à partir des caractéristiques de la population considérée. Or, on a vu que ces formules ne tiennent d'une part pas compte de certains paramètres impactant la précision des valeurs génomiques estimées, et que d'autre part elles peuvent produire des résultats sensiblement différents selon les hypothèses faites sur le nombre de segments chromosomiques indépendants dans la population.

3.2. Modèles stochastiques

Dans le cas d'un modèle stochastique, les critères d'efficacité que l'on souhaite évaluer sont obtenus en calculant des statistiques simples (moyennes, variances) sur des observations générées directement ou indirectement par tirages aléatoires dans des lois statistiques. On peut ainsi par exemple simuler des individus constituant une population présentant une structure familiale réaliste. Chaque individu dispose d'un génome constitué de marqueurs et de gènes déterminant sa valeur génétique vraie, qui, additionnée à des effets environnementaux, détermine son phénotype. Les phénotypes de la population simulée sont ensuite traités comme des données réelles sur lesquelles les différents modèles d'évaluation génétique (BLUP-MA, GBLUP, ...) peuvent être appliqués. Les génotypes aux marqueurs des individus peuvent avoir été générés par simulations, en tentant de reproduire les caractéristiques du génome de la population étudiée (DL, densité de marqueurs, ...) selon les

méthodes décrites dans Daetwyler et al. (2013), ou peuvent être de véritables données de génotypage, reflétant toute la complexité du DL et les fréquences alléliques réellement observées dans la population que l'on souhaite étudier.

Un avantage des modèles stochastiques est que la population simulée peut être gérée selon divers scénarios éventuellement très complexes (gestion optimisée des accouplements, ...), comme le serait une population réelle, ce qui peut difficilement être pris en compte avec un modèle déterministe. Un autre avantage est que la réponse à la sélection est calculée à partir de l'évolution au cours du temps des valeurs génétiques vraies des individus qui sont connues puisque simulées, sans qu'aucune hypothèse *a priori* quant à la précision des valeurs génomiques estimées soit nécessaire. Cette précision dépend directement des caractéristiques de la population de référence et de la population évaluée, qui peuvent évoluer au cours du temps du fait du scénario simulé (par exemple, augmentation de la taille de la population de référence au cours du temps par accumulation d'individus) ou en conséquence du travail de sélection (évolution des fréquences alléliques aux marqueurs ou aux QTL).

L'efficacité de la sélection observée sur des modèles stochastiques reflète donc bien, *a priori*, celle qu'on pourrait espérer obtenir dans une population réelle conduite de la même manière que la population simulée. Cependant, comme dans le cas des modèles déterministes, les résultats dépendent évidemment de la qualité du modèle et des hypothèses faites lors de son élaboration : nombre de QTL et distribution de leurs effets, étendue du déséquilibre de liaison, densité de marqueurs ...

Des simulations stochastiques ont été réalisées par exemple par Jannink (2010) pour étudier l'efficacité à long terme de diverses modalités de sélection génomique sur l'orge, par Lillehammer et al. (2011a ; 2011b) pour comparer l'efficacité de programmes de sélection traditionnels et génomiques chez les bovins laitiers et dans une lignée maternelle porcine, ou encore par Nielsen et al. (2011) pour estimer l'intérêt d'évaluations génomiques dans un programme de sélection aquacole.

L'inconvénient majeur des modèles stochastiques est qu'ils génèrent des calculs très longs puisque chaque cycle de sélection nécessite de réaliser une évaluation génétique comparable à celle d'une véritable population, et qu'ils requièrent des moyens de calculs importants pour gérer les génotypes à des milliers (ou dizaines de milliers) de loci pour un nombre important d'individus. Leur emploi est donc en règle générale limité à l'étude d'un seul caractère, alors que la simulation de plusieurs caractères présentant une structure de corrélation donnée est aisée en ayant recours à un modèle déterministe.

3.3. Modèles pseudo-stochastiques

Une troisième catégorie de modèles tente de combiner l'aptitude des modèles stochastiques à simuler des programmes de sélection réalistes tout en palliant la lenteur de leurs calculs. Dans ces modèles « pseudo-stochastiques », on simule à nouveau les individus d'un programme de sélection comme dans un modèle stochastique. Cependant, les valeurs génomiques estimées ne sont plus obtenues en réalisant une évaluation génomique, mais par tirage aléatoire dans une loi normale en fonction de la précision de l'évaluation fixée à priori ou établie en s'appuyant sur une formule déterministe. Le gain important réalisé sur les temps de calculs se fait donc au prix des mêmes hypothèses fortes sur la précision des valeurs génomiques estimées qu'avec les modèles déterministes (même précision pour tous les individus évalués, non prise en compte de la structure et de la composition de la population de référence, non prise en compte de l'évolution des fréquences alléliques au cours du temps ...). Ce type de modèle a notamment été utilisé par Henryon et al. (2012) chez le porc, ainsi que par de Roos et al. (2010), Wensch-Dorendorf et al. (2011) et Buch et al. (2011) chez les bovins laitiers.

Conclusion

Un modèle, quel qu'il soit, est toujours une représentation plus ou moins simplifiée de la réalité, nécessitant de faire des hypothèses parfois fortes. Les trois types de modèles présentés ci-dessus ont chacun leurs points forts et leurs faiblesses.

Un objectif de cette thèse était d'évaluer le gain (ou la perte) d'efficacité en terme de réponse à la sélection généré par la mise en place d'évaluations génomiques dans un programme de sélection porcin. La précision des valeurs génomiques étant directement liée à la réponse à la sélection, nous avons donc choisi de réaliser des simulations stochastiques, afin d'estimer au mieux la précision des valeurs génomiques dans une population porcine de structure réaliste et de limiter les hypothèses quant à ce paramètre clé de l'efficacité d'un schéma

PARTIE 2.

Article 1.

Perspectives d'application de la selection genomique dans les schemas d'amelioration genetique porcins.

T. Tribout. 2011, INRA Productions Animales 24(4):369-376

Perspectives d'application de la sélection génomique dans les schémas d'amélioration génétique porcins

T. TRIBOUT

INRA, UMR1313 Génétique Animale et Biologie Intégrative, F-78352 Jouy-en-Josas, France
AgroParisTech, Génétique Animale et Biologie Intégrative, 16 rue Claude Bernard, F-75231 Paris, France
Courriel : Thierry.Tribout@jouy.inra.fr

L'évaluation génomique est une nouvelle méthode d'estimation de la valeur génétique des animaux d'élevage, et son implémentation dans les schémas de sélection bovins laitiers a permis d'améliorer fortement leur efficacité. La récente disponibilité d'une puce porcine de 60 000 marqueurs SNP permet aujourd'hui d'envisager une sélection génomique chez le porc. Cet article analyse l'organisation des schémas de sélection porcins actuels, et indique les voies possibles d'amélioration grâce à cette nouvelle méthodologie.

Au cours des 20 dernières années, de nombreux QTL (*Quantitative Trait Loci*, ou locus à effet quantitatif) ont été mis en évidence dans l'ensemble des espèces animales de rente. Des programmes de Sélection Assistée par Marqueurs (SAM) ont été mis en place avec succès dans plusieurs espèces, en France par exemple chez les bovins laitiers (Fritz *et al* 2003) ou encore chez les ovins pour la résistance à la tremblante (Sidani *et al* 2010). Dans l'espèce porcine en revanche, force est de constater que jusqu'à ce jour les applications de SAM ont été très peu nombreuses (Van der Steen *et al* 2005, Luther *et al* 2009, Schwob *et al* 2009), en dépit des centaines de QTL localisés (Hu et Reecy 2007). Cette situation s'explique principalement par le fait que les marqueurs microsatellites (seuls marqueurs disponibles jusqu'à récemment) sont généralement en équilibre de liaison avec les QTL au niveau d'une population (obligeant à raisonner intra-famille ce qui complique les programmes de SAM), et éloignés des mutations causales (d'où une diminution rapide de l'efficacité de la sélection en raison des recombinaisons entre marqueurs et QTL).

Au cours des dernières années, un nouveau type de marqueurs – les marqueurs SNP («*Single Nucleotide Polymorphism*») – répartis sur l'ensemble du génome et en nombre extrêmement élevé (plusieurs centaines de milliers à plusieurs millions de mar-

queurs sur l'ensemble d'un génome) a été mis en évidence. Grâce au développement de puces à SNP, on peut aujourd'hui génotyper pour un faible coût (moins de 200 €) un individu pour plusieurs dizaines (voire centaines) de milliers de ces marqueurs. Parallèlement à ces avancées, des travaux ont montré qu'il était possible, en subdivisant le génome en un très grand nombre de segments définis par un ou plusieurs marqueurs SNP, d'estimer précisément la valeur génétique d'un individu comme la somme des effets de ces segments, l'effet de chaque segment étant au préalable estimé sur un groupe d'individus pour lesquels on dispose à la fois du phénotype et du génotype aux marqueurs.

Cette nouvelle méthodologie d'évaluation, appelée «sélection génomique», est en train de révolutionner les schémas de sélection bovins laitiers depuis quelques années. Elle permet en effet de sélectionner avec une grande précision les jeunes mâles dès leur naissance sans avoir besoin de les soumettre à un testage sur descendance long et coûteux, et d'augmenter les intensités de sélection sur les voies mâles et femelles (Guillaume *et al* 2011).

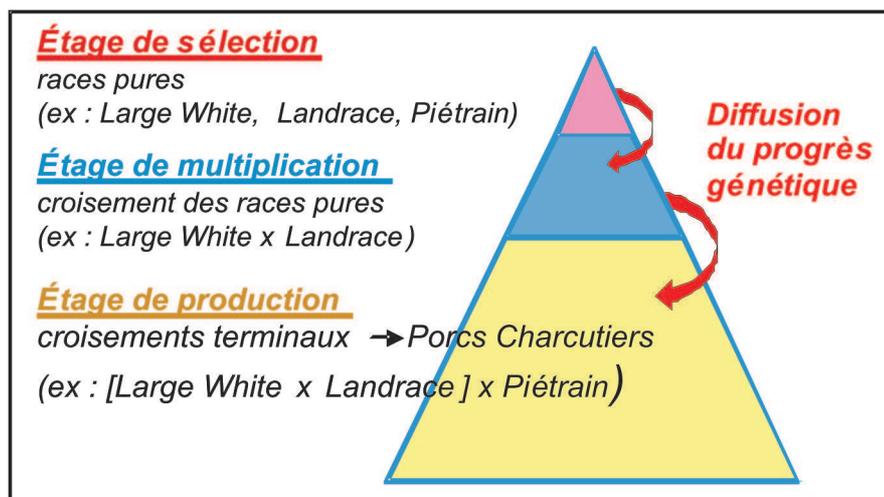
Chez le porc, la disponibilité depuis 2009 d'une puce porcine de 60 000 SNP ouvre de nouvelles perspectives d'utilisation d'information des marqueurs dans les dispositifs d'améliora-

tion génétique. La présente communication a pour objet d'évaluer l'intérêt de la sélection génomique pour l'espèce porcine et d'aborder les difficultés liées à une éventuelle mise en œuvre de cette méthodologie.

1 / Organisation actuelle des schémas d'amélioration génétique porcins ; forces et faiblesses

Chez le porc, les dispositifs d'amélioration génétique ont une structure pyramidale, en général à 3 étages (figure 1). Au sommet de cette pyramide se trouvent les élevages de sélection, dans lesquels on trouve les populations de race pure et les lignées synthétiques. Ces élevages de sélection approvisionnent les élevages des autres niveaux de la pyramide et les centres d'insémination artificielle en reproducteurs de race pure ou de lignées synthétiques. Au niveau intermédiaire se trouvent les élevages de multiplication dans lesquels on produit des cochettes et des verrats croisés à partir d'animaux nés à l'étage de sélection. Ces animaux croisés sont vendus aux élevages de production, qui constituent le dernier niveau de la pyramide, afin de produire les porcs charcutiers (résultats du croisement entre une femelle croisée (la «trui parentale») et un verrat parental croisé ou de race pure).

Figure 1. Organisation générale des dispositifs d'amélioration génétique porcins.



Cette organisation pyramidale s'accompagne d'une spécialisation des populations de l'étage de sélection. Schématiquement, on y trouve des populations maternelles sélectionnées principalement pour améliorer les caractères de prolificité et d'aptitudes maternelles (qui sont combinées en multiplication pour produire des truies parentales croisées), et des populations paternelles sélectionnées pour améliorer les caractères d'engraissement et la valeur de la carcasse (utilisées comme verrat terminal en race pure ou après croisement en multiplication).

L'ensemble du travail d'amélioration génétique est réalisé à l'étage de sélection sur les populations de race pure et les lignées synthétiques ; l'objectif final est cependant d'améliorer les performances de reproduction et d'aptitudes maternelles des truies parentales croisées et les performances d'engraissement et de qualité de la carcasse des porcs charcutiers terminaux des élevages de production. Cette concentration de l'effort de sélection au sommet de la pyramide est fondée sur le principe que le progrès génétique qui y est réalisé se transmet jusqu'à l'étage de production.

À l'exception du schéma bavarois s'appuyant en partie sur un testage sur descendance (Habier *et al* 2007), les dispositifs d'évaluation génétique porcins actuels reposent sur un contrôle en élevage des performances des jeunes candidats mâles et femelles pour quelques caractères d'engraissement et de composition de la carcasse ; ces mesures sont parfois complétées par le contrôle en station puis l'abattage d'un nombre limité de collatéraux des candidats, permettant de disposer de performances pour des caractères trop complexes ou trop coûteux à mesurer en élevage (efficacité alimentaire) ou enre-

gistrés post mortem (caractères de qualité de la viande, composition de la carcasse...). Dans les populations maternelles, les performances de reproduction des truies (prolificité, aptitudes maternelles) sont également considérées. Les valeurs génétiques sont estimées grâce à la méthodologie du BLUP (*Best Linear Unbiased Predictor*, ou meilleur prédicteur linéaire non biaisé) appliquée à un modèle animal. Chaque individu (jeune candidat mâle ou femelle, reproducteur) dispose ainsi d'une valeur génétique estimée pour tous les caractères de l'objectif de sélection de sa population, y compris pour les caractères pour lesquels il ne dispose pas de phénotype propre. En général, seules les performances des individus de la population sélectionnée sont considérées dans l'évaluation génétique, mais certains schémas prennent également en compte des performances d'apparentés croisés.

Les décisions de sélection ou de réforme des candidats mâles et femelles sont prises dès la fin de leur contrôle de performances en ferme (un peu avant 6 mois d'âge), et les intensités de sélection sont fortes. Ainsi, dans la population collective Large White type femelle française par exemple, 1 candidate sur 11 est conservée pour la reproduction en sélection et 1 candidat sur 65 entre en Centre d'Insémination Artificielle (CIA) et sera utilisé en sélection ; dans la population collective Piétrain française, ces taux sont de 1/7 et 1/50, respectivement. Les individus retenus sont mis à la reproduction dès 8 mois d'âge, et leur carrière de reproducteur en sélection est courte, environ 1 an pour les mâles et 2 ans pour les femelles. Ce turnover rapide s'explique d'une part par le progrès génétique réalisé dans les populations, qui fait que la supériorité génétique des reproducteurs

en activité par rapport au niveau des meilleurs candidats s'amenuise rapidement, ce qui impose leur réforme et leur remplacement, et d'autre part par la volonté de limiter la descendance des verrats dans les élevages de sélection afin de préserver la variabilité génétique dans la population. En conséquence directe de ces pratiques, les intervalles de génération sont très faibles, tant sur la voie mâle que sur la voie femelle, et seules des carrières encore plus courtes pourraient les réduire davantage.

Faiblesses des dispositifs actuels :

Cette organisation des schémas d'amélioration porcins, bien que résultant en des intervalles de génération faibles et en des intensités de sélection fortes tous deux propices à un progrès génétique important, présente cependant plusieurs faiblesses.

Tout d'abord, la faible durée de carrière des reproducteurs limite le nombre de leurs produits phénotypés, et la précision d'estimation de leurs valeurs génétiques reste donc globalement faible. Ainsi, le Coefficient de Détermination (CD) moyen des verrats de CIA actifs dans les populations collectives françaises pour les caractères mesurés sur les individus contrôlés en station varie entre 0,18 pour le rendement technologique estimé (caractère le moins héritable, $h^2=0,28$) et 0,50 pour le taux de muscle des pièces (caractère le plus héritable, $h^2=0,75$), et se limite à 0,30 pour le nombre de porcelets nés vivants par portée ($h^2=0,10$) dans les populations maternelles. Il en va de même pour le CD des valeurs génétiques des jeunes candidats en fin de contrôle de performances, en moyenne très modeste pour ces mêmes caractères (0,10, 0,40 et 0,16, respectivement). Cette précision limitée des valeurs génétiques estimées au moment de la sélection des reproducteurs de renouvellement pénalise le progrès génétique, les risques de sélectionner un individu qui se révélera finalement mauvais étant non négligeables. Des schémas fondés sur un testage sur descendance des verrats (comme le schéma bavarois) permettraient certes de disposer de valeurs génétiques estimées plus précises et donc de diminuer les risques lors de la sélection des candidats, mais au prix cependant d'un allongement de l'intervalle entre générations et d'une réduction de l'intensité de sélection venant réduire le progrès génétique réalisé.

Comme évoqué précédemment, l'ensemble du travail d'amélioration génétique est réalisé à l'étage de sélection sur les populations de race pure et les lignées synthétiques, mais l'objectif réel est d'améliorer les performances des truies parentales croisées et des porcs

charcutiers terminaux des élevages de production (qui représentent 95% de la population porcine globale). Le principe sous-jacent à cette organisation est que le progrès génétique réalisé à l'étage de sélection dans les populations de race pure et les lignées se transmet (avec un retard d'une ou deux générations) aux individus croisés de l'étage de production. Généralement, les valeurs génétiques des candidats et des reproducteurs des élevages de sélection sont estimées uniquement à partir des performances d'individus des races pures (ou des lignées) enregistrées en élevage ou dans des stations de contrôle. Cependant, ces performances peuvent parfois se révéler être des prédicteurs imparfaits de celles des descendants croisés élevés dans des conditions de production, pour deux raisons principalement. Tout d'abord, les individus de l'étage de sélection sont élevés dans des environnements extrêmement maîtrisés et d'excellent statut sanitaire, et ces conditions d'élevage favorables peuvent différer sensiblement des conditions de production. Or, les animaux les plus performants dans un milieu peuvent se révéler moyens voire médiocres dans un environnement différent. De telles interactions génotype x milieu ont par exemple été mises en évidence chez le porc par Merks (1989). Par ailleurs, des effets de dominance et d'épistasie inexistantes dans deux races pures peuvent exister chez les individus issus de leur croisement. Ainsi, plusieurs auteurs rapportent des corrélations entre les performances d'individus de race pure d'élevages de sélection et celles de leurs descendants croisés comprises entre 0,3 et 1 selon les caractères et les populations considérées (Merks 1989, Lutaaya *et al* 2001, Habier *et al* 2007). En conséquence, le progrès génétique réalisé peut être sensiblement plus faible à la base de la pyramide qu'à son sommet.

Il est possible d'augmenter le progrès génétique réalisé en élevage de production en combinant les performances des individus de race pure à celles d'apparentés croisés lors de l'évaluation génétique des populations en sélection (Wei et Van der Steen 1991), et plusieurs schémas de sélection porcins se sont engagés dans cette voie. Cependant, la remontée d'informations en provenance des élevages de production s'avère en général difficile faute de généalogies de qualité suffisante et en l'absence de contrôle de performances. De plus, 1 ou 2 générations (parfois plus dans le cas de croisements complexes) séparent les populations sélectionnées des porcs charcutiers, et sauf à vouloir organiser un testage sur descendance (qui augmenterait considérablement l'intervalle entre générations et pénaliserait le progrès génétique réalisé annuellement), le

gain de précision obtenu pour les jeunes candidats grâce à leurs apparentés croisés serait minime.

2 / Opportunités offertes par la sélection génomique dans l'espèce porcine

Une puce pangénomique de 60 000 marqueurs SNP est disponible chez le porc depuis 2009. Les résultats de Du *et al* (2007), de Uimari et Tapio (2011), ainsi que de récents sondages dans les principales populations porcines françaises en sélection (programme de recherche DELISUS ; Boitard communication personnelle) indiquent qu'avec une telle densité de marqueurs, le déséquilibre de liaison moyen (r^2) entre SNP adjacents serait supérieur à 0,20 (valeur de DL en-deçà de laquelle l'évaluation génomique ne semble pas efficace selon le retour d'expérience des schémas bovins laitiers français et étrangers) ; la mise en place d'une sélection génomique dans l'espèce porcine est donc envisageable. De plus, il est probable que des puces beaucoup plus denses seront disponibles dans les prochaines années, qui permettront de tirer parti d'associations encore plus fortes entre marqueurs et QTL. Cependant, avant de se lancer dans une telle évolution des schémas d'amélioration génétique porcins, il est impératif de se demander en quoi la sélection génomique pourrait améliorer leur efficacité.

2.1 / Intervalle entre générations

L'intervalle entre générations dans les dispositifs porcins actuels est très faible, de l'ordre de 2 ans dans les populations collectives françaises (Maignel *et al* 1998). Un changement des modalités d'évaluation génétique n'aurait qu'un impact très limité sur ce paramètre.

2.2 / Intensités de sélection

Dans un dispositif fondé sur l'évaluation génomique, chaque individu génotypé peut se voir calculer une valeur génomique, même s'il n'a pas de performance propre et ne dispose d'aucun apparenté phénotypé. Le choix des reproducteurs de renouvellement peut donc se faire en théorie parmi l'ensemble des porcelets nés, permettant de maximiser l'intensité de sélection et donc d'augmenter le progrès génétique réalisé. En pratique, le nombre de candidats dans un tel dispositif, et par conséquent l'intensité de sélection effective, seraient fonction des éventuels surcoûts liés au génotypage que les sélectionneurs seraient prêts à supporter compte tenu des gains génétiques supplémentaires attendus. Les intensités de sélection

actuellement réalisées dans les schémas porcins sont cependant déjà élevées (le coût des mesures réalisées en ferme sur les candidats est faible, permettant de contrôler une forte proportion des porcelets nés), et le gain attendu en jouant sur ce paramètre pourrait donc ne pas être majeur.

2.3 / Précision des valeurs génétiques estimées

L'évaluation génomique peut en revanche avoir un impact beaucoup plus important sur l'efficacité des schémas par le biais d'une augmentation de la précision des valeurs génétiques estimées. Comme indiqué précédemment, les valeurs génétiques des candidats pour les caractères de reproduction (dans les lignées femelles) et pour les caractères mesurés sur collatéraux (caractères de qualité, d'efficacité alimentaire) sont estimées avec des précisions souvent modestes (CD moyens compris entre 0,10 et 0,40 selon le caractère). Le retour d'expérience des schémas de sélection génomique bovins laitiers et d'études réalisées par simulations (figure 2, établie d'après Meuwissen 2009) montrent qu'il serait possible d'estimer pour de jeunes candidats des valeurs génomiques plus précises que leurs valeurs génétiques actuelles, à la condition toutefois de disposer d'une population de référence de taille suffisante et de marqueurs couvrant le génome de manière suffisamment dense.

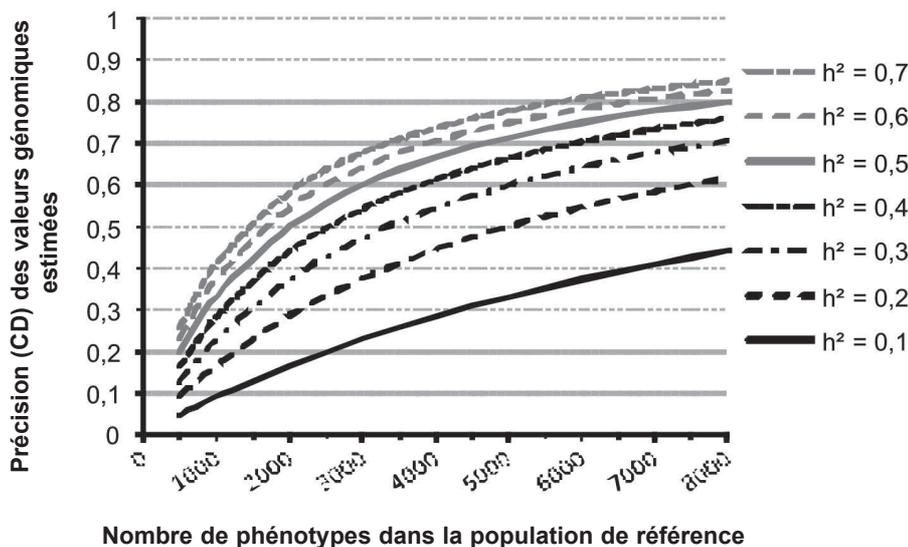
On constate cependant sur la figure 2 que l'obtention de CD supérieurs à 0,3 pour des caractères faiblement héréditaires (ce qui est le cas de la qualité de la viande, de la prolificité et des aptitudes maternelles par exemple) requerrait une population de référence d'environ 3000 à 5000 individus.

2.4 / Sélection sur des caractères non mesurables en routine

L'évaluation génomique pourrait également permettre de sélectionner les populations pour des caractères nouveaux, trop complexes ou trop coûteux à mesurer en routine (y compris dans le cadre d'un contrôle sur collatéraux) ou nécessitant la mise en place de protocoles expérimentaux spécifiques : mesures comportementales, résistance à des maladies, composition des rejets, maturité des porcelets à la naissance, homogénéité des porcelets intra-portée, odeurs sexuelles, qualité des aplombs, etc. On pourrait par exemple imaginer d'enregistrer dans quelques élevages équipés d'automates de pesée les poids des porcelets à la naissance et au sevrage pour un échantillon de truies génotypées ; ces truies constitueraient une population de

Figure 2. Précision théorique (CD) des valeurs génomiques estimées en fonction de la taille de la population de référence et de l'héritabilité (h^2) du caractère considéré, pour une population de taille efficace (N_e) = 100 (d'après Meuwissen 2009).

On fait l'hypothèse que chaque QTL est en déséquilibre de liaison complet avec un marqueur SNP.



référence permettant d'estimer les valeurs génomiques des candidats à la sélection pour l'homogénéité du poids des porcelets intra-portée, la croissance des porcelets en période d'allaitement et la production laitière des truies, etc. Les coûts de génotypage et de phénotypage liés à la constitution initiale et au renouvellement périodique de telles populations de référence seraient cependant élevés, et ce type d'approche se limiterait vraisemblablement à quelques caractères parmi les plus importants pour la filière.

2.5 / Augmentation du progrès génétique réalisé en production

Comme évoqué précédemment, le fait de sélectionner des individus de race pure élevés dans un milieu très contrôlé dans le but d'améliorer les performances de leurs descendants croisés en conditions de production constitue l'une des principales faiblesses des dispositifs d'amélioration génétique porcins, et c'est probablement sur ce point que la sélection génomique pourrait présenter son plus grand intérêt chez le porc. Il serait en effet possible, en s'appuyant sur une population de référence constituée d'animaux croisés, idéalement phénotypés dans des conditions de milieu semblables à celles des élevages de production, d'estimer la valeur génomique de candidats de race pure pour des caractères exprimés par des animaux commerciaux (truies parentales ou porcs charcutiers). Dekkers (2007) et Kinghorn *et al* (2010) ont montré par simulation qu'une telle approche permettrait un progrès génétique plus important dans la population croisée qu'une sélection génomique s'appuyant sur une population de référence consti-

tuee d'animaux de race pure. Elle serait également plus efficace qu'une sélection classique (BLUP-modèle animal) combinant les informations des individus de race pure et de leurs apparentés croisés (Dekkers 2007) sans nécessiter l'organisation d'un enregistrement en routine des généalogies et performances des descendants de production et sans augmenter l'intervalle entre générations.

Deux modèles ont été proposés pour estimer les effets des segments chromosomiques à partir d'une population de référence constituée d'animaux croisés (Ibanz-Escriche *et al* 2009, Kinghorn *et al* 2010). Pour décrire ces modèles, considérons par exemple que les individus de la population de référence sont issus du croisement entre une truie parentale croisée ayant des parents des races pures A et B et un verrat terminal de race pure C, et que chaque segment chromosomique est défini par un seul marqueur biallélique.

Le premier modèle, le plus simple, fait l'hypothèse que l'effet du génotype à un marqueur sur la performance d'un animal croisé est identique quelles que soient les populations d'origine (A, B ou C) des allèles reçus par l'animal pour ce marqueur :

$$y_i = \mu + \sum_j X_{ij} \beta_j + e_i,$$

où y_i est la performance de l'individu croisé i , μ est la performance moyenne, X_{ij} est le nombre d'allèles 1 (0, 1 ou 2) au locus du $j^{\text{ème}}$ marqueur de l'individu i , β_j est l'effet de substitution de l'allèle 1 au locus du $j^{\text{ème}}$ marqueur dans la population de référence, et e_i est le résidu pour l'individu i .

La valeur génomique estimée VGE d'un candidat q d'une quelconque des

3 races pures A, B ou C se calcule alors simplement comme la somme sur l'ensemble des marqueurs considérés du produit du nombre d'allèles 1 portés par l'individu q au locus j (X_{qj}) par l'effet de substitution estimé pour cet allèle :

$$VGE_q = \sum_j X_{qj} \beta_j$$

Le second modèle considère au contraire que l'effet du génotype à un marqueur peut être différent en fonction de la population d'origine des allèles au marqueur :

$$y_i = \mu + \sum_j (A_{ijk}^P \beta_{jk}^P + A_{ijl}^M \beta_{jl}^M) + e_i$$

où y_i est la performance de l'individu croisé i , μ est la performance moyenne, A_{ijk}^P est l'allèle porté par l'individu au $j^{\text{ème}}$ SNP provenant de la race k transmis par le père de l'individu (ici, l'allèle provient forcément de la race C puisque le père est de race pure), et β_{jk}^P est l'effet de substitution de l'allèle A_{ijk}^P . De même, A_{ijl}^M est l'allèle porté par l'individu au $j^{\text{ème}}$ SNP provenant de la race l transmis par la mère de l'individu (ici, l'allèle provient de la race A ou de la race B puisque la mère de l'individu est croisée), et β_{jl}^M est l'effet de substitution de l'allèle A_{ijl}^M . Dans le cas d'une population croisée constituée d'individus (A x B) x C, on estimera donc pour chaque SNP 3 effets : un effet β_{jc}^P pour l'allèle transmis par les pères, et deux effets β_{ja}^M et β_{jb}^M pour l'allèle transmis par les mères selon qu'il provient à l'origine de la population A ou B, respectivement.

La valeur génomique des candidats de chacune des 3 races pures constituant le croisement est ensuite calculée à partir des effets des marqueurs estimés pour la race correspondante ; par exemple, pour un individu q de la race pure A :

$$VGE_q = \sum_j X_{qjA} \beta_{jA}$$

La sélection génomique s'appuie sur le déséquilibre de liaison existant entre marqueurs et QTL. Cependant, les associations entre marqueurs et QTL, et par conséquent les effets des génotypes aux marqueurs, peuvent varier entre populations. Le modèle estimant les effets des marqueurs intra-race semble donc plus pertinent, en particulier si les populations de race pure considérées ont divergé depuis de nombreuses générations (par exemple, populations paternelles vs populations maternelles) ou si la densité de marqueurs est faible (augmentant la probabilité de recombinaisons entre marqueurs et QTL). Cependant, le nombre de paramètres à estimer avec ce modèle est plus important que lorsqu'on estime un seul effet par SNP commun à toutes les races, et une population de

Tableau 1. Synthèse des apports potentiels d'une évaluation génomique dans les schémas d'amélioration génétique porcins.

Paramètre considéré	Apport potentiel de la sélection génomique	Conditions nécessaires / facteurs limitants
Intervalle entre générations	Très limité à nul	
Intensités de sélection	Limité (dégradation ?)	Coût du génotypage des candidats
Précision des valeurs génétiques estimées	Fort	Population de référence de taille suffisante / coût
Sélection sur de nouveaux caractères non mesurables en routine	Fort	Disposer de phénotypes pertinents / coût de constitution des populations de référence
Prise en compte des interactions génotype x milieu	Fort	Phénotypage des populations de référence en milieu de production
Estimation de la valeur en croisement d'individus de races pures (ou de lignées)	Fort	Constitution de populations de référence constituées de truies parentales et de porcs charcutiers / modélisation statistique adéquate

référence de plus grande taille est donc nécessaire pour conserver une bonne précision d'estimation des effets des marqueurs (Ibanz-Escriche *et al* 2009). De plus, le modèle «intra-race» nécessite de connaître l'origine raciale de chaque allèle porté par les individus de la population de référence croisée, mais des erreurs d'attribution seraient en pratique inévitables, réduisant peut-être son efficacité par rapport à un modèle simple.

Ces deux modèles, appliqués à une population de référence croisée, permettraient donc d'estimer l'effet de substitution des allèles présents dans les races pures dans le fond génétique de la population croisée que l'on cherche à améliorer. Selon les auteurs (Ibanz-Escriche *et al* 2009, Kinghorn *et al* 2010), une partie des effets de dominance et d'épistasie s'exprimant en croisement, bien que non explicitement modélisés, seraient implicitement inclus dans ces estimations. Une meilleure modélisation des effets génétiques non additifs, sujet qui fait actuellement l'objet de travaux de recherche, permettrait sans doute d'accroître l'efficacité d'une telle approche.

Par ailleurs, l'enregistrement des performances de la population de référence dans un environnement représentatif du milieu de production permettrait d'estimer pour les individus des races pures des valeurs génomiques pour les caractères exprimés en conditions commerciales, et de s'affranchir ainsi du problème d'éventuelles interactions génotype x milieu. Le progrès génétique réalisé dans la population croisée à l'étage de production serait donc plus important grâce à un tel dispositif ; en revanche, le progrès réalisé à l'étage de sélection serait nécessairement plus faible qu'avec des objectifs de sélection définis en race pure.

Bien que ces modèles n'aient pour l'instant pas été testés sur de vraies populations mais aient uniquement fait l'objet d'études par simulation, la possibilité de sélectionner les individus de races pures en se fondant sur des valeurs génomiques estimées à l'aide d'une population de référence croisée semble néanmoins prometteuse pour les schémas porcins. Le tableau 1 synthétise les apports potentiels d'une évaluation génomique pour les différents paramètres considérés.

3 / Mise en place d'une évaluation génomique chez le porc : de l'idéal à la réalité

Compte tenu de l'ensemble des éléments précédents, on pourrait imaginer faire évoluer les actuels dispositifs d'amélioration génétique porcins organisés autour d'un contrôle de performances des candidats et de collatéraux et d'évaluations génétiques BLUP-modèle animal vers des dispositifs fondés sur des évaluations génomiques. Dans l'idéal (figure 3), des populations de référence constituées de porcs charcutiers croisés élevés dans des conditions de production et phénotypés pour l'ensemble des caractères d'intérêt de production et de qualité permettraient d'établir les équations de prédiction de la valeur génomique des populations en sélection. Ce dispositif serait complété dans les populations maternelles par une évaluation génomique pour la prolificité et les aptitudes maternelles s'appuyant sur des populations de référence constituées de truies parentales croisées d'élevages de production.

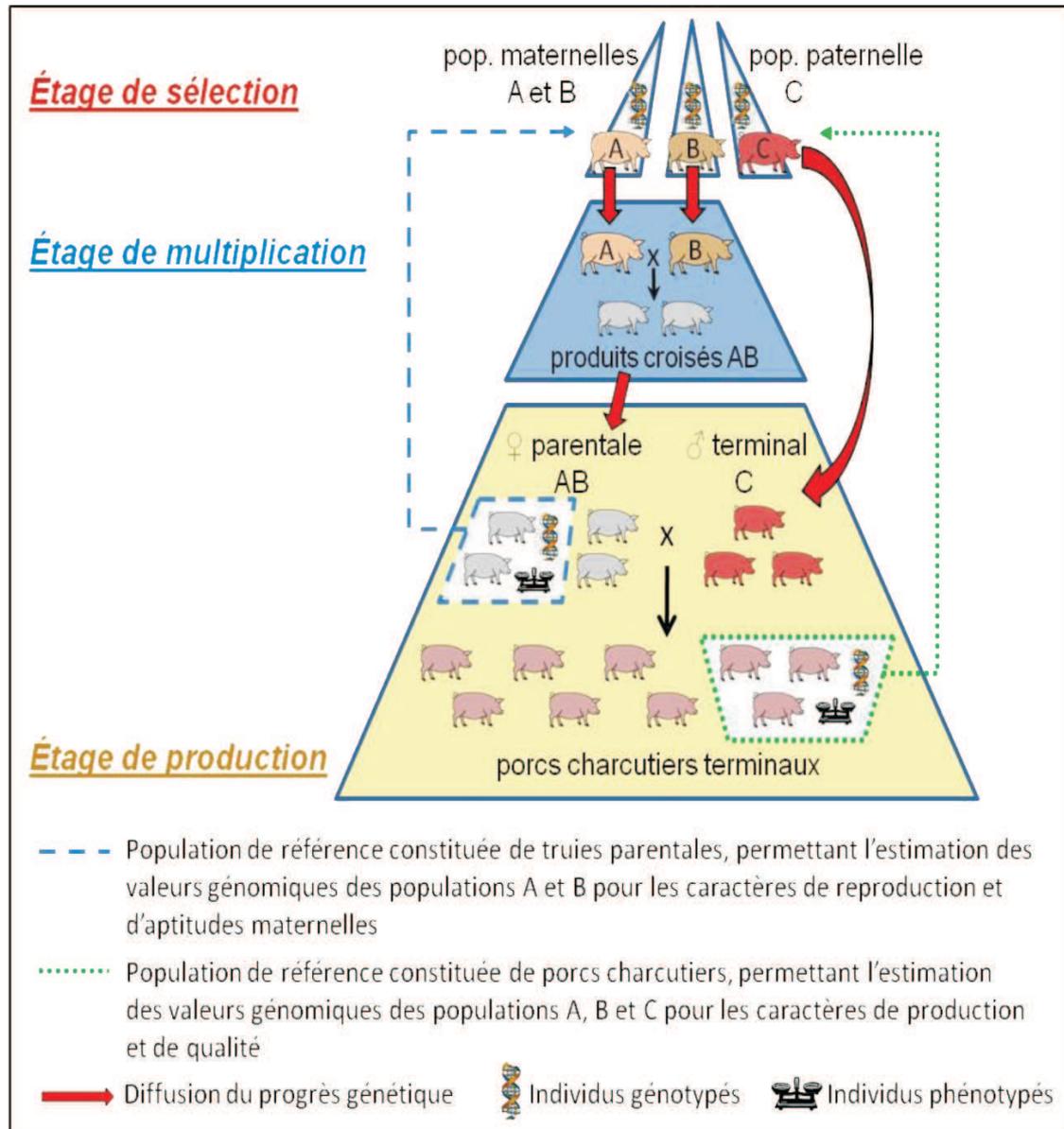
Cependant, la mise en place effective d'un tel scénario nécessite de

résoudre un certain nombre de difficultés pratiques.

3.1 / Coût de la constitution des populations de référence et du génotypage des candidats

On ne dispose pour l'instant pas des éléments nécessaires au dimensionnement précis des populations de référence à mettre en place en vue d'une sélection génomique chez le porc. Cependant, compte tenu des résultats de simulations, du retour d'expérience des schémas génomiques bovins laitiers, et du niveau présumé du déséquilibre de liaison entre les marqueurs de la puce porcine 60 K actuellement disponible, une population de référence de plusieurs milliers d'individus semble nécessaire pour atteindre des précisions satisfaisantes (c'est-à-dire au moins égales aux précisions obtenues dans les dispositifs actuels) pour une population sélectionnée dont la taille efficace est proche de 100, en particulier si l'on s'intéresse à des caractères peu héréditaires comme la qualité de la viande ou les aptitudes maternelles. Il est probable que les coûts de génotypage diminueront au cours des prochaines années, mais en considérant un prix approximatif de 150 € par individu, la constitution initiale d'une population de référence s'élèverait donc à plusieurs centaines de milliers d'euros, auxquels viendraient s'ajouter les coûts de phénotypage, évidemment dépendants de la liste des caractères considérés. De plus, une population de référence doit être renouvelée régulièrement afin de préserver la qualité de prédiction des valeurs génomiques des candidats (Solberg *et al* 2009). La vitesse de perte d'efficacité de la prédiction est inversement liée au niveau du déséquilibre de liaison et à la densité des marqueurs, et même si des références précises manquent encore pour l'espèce porcine, on peut néanmoins faire l'hypothèse de la validité d'une population de

Figure 3. Organisation théorique d'un dispositif d'amélioration génétique porcine s'appuyant sur des évaluations génomiques et des populations de référence constituées d'individus croisés, dans le cas d'un croisement 3 voies de deux populations maternelles A et B et d'une population paternelle C.



référence pendant 3 générations, ce qui impliquerait de renouveler la dépense initiale environ tous les 5 ans. De plus, ces coûts seraient approximativement doublés pour les populations maternelles, puisque les phénotypes pour les caractères de production et de reproduction ne pourraient pas être mesurés dans une même population de référence. Enfin, l'évaluation génomique d'éventuels nouveaux caractères dont le phénotypage serait incompatible avec des conditions d'élevage classiques pourrait nécessiter la constitution de populations de référence spécifiques supplémentaires.

A ces charges viendraient s'ajouter celle du génotypage des candidats à la sélection, directement liées au nombre de candidats à évaluer, et donc à l'intensité de sélection et au progrès gé-

tique réalisés. Des études (Druet et Georges 2010) montrent que sous certaines conditions (par exemple, génotypage d'une partie des reproducteurs de la population pour l'ensemble des marqueurs considérés dans la population de référence), il serait possible de ne génotyper les candidats que pour quelques milliers de marqueurs informatifs répartis de manière homogène sur le génome, et « d'imputer » avec un très faible taux d'erreur leur génotype aux marqueurs manquants. Ces techniques d'imputation permettraient de diminuer sensiblement les coûts de génotypage des candidats, tout en préservant l'efficacité de l'évaluation génomique. Néanmoins, même sous l'hypothèse favorable d'un coût de 30 € par individu, le génotypage pour une puce de faible densité des 9200 candi-

dats mâles et des 35 000 candidates femelles contrôlés en ferme en 2008 dans la population collective française Large White type femelle s'élèverait par exemple à plus de 1300 K€.

Les sommes à engager pour mettre en place une sélection génomique dans les dispositifs porcins sont donc importantes. Contrairement à la situation des schémas bovins laitiers pour lesquels les économies réalisées grâce à l'arrêt du testage sur descendance (40 000 € par candidat mâle testé en France) ont compensé les frais liés à l'évaluation génomique, un mode de financement reste à trouver chez le porc. Le coût des mesures actuellement réalisées sur les candidats en ferme, qui se limitent en général à des pesées et des mesures par ultrasons, est faible, et un éventuel arrêt

du contrôle en ferme permis par une évaluation génomique ne conduirait pas à des économies substantielles. Le contrôle des collatéraux est en revanche plus coûteux, nécessitant l'équipement des élevages ou des stations en automates d'alimentation et le suivi des animaux à l'abattoir pour les mesures de qualité, avec parfois une diminution de la valeur de la carcasse en cas de prélèvement de tissus. Cependant, on peut dans une première approche considérer que le phénotypage des individus des populations de référence remplacerait celui des collatéraux, et là encore aucun report de financement vers des génotypes n'est à espérer.

Toutefois, si on tient compte du volume annuel de la production porcine (25 millions de porcs charcutiers ont été abattus en France en 2008), l'ensemble des coûts liés à la constitution des populations de référence, au génotypage des candidats et à l'encadrement des dispositifs serait a priori faible comparé au profit supplémentaire réalisé à l'étape de production grâce à un accroissement du progrès génétique permis par la sélection génomique. On pourrait en conséquence imaginer de financer l'augmentation du coût de la sélection par une hausse du prix de vente des reproducteurs et de la semence. Les valeurs individuelles d'une cochette et d'un verrat sont certes faibles (entre 260 et 300 € pour une cochette, environ 650 € pour un verrat, IFIP 2009), mais le nombre de reproducteurs vendus annuellement est très élevé (environ 400 000 cochettes et 9400 verrats ont été vendus en France en 2008, tous types génétiques confondus, IFIP 2009), si bien que cette augmentation serait a priori modérée. Des études économiques devront cependant être conduites afin de quantifier de manière plus précise les sommes concernées et trouver le mode de financement le plus adapté à une évolution des dispositifs porcins.

3.2 / Cas des petites populations

De nombreuses lignées et variétés de races en sélection comptent moins de 200 femelles reproductrices, et même si la taille de la population de référence nécessaire à l'estimation de valeurs génomiques avec une précision donnée est plus faible dans les populations de petite taille efficace (N_e), la constitution de dispositifs suffisamment puissants au sein de chacune de ces populations paraît difficile. Il semble néanmoins envisageable dans de tels cas d'utiliser des équations de prédiction des valeurs génomiques établies dans une autre race. L'efficacité d'une telle solution nécessite cependant une densité de marqueurs beaucoup plus forte que lorsque

la population de référence et la population sélectionnée sont du même type génétique, afin de disposer d'associations entre marqueurs et QTL valides dans l'ensemble des populations. De Roos *et al* (2008) ont ainsi estimé à 300 000 le nombre de SNP nécessaires à une évaluation efficace de la population bovine laitière Jersey à l'aide d'effets de marqueurs estimés en Holstein. Il est possible que la puce porcine de 60 000 SNP actuellement sur le marché soit insuffisante pour de telles applications, mais des puces plus denses seront certainement bientôt disponibles. Par ailleurs, une représentation, même faible, d'individus de la population à évaluer au sein de la population de référence semble permettre d'améliorer sensiblement la précision d'évaluation (Harris *et al* 2008, de Roos *et al* 2008), et les organisations de sélection détentrices de petites populations seront donc vraisemblablement amenées à collaborer afin de constituer des dispositifs multiraciaux.

3.3 / Partages des ressources

Cette mise en commun de ressources ne se limitera vraisemblablement pas aux seules petites populations. En effet, compte tenu du nombre élevé de populations porcines sélectionnées (par exemple, en France, 5 grandes races, une dizaine de variétés de ces races et une quinzaine de lignées composites), de la variété des truies parentales et des types de porcs charcutiers produits, des capacités de phénotypage limitées pour certains caractères (efficacité alimentaire) et des sommes à engager, il semble peu probable de pouvoir constituer et entretenir des populations de référence spécifiques pour un grand nombre de populations sélectionnées ou de croisements pratiqués. Une mise en commun des phénotypes et génotypes de populations génétiquement proches détenues par plusieurs Organisations de Sélection semble donc nécessaire, permettant d'augmenter la puissance des dispositifs tout en limitant les coûts engagés par chaque partenaire. Des collaborations internationales sont également envisageables, à l'image du projet Euro-genomics destiné à mettre en commun les phénotypes et génotypes de 16 000 taureaux Holstein de 6 pays européens (Lund *et al* 2010). En pratique, un tel partage des ressources suppose cependant une homogénéité des phénotypes mis en commun par les partenaires potentiels, ce qui était le cas chez les bovins laitiers de par l'existence de l'évaluation génétique internationale INTERBULL. Celle-ci n'a malheureusement pas son équivalent actuellement chez le Porc, espèce dans laquelle un effort préalable serait donc nécessaire.

Conclusions

La sélection génomique offre plusieurs pistes d'amélioration de l'efficacité des schémas d'amélioration génétique porcins, bien qu'il soit peu probable que sa mise en place entraîne des mutations aussi profondes que les changements auxquels nous assistons dans les schémas bovins laitiers, principalement avec l'arrêt du testage sur descendance. Les perspectives les plus intéressantes semblent être la possibilité d'estimer plus précisément la valeur génétique des individus pour les caractères peu héréditaires et non observables sur les candidats (qualité de la viande, efficacité alimentaire, aptitudes maternelles) et de sélectionner plus efficacement les individus de race pure pour les caractères exprimés en croisement et dans des conditions de production.

Bien que cette méthodologie permette d'estimer la valeur génétique de candidats sans performance propre et sans performance d'apparenté, l'enregistrement de phénotypes pour les caractères que l'on souhaite améliorer reste tout de même primordial afin d'établir les équations de prédiction des valeurs génomiques. Les stations publiques porcines, bien que les conditions d'environnement y soient sensiblement différentes de celles des élevages commerciaux, seraient naturellement un outil de choix pour la constitution des populations de référence. Elles permettent en effet de mesurer de manière précise les performances des individus pour un grand nombre de caractères, et la nécessité de leur maintien dans des schémas d'amélioration génétique s'appuyant sur la sélection génomique est donc primordiale. La création d'un réseau d'élevages de référence dont la fonction serait analogue à celles des stations pour les caractères de reproduction et d'aptitudes maternelles semble également nécessaire.

De nombreuses incertitudes demeurent cependant quant aux modalités de mise en œuvre de la sélection génomique dans les schémas porcins, qui devraient être levées par les prochaines avancées technologiques (puces plus denses et à moindre coût) et méthodologiques (meilleure prise en compte des effets génétiques non additifs, mélanges de populations). Des études par simulations devront également être conduites afin de déterminer la manière optimale de réorganiser les schémas d'amélioration génétique porcins autour de cette nouvelle méthodologie.

Références

- Dekkers J.C.M., 2007. Marker assisted selection for commercial crossbred performance. *J. Anim. Sci.*, 85, 2104-2114.
- De Roos A.P.W., Hayes B.J., Spelman R., Goddard M.E., 2008. Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. *Genetics*, 179, 1503-1512.
- Druet T., Georges M., 2010. A hidden Markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and Quantitative Trait Locus fine mapping. *Genetics*, 184, 789-798.
- Du F.X., Clutter A.C., Lohuis M.M., 2007. Characterizing linkage disequilibrium in pig populations. *Int. J. Biol. Sci.*, 3, 166-178.
- Fritz S., Colleau J.J., Druet T., Boscher M.Y., Rossignol M.N., Malafosse A., Boichard D., 2003. Mise en place d'une Sélection Assistée par Marqueurs dans les trois principales races bovines laitières françaises. *Renc. Rech. Rum.*, 10, 53-56.
- Guillaume F., Boichard D., Ducrocq V., Fritz S., 2011. Utilisation de la sélection génomique chez les bovins laitier. In : Numéro spécial, Amélioration génétique. Mulsant P., Bodin L., Coudurier B., Deretz S., Le Roy P., Quillet E., Perez J.M. (Eds). INRA Prod. Anim., 24, 363-368.
- Habier D., Götz K.U., Dempfle L., 2007. Estimation of genetic parameters on test stations using purebred and crossbred progeny of sires of the Bavarian Piétrain. *Livest. Sci.*, 107, 142-151.
- Harris B.L., Johnson D.L., Spelman R.J., 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Proc. Interbull Meeting, Niagara Falls, Canada. <http://www.cabi.org/cabdirect/FullTextPDF/2010/20103193083.pdf>
- Hu Z.L., Reedy J.M., 2007. Animal QTLdb: beyond a repository - A public platform for QTL comparisons and integration with diverse types of structural genomic information. *Mamm. Genome*, 18, 1-4.
- Ibanz-Escriche N., Fernando R.L., Toosi A., Dekkers J.C.M., 2009. Genomic selection of purebreds for crossbred performance. *Genet. Sel. Evol.*, 41, 12-21.
- IFIP, Institut du Porc, 2009. Le porc par les chiffres. Ed. IFIP, France, 54p.
- Kinghorn B.P., Hickey J.M., van der Werf J.H.J., 2010. Reciprocal recurrent genomic selection for total genetic merit in crossbred individuals. Proc 9th WCGALP, August 1-6, Leipzig, Germany. Comm. 36.
- Lund M.S., de Roos A.P.W., de Vries A.G., Druet T., Ducrocq V., Fritz S., Guillaume F., Gulbrandtsen B., Liu Z., Reents R., Schrooten C., Seefried M., Su G., 2010. Improving genomic prediction by EuroGenomics collaboration. Proc 9th WCGALP, August 1-6, Leipzig, Germany. Comm. 0880.
- Lutaaya E., Misztal I., Mabry J.W., Short T., Timm H.H., Holzbauer R., 2001. Genetic parameter estimates from joint evaluation of purebreds and crossbreds in swine using the crossbred model. *J. Anim. Sci.*, 79, 3002-3007.
- Luther H., Vögeli P., Hofer A., 2009. Increasing genetic E. coli F18 resistance in Swiss pigs. Proc. 60th EAAP Meeting, 24-27 August, Barcelona, Spain. Comm. 18-03.
- Maignel L., Tribout T., Boichard D., Bidanel J.P., Guéblez R., 1998. Analyse de la variabilité génétique des races porcines Large White, Landrace Français et Piétrain, sur la base de l'information généalogique. *Journ. Rech. Porcine Fr.*, 30, 109-116.
- Merks J.W.M. 1989. Genotype × environment interactions in pig breeding programmes.
- VI. Genetic relations between performances in central test, on-farm test and commercial fattening. *Livest. Prod. Sci.*, 22, 325-339.
- Meuwissen T.H.E., 2009. Accuracy of breeding values of unrelated individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.*, 41, 35.
- Sidani C., Astruc J.M., Baelden M., Barillet F., Bibé B., Bonnot A., Boscher M.Y., Bouchel D., Bouffartigue B., Bouix J., Brochard M., Dion F., Francois D., Jouhet E., Jullien E., Leymarie C.R., Moreno C., Orlianges M., Palhière I., Perret G., Raoul J., Raynal A., Tiphine L., Tribon P., 2010. The french ovine scrapie plan: results and prospects. Proc. 9th WCGALP, August 1-6, Leipzig, Germany. Comm. 412.
- Schwob S., Riquet J., Bellec T., Kernalguen L., Tribout T., Bidanel J.P., 2009. Mise en place d'un programme de sélection assistée par marqueurs dans la population sino-européenne Duochan. *Journ. Rech. Porcine Fr.*, 41, 29-30.
- Solberg T.R., Sonesson A.K., Woolliams J.A., Odegard J., Meuwissen T.H.E., 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genet. Sel. Evol.*, 41, 53.
- Uimari P., Tapio M., 2011. Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *J. Anim. Sci.*, 89, 609-614.
- Van der Steen H.A.M., Prall G.F.W., Plastow G.S., 2005. Application of genomics to the pork industry. *J. Anim. Sci.*, 83 (E. Suppl.), E1-E8.
- Wei M., Van der Steen H.A.M., 1991. Comparison of reciprocal recurrent selection with pure-line selection systems in animal breeding (a review). *Anim. Breed. Abstr.*, 59, 281-298.

Résumé

L'évaluation génomique est une nouvelle méthode d'estimation de la valeur génétique des animaux d'élevage qui est en train de révolutionner l'organisation des schémas d'amélioration génétique bovins laitiers. La récente disponibilité d'une puce porcine de 60 000 marqueurs SNP permet d'envisager la mise en place de cette méthodologie chez le porc. Dans cette espèce, les perspectives les plus intéressantes de l'évaluation génomique semblent être la possibilité d'améliorer les populations pour des caractères non mesurables en routine, d'augmenter la précision des valeurs génétiques estimées, et de sélectionner les individus de l'étage de sélection pour leur valeur en croisement et en milieu de production. Les coûts de mise en œuvre d'une sélection génomique chez le porc seraient élevés, en raison principalement de la taille des populations de référence nécessaires à une évaluation précise, du nombre élevé de candidats à génotyper, et de la variété des populations et lignées sélectionnées. Le partage de ressources entre Organisations de Sélection et le recours aux techniques d'imputation génotypiques permettraient de maîtriser ces coûts.

Abstract

Potential applications of genomic selection in pig breeding schemes

Genomic evaluation, a new methodology to estimate the breeding value of livestock, is currently revolutionizing dairy cattle breeding, and its implementation in pigs is being considered. In this species, genomic selection could be used to improve populations for traits too costly to be recorded routinely, to increase the accuracy of the estimated breeding values, and to select the nucleus populations for their crossbred value expressed in a commercial environment. The implementation of genomic selection in pigs would be costly, mainly because of the size of the reference populations required to estimate accurate breeding values, the high number of candidates that should be genotyped, and the large number of selected lines. The sharing of reference populations by several breeding organizations and the use of genotype imputation techniques would help reduce the implementation costs.

TRIBOUT T., 2011. Perspectives d'application de la sélection génomique dans les schémas d'amélioration génétique porcins. In : Numéro spécial, Amélioration génétique. Mulsant P., Bodin L., Coudurier B., Deretz S., Le Roy P., Quillet E., Perez J.M. (Eds). INRA Prod. Anim., 24, 369-376.

PARTIE 3.

Article 2.

Efficiency of genomic selection in a purebred pig male line.

Tribout, T., C. Larzul, and F. Phocas. 2012. *Journal of Animal Science* 90(12):4164-4176.

JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Efficiency of genomic selection in a purebred pig male line

T. Tribout, C. Larzul and F. Phocas

J ANIM SCI 2012, 90:4164-4176.

doi: 10.2527/jas.2012-5107 originally published online August 2, 2012

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://www.journalofanimalscience.org/content/90/12/4164>



American Society of Animal Science

www.asas.org

Efficiency of genomic selection in a purebred pig male line¹

T. Tribout,*^{†2} C. Larzul,*[†] and F. Phocas*[†]

*INRA, UMR1313 Génétique Animale et Biologie Intégrative, F-78350 Jouy-en-Josas, France;
and [†]AgroParisTech, UMR1313 Génétique Animale et Biologie Intégrative, F-75231 Paris, France

ABSTRACT: Stochastic simulation was used to compare the efficiency of 3 pig breeding schemes based on either traditional genetic evaluation or genomic evaluation. The simulated population contained 1,050 female and 50 male breeding animals. It was selected for 10 yr for a synthetic breeding goal that included 2 traits with equal economic weights and heritabilities of 0.2 or 0.4. The reference breeding scheme, named BLUP-AM, was based on the phenotyping of all candidates (13,770 animals/yr) for Trait 1 and of relatives from 10% of the litters (270 animals/yr) for Trait 2 and on BLUP-Animal Model genetic evaluations. Under the first alternative scenario, named GE-1TP, selection was based on genomic breeding values (GBV) estimated with one training population (TP) made up of candidate relatives phenotyped for both traits, with a size increasing from 1,000 to 3,430 over time. Under the second alternative scenario, named GE-2TP, the GBV for Trait 2 were estimated using a TP identical to that of GE-1TP, but the GBV for Trait 1 were estimated using a large TP made up of candidates that increased in number from 13,770 to 55,080 over time. Over the simulated period, both genomic breeding schemes generated 39 to 58% more accurate EBV for Trait 2 than the reference scheme,

resulting in 78 to 128% (GE-1TP) and 63 to 84% (GE-2TP) greater average annual genetic trends for this trait. For Trait 1, GE-1TP was 18 to 24% less accurate than BLUP-AM, reducing average annual genetic trends by 27 to 44%. By contrast, GE-2TP generated 35 to 43% more accurate EBV and 8 to 22% greater average annual genetic trends for Trait 1 than the reference scheme. Consequently, GE-2TP was 27 to 33% more efficient in improving the global breeding goal than BLUP-AM whereas GE-1TP was globally as efficient as the reference scheme. Both genomic schemes reduced the inbreeding rate, the greatest decrease being observed for GE-2TP (−49 to −60% compared with BLUP-AM). In conclusion, genomic selection could substantially and durably improve the efficiency of pig breeding schemes in terms of reliability, genetic trends, and inbreeding rate without any need to modify their current structure. Even though it only generates a small TP, limited annual phenotyping capacity for traits currently only recorded on relatives would not be prohibitive. A large TP is, however, required to outperform the current schemes for traits recorded on the candidates in the latter.

Key words: breeding scheme, genomic selection, pig, production traits, simulation

© 2012 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2012.90:4164–4176

doi:10.2527/jas2012-5107

INTRODUCTION

Genomic selection (GS) has successfully been implemented for dairy cattle (e.g., in Holstein, Jersey,

and Brown-Swiss populations; Hayes et al., 2009). Compared with traditional schemes based on progeny testing, GS enables earlier accurate selection from a larger number of candidates. The ensuing shorter generation intervals and increased selection intensities have resulted in markedly greater annual genetic trends.

Such an improvement is tempting in other species. Implementation of GS can be envisaged in pigs since the availability of the *SNP60* Illumina BeadChip (Illumina, Inc., San Diego, CA). However, current pig breeding schemes are characterized by high selection

¹The authors acknowledge A. Legarra for providing the program used for the GBLUP evaluations and I. Misztal for supplying the BLUPF90 software. The authors would also like to thank J.C.M. Dekkers for his valuable advices. The authors wish to thank the ANR for financial support (ANR_10_GENOM_BTV_015).

²Corresponding author: thierry.tribout@jouy.inra.fr

Received January 10, 2012.

Accepted June 19, 2012.

intensities and very short generation intervals. Therefore, the impact of GS on these 2 characteristics is expected to be small. However, accuracy of EBV is generally poor in pigs, and GS could potentially improve scheme efficiency through this parameter (Muir, 2007).

Therefore, the benefits of GS should be evaluated carefully before any implementation. Lillehammer et al. (2011) showed that GS could increase the genetic gain in dam lines for traits only recorded on females. Nevertheless, no other study based on a pig nucleus structure has been conducted.

Most pig male line breeding schemes are based on the on-farm phenotyping of candidates for a few fattening traits and on phenotyping in testing stations of a limited number of relatives or candidates for traits that are expensive or difficult to measure. The accuracy of EBV is particularly low for these latter traits, resulting in limited genetic trends, and GS might increase selection efficiency. Nevertheless, the limited phenotyping capacity might not allow for the establishment of an adequate training population (TP) for these traits.

The objective of the present study was, therefore, to compare, in terms of genetic trends, accuracy, and inbreeding, the efficiency of a current pig male line breeding scheme based on the combined phenotyping of candidates and relatives with alternative schemes based on GS.

MATERIALS AND METHODS

The comparison of the efficiency of different breeding schemes was ensured by the stochastic simulation of a purebred paternal pig population made up of 1,050 breeding females and 50 breeding males, with a family structure similar to the French Piétrain breed. The simulated population was selected for a synthetic breeding goal that included 2 traits with equal economic weights per genetic SD. Trait 1 represents one of the fattening traits that are cheap and easy to record on farm on a large number of candidates (e.g., growth rate, ultrasonic backfat thickness, or loin depth) whereas Trait 2 represents a trait that is too difficult or too expensive to measure on a large scale (e.g., meat quality, feed efficiency, and intramuscular fat).

Three breeding scenarios were considered in this study, differing in terms of phenotyping procedures and breeding value estimations. Under the first scenario, named **BLUP-AM** and representing a current traditional male line breeding scheme, the breeding value estimation was based on a conventional BLUP-Animal Model (Quaas and Pollak, 1980), the phenotypes for Trait 1 being recorded on the candidates, and the phenotypes for Trait 2 being recorded on a small number of relatives. Under the second scenario, named **GE-1TP**, the candidates were not phenotyped and selection decisions were based on

genomic estimated breeding values (**GEV**; Meuwissen et al., 2001), which were estimated using a single TP that was genotyped and phenotyped for both traits and only comprised relatives of the candidates. Under the third scenario, named **GE-2TP**, Trait 1 was recorded on the candidates and Trait 2 was recorded on their relatives as in the BLUP-AM scenario, but the former and latter animals constituted 2 distinct TP to estimate the GEV for Trait 1 (with a large TP) and for Trait 2 (with a small TP), respectively.

Each replicate of the simulation consisted in 2 successive main steps: i) creation of a base population before selection presenting a linkage disequilibrium (**LD**) of a desired level and ii) simulation of the genetic evolution of the population under various phenotyping and selection procedures.

Generation of the Base Population

At the beginning of each replicate, a base population of 400 males and 400 females was created. Each animal had 10 pairs of 100-cM chromosomes, each of them carrying 3,600 equally spaced biallelic loci that might become SNP or QTL thereafter. In Generation 0, the genotype of each animal for each locus was randomly sampled from a Bernoulli distribution with frequency of 0.5. Recombination was modeled by sampling in a binomial function the number of chromosomal intervals (constituted by the segment between 2 adjacent loci) where crossing over occurred during gametogenesis and then sampling the locations of these intervals under a uniform distribution. No mutation process was implemented.

The 400 females were randomly mated to one of the 400 males for 350 generations, each female producing 1 male offspring and 1 female offspring, and the new generation completely replacing the previous one. After 350 generations, a bottleneck was created by reducing the population size to 45 males and 45 females. The population was randomly mated for the next 35 generations with complete renewal at each generation and was then extended to 1,200 females and 400 males for the next 3 generations.

The number of random generations simulated, and the evolution of the size of the base population over generations of mating were determined in order to produce a population presenting a short- and long-range LD structure comparable to the LD structure actually observed in French pig populations (Figure 1; S. Boitard, INRA, Toulouse, France, personal communication).

In the last generation, 1,500 loci with a minor allele frequency above 0.05 were randomly sampled on each chromosome to be SNP markers that would be used for subsequent genomic evaluation procedures. Consequently, the SNP were unevenly spaced, the

minimum distance between 2 adjacent markers being 28 kb. The average distance was 67 kb, corresponding to the average distance between adjacent markers on the *porcineSNP60* Illumina BeadChip (Illumina, Inc.) after the usual elimination of noninformative markers and markers with technical problems. Among the remaining loci with a minor allele frequency above 0.05, 2 series of n loci were randomly sampled on each chromosome to be QTL, one series being QTL for Trait 1, the other series being QTL for Trait 2. Values of n were 10, 30, or 60, depending on the QTL density considered in the simulations. To summarize, each of the 10 pairs of 100-cM chromosomes carried 1,500 unevenly spaced SNP, n QTL for Trait 1 and n QTL for Trait 2.

By convention, the allele 1 of each QTL had no effect. A gamma distribution with shape and scale parameters equal to 0.4 and 1/1.66 (Hayes and Goddard, 2001) was used to sample the absolute value of effects of allele 2 of the QTL. The sign of the effect of allele 2 was positive or negative, with a probability of 0.5. For both traits, the effects of the QTL were rescaled to result in a genetic variance of 1. In the following, the true breeding value (TBV) TBV_{t_i} of animal i for trait t ($t = 1$ or 2) was calculated as

$$TBV_{t_i} = \sum_{j=1}^{10} \left(\sum_{k=1}^n g_{t_{jki}} \times q_{t_{jk}} \right),$$

in which n is the number of QTL carried by each of the 10 chromosomes for each trait, $g_{t_{jki}}$ is the genotype (i.e., 0, 1, or 2 copies of allele 2) of animal i for the k th QTL for trait t on chromosome j , and $q_{t_{jk}}$ is the effect of the k th QTL for trait t on chromosome j .

Breeding Scenarios

The second step in each replicate of the simulation started with the random sampling of 1,050 females and 50 males from the 1,200 females and 400 males of the last generation of the base population before selection to constitute the founders of the selected population. The 1,050

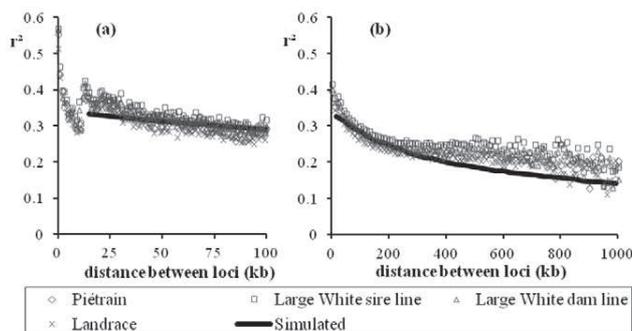


Figure 1. Linkage disequilibrium (r^2) observed in the 4 collective French pig breeds (S. Boitard, personal communication) and simulated in the present study for 0- to 100-kb intervals (a) and 0- to 1,000-kb intervals (b), assuming 1 Mb/cM.

females were randomly assigned to 5 herds of equal size, and within each herd into 7 batches of 30 females. To boot the system, each of the 1,050 females was mated randomly to one of the 50 boars at the beginning of the replicate.

The demographic and breeding characteristics of the simulated population are summarized in Figure 2, and were modeled on the average statistics of the French Piétrain nucleus population, established from the database used for national genetic evaluations.

The time step (TS) in our simulations was 3 wk, which corresponds to the period elapsing between the farrowing of females from 2 consecutive batches. Consequently, 1 yr of simulation represented 18 TS, and a sow farrowed every 7 TS until culling. At each TS, several tasks were performed within each herd.

Farrowing of the Batch(0) Sows and Birth of the Piglets. The 30 sows of the current batch, Batch(0), farrowed. Three litters from each herd (i.e., 10% of the litters born) were randomly sampled and left out to simulate the various unexpected events that can cause some litters to be unsuitable for genetic purposes (e.g., abortion or misidentification of the mated boar). All of the 90% remaining litters (i.e., 27 litters per herd) contained 3 female candidates. Among these 27 litters, the best 90% were selected according to the mean EBV of their parents (i.e., 24 litters per herd) and each contained 3 male candidates. Each male and female candidate i received a simulated record y_{1_i} for Trait 1 available 8 TS after its birth (i.e., when the animal is 5.5 mo old, corresponding to the end of the on-farm test period when selection decisions are made), simulated as $y_{1_i} = TBV_{1_i} + BL_{1_i} + e_{1_i}$, in which BL_{1_i} is the random environmental effect of the z th litter in which animal i is born, sampled from a normal distribution with variance

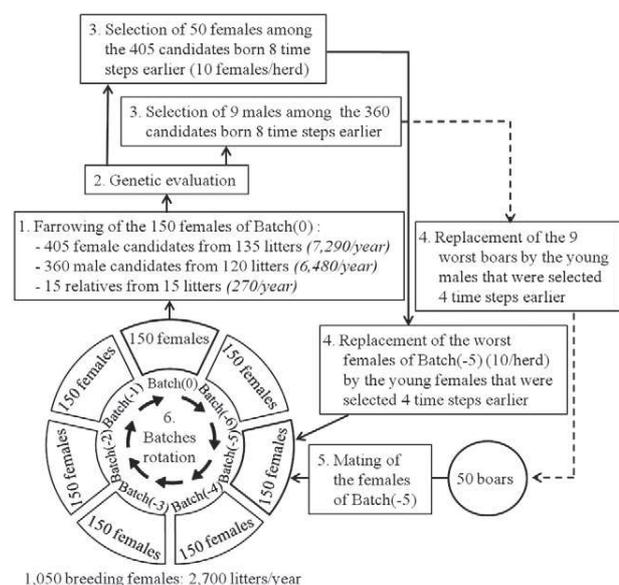


Figure 2. Population structure of the simulated breeding scheme and chronological order of events occurring at each simulated time step.

$\sigma_{BL}^2 = 0.2$, e_{1i} is a random residual sampled from a normal distribution with variance $\sigma_{e_1}^2 = (1/h_1^2) - 1 - \sigma_{BL}^2$, h_1^2 being the heritability of Trait 1, and assuming a genetic variance $\sigma_{a_1}^2$ of 1.

Ten percent of the litters born (i.e., a total of 15 litters per batch in all 5 herds) also contained 1 piglet destined to be phenotyped for Traits 1 and 2 (these animals are referred to as “relatives” below). These 10% litters were chosen from the litters suitable for genetic purposes by selecting the best litters as a function of their parental EBV, giving priority to the litters of boars with the fewest offspring among “relatives” and balancing the number of litters chosen in each herd. As for the candidates, the simulated records y_{1i} and y_{2i} for Traits 1 and 2 were available 8 TS after the birth of “relative” i , and y_{2i} was simulated as $y_{2i} = TBV_{2i} + e_{2i}$, in which e_{2i} is a random residual sampled from a normal distribution with variance $\sigma_{e_2}^2 = (1 - h_2^2) / h_2^2$, h_2^2 being the heritability of Trait 2 and assuming a genetic variance $\sigma_{a_2}^2$ of 1.

Breeding Value Estimation. Starting from the 26th TS (TS26; to allow time to accumulate sufficient phenotypes for Trait 2), breeding values for Trait 1 and Trait 2 (BLUP_EBV₁ and BLUP_EBV₂) were estimated for the complete population at each TS using a standard BLUP-Animal Model procedure and the BLUPF90 software (<http://nce.ads.uga.edu/~ignacy/newprograms.html>), considering all the phenotypes and pedigrees recorded since creation of the herds (the most recent available phenotypes being recorded on candidates and relatives born 8 TS earlier). The genetic parameters were the same as those used to simulate the data, and the 2 traits were assumed to be genetically independent. For each animal i , the BLUP_EBV for Traits 1 and 2 were then combined in a global EBV (GL_EBV) giving equal weight to both traits as

$$GL_EBV_i = 1 \times BLUP_EBV_{1i} + 1 \times BLUP_EBV_{2i}.$$

Selection Among Candidates. In each herd, the 10 best female candidates as a function of their GL_EBV were selected from the candidates in the herd available to renew the selection nucleus, after randomly eliminating 20% of the 81 candidates to simulate the uncontrolled culling events that can occur in real life and reduce selection efficiency (e.g., disease or leg problems). Similarly, the 9 best male candidates from the 360 candidates in the 5 herds were selected (a 20% random loss of candidates was also applied) to renew the pool of boars. These selected animals would be available for reproduction 4 TS later. No attention was paid to relationships during the selection procedure.

Culling of Active Boars and Sows. In each herd, the worst third of the 30 sows in batch Batch(-5), according

to their GL_EBV, were culled and replaced by the 10 young females that had been selected 4 TS earlier. Similarly, the 9 worst active boars were replaced by the 9 young males that had been selected 4 TS earlier.

Mating. In each herd, the 30 females in batch Batch(-5) were randomly mated to 1 of the 50 available boars. Kinship relationships between males and females were not taken into account, which could thus lead to sib or half-sib matings.

Each TS ended with moving the active sows from their current batch to the next batch to prepare the next TS. Each task was repeated for 89 TS (i.e., 5 yr of BLUP-AM selection after creation of the herds).

At TS90, the 3 scenarios diverged and were run for 180 additional TS (i.e., 10 yr). Under the BLUP-AM scenario, the above procedure was executed 180 more times (i.e., 10 more yr). The 2 genomic scenarios considered in the study only differed from the BLUP-AM scenario in terms of the procedure used to estimate breeding values. The number of candidates, phenotyping capacity for Trait 2, and selection rates among candidates and breeding males and females were the same under all 3 scenarios.

Simulation of Training Populations. The accuracy of GS can be expressed as (Daetwyler et al., 2008; Meuwissen, 2009)

$$r = \left[nh^2 / (nh^2 + 4N_eLv) \right]^{1/2}, \quad [1]$$

in which n is the number of individuals in the TP, h^2 is the heritability of the trait, L is the genome size (in Morgan), N_e is the effective size of the population, and $4N_eLv$ is the effective number of QTL loci in the genome. Consequently, considering constant SNP and QTL densities in all chromosomes of the genome, the number of phenotypes in the TP required to achieve the same accuracy of predicting breeding values when simulating a K Morgan genome rather than a 30 Morgan true porcine genome must be $30/K$ times smaller (Meuwissen, 2009). Because the genome simulated in this study was approximately one-third the actual pig genome (10 Morgan instead of 30 Morgan), the same ratio (1:3) was applied when simulating the TP to produce GEBV with appropriate accuracies.

To illustrate the relevance of Eq. [1], independently from the main study we simulated 5 yr of the BLUP-AM scenario (TS1 to TS89) followed by 3 yr of the GE-2TP scenario (TS90 to TS143), using the same population parameters as described above. Three situations were considered: i) 5 pairs of 100-cM chromosomes applying a scaling factor of 1/6 (Sit1a), ii) 10 pairs of 100-cM chromosomes applying a scaling factor of 1/3 (Sit2), and iii) 5 pairs of 100-cM chromosomes applying an inadequate scaling factor of 1/3 (Sit1b). Because the only

concern regarding this illustration of Eq. [1] was the effect of the scaling factor on the accuracy of genomic evaluation, culling and replacements were random between TS1 and TS143, so that genetic variance was maintained as being equal to 1 on average for both traits in the 3 alternative situations. The results were averaged over 30 replicates.

GE-1TP Scenario. Under the GE-1TP scenario, male and female candidates were no longer phenotyped for Trait 1 after TS90. After that time, all candidates and breeding animals were genotyped, and all selection and culling decisions were based on GEBV, estimated as follows. At TS90, the effect of each of the 15,000 SNP on Trait 1 and Trait 2 was estimated using the BLUP methodology described by Meuwissen et al. (2001) on a TP of 1,000 animals, genotyped for the 15,000 SNP and phenotyped for Traits 1 and 2. This TP was made up of relatives of the candidates born between TS16 and TS82; in practice, according to the scaling argument described above, one-third of the relatives were randomly sampled at each TS to constitute the TP. The estimated SNP effects were used during the next 18 TS (i.e., 1 yr) to calculate the genomic GEBV_{t_i} of the breeding animals and candidates for Trait *t* as

$$\text{GEBV}_{t_i} = \sum_{j=1}^{10} \left(\sum_{k=1}^{1500} g_{t_{jki}} \times m_{t_{jk}} \right),$$

in which $g_{t_{jki}}$ is the genotype (i.e., number of copies of allele 2) of animal *i* for the *k*th SNP on chromosome *j* and $m_{t_{jk}}$ is the estimated effect of the *k*th SNP on chromosome *j* for Trait *t*. As under the BLUP-AM scenario, GEBV for Trait 1 and 2 were combined as $\text{GL_GEBV}_i = 1 \times \text{GEBV}_{1_i} + 1 \times \text{GEBV}_{2_i}$, and all selection decisions were based on the GL_GEBV_i . Every year (i.e., in TS108, TS126, TS144, TS162, TS180, TS198, TS216, TS234, and TS252), the effects of the SNP on the traits were estimated again on the previous TP augmented by the 270 new animals (i.e., 90 simulated animals) phenotyped during the past year, and these new estimates were then used to calculate the GEBV_{t_i} of breeding animals and candidates for the next 18 TS.

GE-2TP Scenario. Under the GE-2TP scenario as with GE-1TP, all selection and culling decisions were based on GEBV, requiring all animals to be genotyped. The estimation procedure for GEBV for Trait 2 was the same as under the GE-1TP scenario, with a TP of 1,000 at TS90, increasing by 270 animals annually. Unlike the GE-1TP scenario, the male and female candidates were still phenotyped for Trait 1 after TS90, and consequently could be used to constitute and increase a TP for Trait 1. As for Trait 2, an initial estimate of the SNP effects on Trait 1 was made using the BLUP methodology at TS90 with a TP of 13,770, corresponding to the male and female candidates phenotyped during the fifth yr of the simulation (TS73 to TS90); at each TS, one-third of the candidates

were randomly sampled to constitute the TP. The TP for Trait 1 increased by 13,770 animals every year (i.e., 4,590 simulated animals) until TS144; after that, the oldest 13,770 animals were removed each year and replaced by the 13,770 most recent ones, so as to maintain the TP at a constant size of 55,080 animals (i.e., 4 yr of candidate testing) and keep the computation time at a reasonable level.

For each scenario, 3 couples of medium and low heritabilities for Trait 1 and Trait 2 (0.4 and 0.4, 0.4 and 0.2, and 0.2 and 0.2 referred to, respectively, as **MMh**², **MLh**², and **LLh**² below) and 3 QTL densities (10 QTL/Morgan per trait, 30 QTL/Morgan per trait, or 60 QTL/Morgan per trait, which are referred to, respectively, as **LowQd**, **MedQd**, and **HighQd**) were considered at TS0. Nine parameter cases were thus simulated under each scenario.

For each TS, the mean and variance of the TBV_{1_i} , TBV_{2_i} and $\text{GL_TBV}_i = \text{TBV}_{1_i} + \text{TBV}_{2_i}$ values of candidates born in the TS considered as well as the mean of their inbreeding coefficients based on pedigree information were calculated to measure the genetic gain and the evolution of genetic variance and inbreeding over time. The correlation between the TBV and the BLUP or genomic EBV of candidates available for selection was also calculated at each TS to estimate the accuracy of the selection. All results were averaged over 200 replicates per simulated case.

RESULTS

Relevance of the Scaling Relation Between the Length of the Simulated Genome and the Size of the Training Population

Figure 3 presents the accuracy of the GEBV (i.e., correlation between TBV and GEBV) for the 765 candidates available for selection in each TS from TS90 to TS143, under the Sit1a, Sit2, and Sit1b situations. Overall, the time course of the accuracy of the GEBV for both traits under the 3 situations displayed similar profiles. The accuracy increased markedly every 18 TS (TS108 and TS128), and decreased slowly for the next 17 TS. The increases in accuracy corresponded to the yearly reestimation of the SNP effects considering a larger TP whereas the subsequent regular decreases reflected the progressive reduction in the degree of relationship between the population under evaluation and the TP. However, the consequences of the decreasing relationships between the candidates and the TP on the accuracy of the GEBV were small, and diminished from year to year. For example, the accuracy for Trait 1 decreased by 5.5, 3.8, and 3.3% during the first, second, and third years of selection, respectively. The presence of close relatives of the candidates in the TP became less and less important as TP size increased.

Table 1. Reduction in genetic variance (%) for Traits 1 and 2 under 3 scenarios during the last 10 yr simulated, according to the initial heritabilities of the traits and QTL densities¹

Initial QTL density	Scenario ²	Reduction in genetic variance for:					
		Trait 1, genetic variance unit			Trait 2, genetic variance unit		
		Initial heritabilities of traits ^{3,4}					
		0.4/0.4	0.4/0.2	0.2/0.2	0.4/0.4	0.4/0.2	0.2/0.2
10QTL/ Morgan	BLUP-AM	-79	-79	-75	-34	-22	-29
	GE-1TP	-64	-67	-59	-67	-53	-60
	GE-2TP	-84	-86	-83	-61	-47	-51
30QTL/ Morgan	BLUP-AM	-59	-61	-57	-22	-21	-23
	GE-1TP	-47	-51	-45	-53	-44	-49
	GE-2TP	-67	-70	-68	-44	-35	-36
60QTL/ Morgan	BLUP-AM	-48	-50	-46	-22	-18	-22
	GE-1TP	-41	-44	-41	-47	-41	-48
	GE-2TP	-57	-61	-58	-38	-31	-35

¹Results are averages of 200 replicates.

²Simulated scenarios: BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

³Initial heritability for Trait 1/initial heritability for Trait 2.

⁴For a given QTL density, all values within a column differ ($P < 0.001$).

The accuracies of selection in Sit1a and Sit2, applying to the TP size a scaling factor proportional to the fraction of the length of the simulated genome, were close as expected from Eq. [1]. Conversely, the results for Sit1b showed that use of an inappropriate scaling factor (in Sit1b, the scaling factor was 1/3 rather than 1/6) resulted in significantly different correlations ($P < 0.001$) between the GEBV and true breeding values of the candidates. As a consequence, ignoring this scaling relation when simulating a genome shorter than the actual genome of the considered species would yield an overestimate of the accuracy of genomic evaluation and of the efficiency of GS because this expected response would actually require a larger TP.

Comparison of the Efficiency of the Three Scenarios

Overall, the relative results for the 3 scenarios from TS90 to TS270 in terms of genetic gain, accuracy, genetic variability, and inbreeding yielded the same general trends in the 9 combinations of heritabilities and QTL densities considered. Consequently, detailed results over time are only presented below for the MLh²-HighQd parameter case, supplemented by tables of the time-averaged results per scenario for the 9 cases.

Genetic Variance. The genetic variance of the 2 traits decreased over time (Table 1) under all 3 scenarios. This decrease can be explained by the increase in frequency of favorable QTL alleles in the population due to selection. This decrease was even more rapid when the initial QTL density was less as shown in Figures 4a and 4b for the 60 QTL/Morgan and 10 QTL/Morgan initial densities, respectively. An initial loss of genetic variance, particularly

large for Trait 1, occurred during the first year after the implementation of selection (TS26 to TS44), its magnitude being consistent with the expected theoretical Bulmer effect (Dekkers, 1992) considering the applied selection intensities and the accuracy of the EBV for both traits. The decrease in genetic variance then continued gradually (from the slowest to the most rapid decrease: GE-1TP, BLUP-AM, and GE-2TP for Trait 1 and BLUP-AM, GE-2TP, and GE-1TP for Trait 2).

Accuracy. The time course of the accuracy of EBV displayed different profiles under the 3 scenarios as shown in Figure 5 for the MLh²-HighQd parameter case. The accuracies under the BLUP-AM scenario decreased steadily from TS90 to TS270, with the rate of this decrease being slow for Trait 2 but more rapid for Trait 1. This trend could be explained by the progressive decrease of genetic variance in the population: since residual variance was constant over time in the simulations, the proportion of the phenotypic variance of genetic origin decreased as did the ability to predict breeding values with a BLUP-AM evaluation. Because more data were available for Trait 1 than Trait 2 under the BLUP-AM scenario, selection was more efficient and the decrease in genetic variance and therefore in accuracy was more marked for Trait 1 (Figures 4 and 5).

Conversely, the accuracy of the genomic values estimated by considering the TP made up of relatives (i.e., Traits 1 and 2 under the GE-1TP scenario and Trait 2 under the GE-2TP scenario) increased globally over time in line with the size of the TP, but the rate of this increase gradually slowed down. The accuracy of the genomic values for Trait 1 under the GE-2TP scenario displayed a similar trend for as long as the number of candidates in the TP increased each year (i.e., until

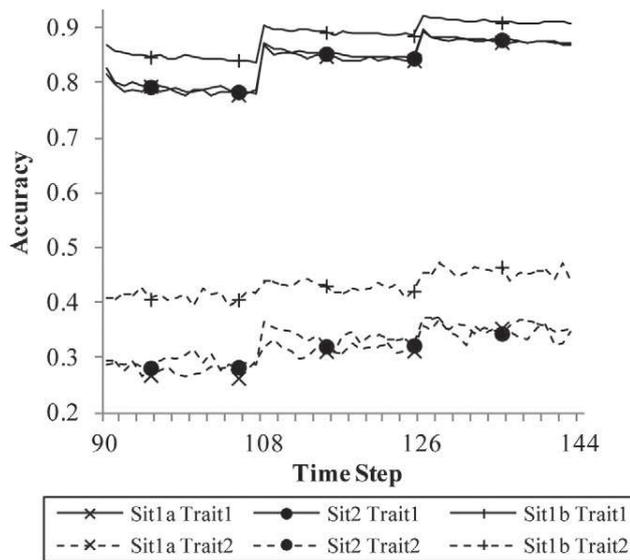


Figure 3. Evolution of the accuracy of genomic EBV of candidates available for selection at each time step, considering different lengths of simulated genome and different values for the scaling factor applied to the size of the training population. The initial QTL density was 30 QTL/M for each trait, and heritabilities of Trait 1 and Trait 2 were 0.4 and 0.2, respectively. The results are averages of 30 replicates: Sit1a, 5 pairs of 100-cM chromosomes applying a scaling factor of 1/6; Sit1b, 5 pairs of 100-cM chromosomes applying an inadequate scaling factor of 1/3; and Sit2, 10 pairs of 100-cM chromosomes applying a scaling factor of 1/3.

TS144), but decreased globally after that. A second annual pattern appeared alongside these general trends in all the cases; accuracies improved abruptly every year when SNP effects were reestimated on the updated TP and then decreased progressively for the next 17 TS.

The average accuracies from TS90 to TS270 are presented in Table 2 for the 3 scenarios. Both genomic scenarios produced significantly ($P < 0.001$) more accurate EBV for Trait 2 than the BLUP-AM scenario. Whereas the accuracy of the BLUP-AM EBV of candidates for selection was capped at 0.27 or 0.33 when the initial heritability of Trait 2 was 0.2 or 0.4, respectively, the accuracy of GEBV rose to 0.36 ($h^2 = 0.2$) or 0.46 ($h^2 = 0.4$) as early as TS90 and then increased yearly with the addition of new animals to the TP (Figure 5). On average, between TS90 and TS270, the accuracy of GEBV for Trait 2 was about 50% greater than the accuracy of BLUP-AM EBV.

The 2 genomic scenarios generated divergent results for Trait 1. As for Trait 2, the GE-2TP scenario provided more accurate EBV for Trait 1 ($P < 0.001$) when compared with the BLUP-AM scenario. Accuracy increased from approximately 0.30 to 0.50 (for h^2 of Trait 1 = 0.2) and from approximately 0.40 to 0.60 (for h^2 of Trait 1 = 0.4) at TS90 when switching from the BLUP-AM evaluation to the genomic evaluation. On average, between TS90 and TS270, the accuracies of EBV for Trait 1 were 35 to 43% greater under the GE-2TP scenario than under BLUP-AM. By contrast, the

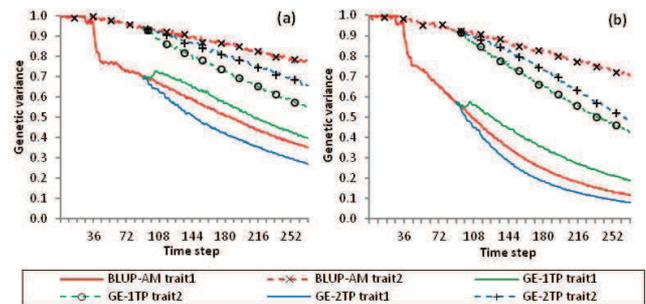


Figure 4. Evolution of genetic variance under 3 scenarios during the 15 yr simulated. Initial QTL density = 60 QTL/M (a) or 10 QTL/M (b) for each trait, initial heritability of Trait 1 = 0.4, and initial heritability of Trait 2 = 0.2. The results are averages of 200 replicates. BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

accuracies of the EBV for Trait 1 were less ($P < 0.001$) under the GE-1TP scenario than under BLUP-AM at TS90 (e.g., 0.40 vs. 0.64 in the ML h^2 -HighQd case) when TP only contained 1,000 animals, and remained less until the end of the simulation, despite TP size increasing every year by 270 animals (Figure 5).

Genetic Trends. The above differences in genetic variances and in EBV accuracies yielded differences in terms of genetic trends between the 3 scenarios being compared (Table 3). Both genomic scenarios produced significantly ($P < 0.001$) greater average annual genetic trends for Trait 2 than the BLUP-AM scenario, this superiority tending to be greater for ML h^2 than for MM h^2 cases (i.e., for a lesser heritability of Trait 2). The average annual genetic trends between TS90 and TS270 were 78 to 128% greater under GE-1TP than under BLUP-AM, and 63 to 84% greater under GE-2TP than under BLUP-AM. For example, in the ML h^2 -HighQd case, Trait 2 improved by +0.13, +0.30, and +0.24 genetic SD per year (σ_g/yr) under the BLUP-AM, GE-1TP, and GE-2TP scenarios, respectively (Figure 6).

The results of the 2 genomic scenarios differed for Trait 1. The annual genetic trends for this Trait were 27 to 44% less under the GE-1TP scenario than under BLUP-AM, this difference ($P < 0.001$) tending to be greater when the heritability of Trait 1 was less (ML h^2 case vs. LL h^2 case). For example, in the ML h^2 -HighQd case, the genetic level of the population for Trait 1 increased by +0.62 σ_g/yr under the reference scenario, but only by +0.42 σ_g/yr under the GE-1TP scenario (Figure 6). Conversely, the GE-2TP scenario produced 8 to 22% greater average annual genetic trends for Trait 1, this improvement being significant in all the cases considered ($P < 0.001$) and, as for Trait 2, was greater with a smaller heritability of Trait 1.

As a consequence, scenario GE-2TP was globally 27 to 33% more efficient than scenario BLUP-AM in improving the global breeding goal (Trait 1 + Trait 2). Despite the sampling variance of the results among the 200 replicates, the average TBV of the population at TS270 was greater under the GE-2TP scenario than under BLUP-AM in 97.8, 98.1, and 99.7% of the replicates for Trait 1, Trait 2, and the global breeding goal, respectively. By contrast, scenario GE-1TP was nearly as efficient as the reference scenario, its advantage for Trait 2 balancing out its poor result for Trait 1.

Inbreeding. Both genomic scenarios resulted in a smaller global increase in the average inbreeding coefficient of the population than the BLUP-AM scenario (Table 4). Compared with the latter, average inbreeding at the end of the simulation was 13 to 30% less under the GE-1TP scenario and 49 to 60% lower under GE-2TP, depending on the initial QTL density and initial heritabilities considered. Nevertheless, whereas the GE-2TP and BLUP-AM scenarios displayed linear time trends for this parameter, the increase in the inbreeding coefficient was much more marked under the GE-1TP scenario during the first time period, before slowing down and stabilizing to approximately the same trend as that seen with GE-2TP (Figure 7). Therefore, as early as 3 yr after the beginning of GS and up to the end of the simulated period, the inbreeding was about 30% less under the GE-2TP scenario than under GE-1TP.

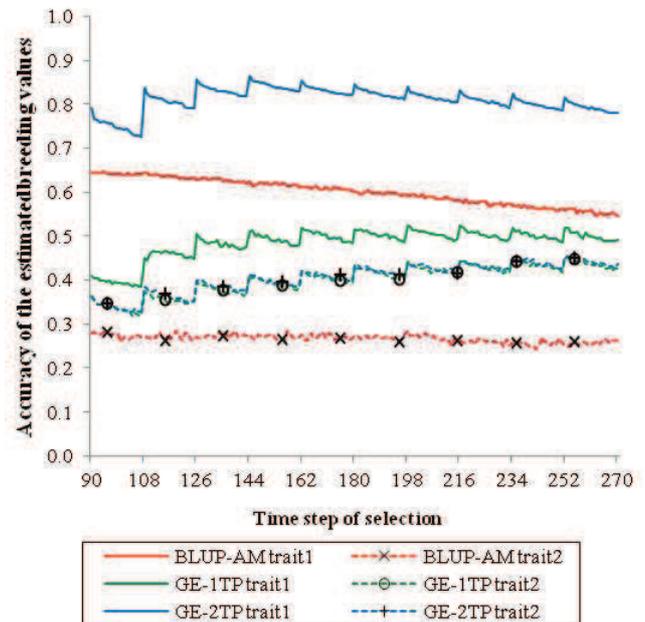


Figure 5. Accuracy of the EBV of candidates under 3 scenarios during the last 10 yr simulated, with an initial QTL density of 60 QTL/M for each trait, initial heritability of Trait 1 = 0.4, and initial heritability of Trait 2 = 0.2. The results are averages of 200 replicates. BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

Table 2. Average accuracy of EBV of young candidates under 3 scenarios during the last 10 yr of the simulated breeding schemes, according to the initial heritabilities of traits and QTL densities¹

Initial QTL density	Scenario ²	Average accuracy of EBV for:					
		Trait 1			Trait 2		
		Initial heritabilities of traits ^{3,4}					
		0.4/0.4	0.4/0.2	0.2/0.2	0.4/0.4	0.4/0.2	0.2/0.2
10QTL/ Morgan	BLUP-AM	0.51	0.50	0.45	0.31 ^a	0.26 ^a	0.26
	GE-1TP	0.42	0.40	0.34	0.46 ^b	0.37 ^b	0.37
	GE-2TP	0.73	0.70	0.64	0.45 ^b	0.37 ^b	0.38
30QTL/ Morgan	BLUP-AM	0.58	0.57	0.50	0.33 ^a	0.27 ^a	0.27 ^a
	GE-1TP	0.47	0.46	0.38	0.50 ^b	0.41 ^b	0.40 ^b
	GE-2TP	0.80	0.78	0.71	0.50 ^b	0.41 ^b	0.41 ^b
60QTL/ Morgan	BLUP-AM	0.61	0.60	0.52	0.33	0.27 ^a	0.27
	GE-1TP	0.49	0.48	0.40	0.51	0.40 ^b	0.41
	GE-2TP	0.82	0.81	0.74	0.51	0.41 ^b	0.42

¹Results are averages of 200 replicates.

²Simulated scenarios: BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

³Initial heritability for Trait 1/initial heritability for Trait 2.

⁴For a given QTL density, all values within a column differ ($P < 0.001$) except values with common superscripts.

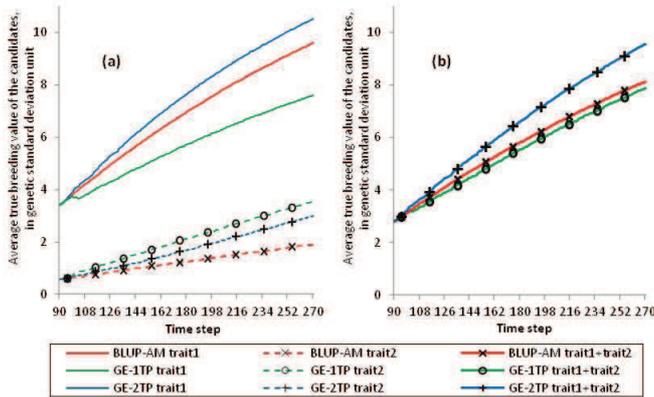


Figure 6. Realized genetic trends (in genetic SD units) for Traits 1 and 2 (a) and for the global breeding goal (Trait 1 + Trait 2) (b) under 3 scenarios during the last 10 yr simulated, with an initial QTL density of 60 QTL/M for each trait, initial heritability of Trait 1 = 0.4, and initial heritability of Trait 2 = 0.2. The results are averages of 200 replicates. BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

DISCUSSION

The representativeness of stochastic simulation results compared with the results of real breeding schemes is obviously dependent on the quality of modeling. Attention was paid in this study to simulating a population with a realistic structure already selected for several years when the 3 scenarios started to diverge, and the use of adequate short- and long-range LD structure and marker density. However, some other assumptions were made that might have departed from reality. For instance, because the actual number of genes

affecting a trait in a population is unknown, 3 arbitrary QTL densities were considered, but they nonetheless generated consistent results on the relative efficiencies of the 3 scenarios. Furthermore, the management of inbreeding was not considered in the selection and mating procedures whereas inbreeding is a major concern for pig breeders. Also, simplifications were made in the genomic evaluation procedures to keep the computation time reasonable. In particular, the yearly renewal of the TP and SNP effects reestimation yielded slightly smaller accuracies of GEBV than those that would have been obtained if the phenotypes of the candidates had been added to the TP at every TS. All these assumptions may have affected the results, which should therefore be considered as trends rather than precise predictions.

The results of this study tend to show that GS could substantially and durably increase the genetic gains achieved in a purebred pig male line based on the combined phenotyping of candidates and relatives for lowly to moderately heritable traits while significantly reducing the annual increase in inbreeding in the population without it being necessary to modify the structure of the current breeding scheme. In particular, although it may result in a small TP of limited power, a small annual phenotyping capacity for traits only recorded on relatives may be sufficient to implement a genomic scheme more efficient than the current breeding scheme based on BLUP-AM evaluation, the accuracy of which is poor for these traits. This situation is quite different from that of dairy cattle breeding schemes where a TP of several thousand bulls is needed to achieve the same high degree of accuracy as conventional progeny testing (Hayes et al., 2009). In the current study, the initial TP for Trait 2 under the GE-1TP and

Table 3. Average realized annual genetic trends under 3 scenarios during the last 10 yr simulated, according to the initial parameter values for the heritabilities of traits and QTL densities¹

Initial QTL density	Scenario ²	Average annual genetic trends, in genetic SD units								
		Trait 1			Trait 2			Global breeding goal		
		0.4/0.4	0.4/0.2	0.2/0.2	Initial heritabilities of traits ^{3,4}			0.4/0.4	0.4/0.2	0.2/0.2
10QTL/ Morgan	BLUP-AM	0.36	0.37	0.36	0.20	0.15 ^a	0.16	0.40	0.37	0.37 ^a
	GE-1TP	0.26	0.27	0.23	0.35	0.28 ^b	0.29	0.43	0.39	0.37 ^a
	GE-2TP	0.39	0.40	0.41	0.34	0.27 ^b	0.26	0.52	0.47	0.48 ^b
30QTL/ Morgan	BLUP-AM	0.52	0.53	0.49	0.20	0.14	0.14	0.50	0.47 ^a	0.45
	GE-1TP	0.33	0.36	0.27	0.40	0.30	0.32	0.52	0.46 ^a	0.42
	GE-2TP	0.57	0.60	0.58	0.35	0.25	0.25	0.65	0.60 ^b	0.58
60QTL/ Morgan	BLUP-AM	0.60	0.62	0.55	0.19	0.13	0.14	0.56 ^a	0.54	0.49
	GE-1TP	0.38	0.42	0.32	0.41	0.30	0.33	0.56 ^a	0.51	0.46
	GE-2TP	0.68	0.71	0.67	0.35	0.24	0.25	0.73 ^b	0.68	0.66

¹Results are averages of 200 replicates.

²Simulated scenarios: BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

³Initial heritability for Trait 1/initial heritability for Trait 2.

⁴For a given QTL density, all values within a column differ ($P < 0.001$) except values with common superscripts.

Table 4. Average annual increase in inbreeding among young candidates under 3 scenarios during the last 10 yr simulated, according to the initial heritabilities of traits and QTL densities¹

Initial QTL density	Scenario ²	Average annual increase in inbreeding coefficient, %		
		Initial heritabilities of traits ^{3,4}		
		0.4/0.4	0.4/0.2	0.2/0.2
10QTL/ Morgan	BLUP-AM	+0.82	+0.84	+1.17
	GE-1TP	+0.65	+0.65	+0.81
	GE-2TP	+0.40	+0.43	+0.50
30QTL/ Morgan	BLUP-AM	+0.90	+0.79	+1.10
	GE-1TP	+0.65	+0.68	+0.77
	GE-2TP	+0.39	+0.38	+0.45
60QTL/ Morgan	BLUP-AM	+0.88	+0.81	+1.17
	GE-1TP	+0.69	+0.67	+0.83
	GE-2TP	+0.40	+0.37	+0.47

¹Results are averages of 200 replicates.

²Simulated scenarios: BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

³Initial heritability for Trait 1/initial heritability for Trait 2.

⁴For a given QTL density, the 3 values within a column differ ($P < 0.001$).

GE-2TP scenarios contained 1,000 animals phenotyped during the 4 yr preceding the implementation of genomic evaluation, assuming that the DNA of these animals was available. In practice, pig breeding organizations have in recent years been building up DNA banks with the most important animals in their nucleus populations (including animals tested in central testing stations) to prepare for genomic research programs. Such an initial TP should therefore become available quite rapidly.

Conversely, for traits where a large quantity of phenotypes may be available and directly recorded on candidates, a breeding scheme based on the BLUP-AM methodology is already highly efficient. In the present study, the GE-2TP scenario only outperformed the BLUP-AM scenario by 8 to 22% with respect to genetic trends for Trait 1, despite the fact that the TP contained several tens of thousands of animals. By contrast, a breeding scheme replacing the phenotyping of candidates with a TP of limited size (as under the GE-1TP scenario) would clearly be less efficient for field data. To conclude, GS should not be substituted for the direct phenotyping of candidates when the latter is possible.

Under the present simulations, TP were updated only once a year by adding relatives (GE-1TP and GE-2TP scenarios) and candidates (GE-2TP) phenotyped during the current TS and previous 17 TS. Consequently, whereas candidates available for selection in the TS in which the SNP effects were estimated or reestimated had closely related animals in the TP (sibs or half sibs) or were

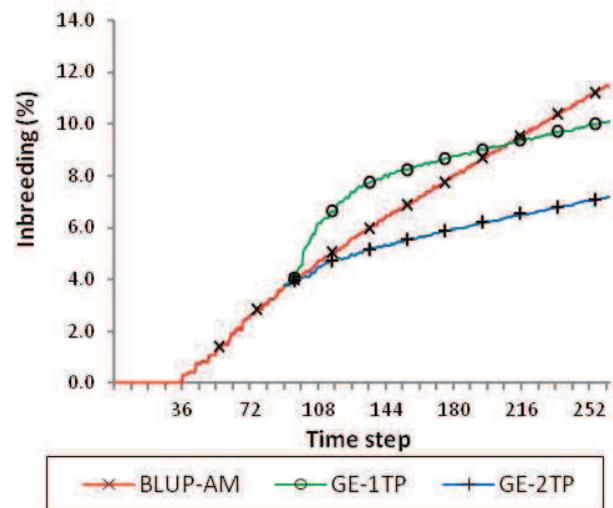


Figure 7. Evolution of inbreeding among young candidates under 3 scenarios during the 15 yr simulated. Initial QTL density = 60 QTL/M for each trait, initial heritability of Trait 1 = 0.4, and initial heritability of Trait 2 = 0.2. The results are averages of 200 replicates. BLUP-AM = breeding scheme based on traditional BLUP-Animal Model genetic evaluations using phenotypes of the candidates for Trait 1 and phenotypes of relatives for Trait 2; GE-1TP = breeding scheme based on genomic evaluations using a single training population composed of relatives; GE-2TP = breeding scheme based on genomic evaluations using a training population composed of candidates for Trait 1 and a training population composed of relatives for Trait 2.

even themselves part of the TP (for Trait 1 under the GE-2TP scenario), candidates in the next 17 TS were less and less closely related to this TP because of the continuous renewal of breeding animals. The progressive annual decrease in the accuracy of GEBV observed between 2 consecutive reestimates of the SNP was therefore consistent with the findings of several other studies (e.g., Solberg et al., 2009; Habier et al., 2010; Wolc et al., 2011), which showed that GEBV precision decreases in line with the degree of relationship between the TP and the population under evaluation, thus confirming the importance of continuously renewing the TP.

However, the reason for the global slowdown in the increase of GEBV accuracy and for the decrease after TS144 for Trait 1 under the GE-2TP scenario was different. As shown in Eq. [1], the accuracy of GEBV depends on the heritability of the trait, which decreased over time during the simulations because of the increase in the incidence of favorable QTL alleles, thus reducing the genetic variance of the traits. Consequently, the impact of adding new animals to the TP became weaker, and the rate of increase in accuracy gradually decelerated. In the case of Trait 1 under the GE-2TP scenario, where the TP size remained constant after TS144, renewal of the TP by replacing the 13,770 oldest animals with 13,770 more recent ones did not compensate for the loss of power due to less heritability, and the accuracy of the GEBV for this trait eventually decreased.

The observed decrease over time in the genetic variance for both traits was actually very marked, whichever methodology was used to estimate breeding value. For example, only 27 and 35% of the initial variance for Trait 1 remained under the GE-2TP and BLUP-AM scenarios after 15 yr of selection in the MLh²-HighQd case, these values decreasing to 8 and 12%, respectively, in the MLh²-LowQd case. The overall decrease in the genetic variance was less for Trait 2, because the EBV were globally less accurate than for Trait 1, resulting in a smaller Bulmer effect and a less efficient fixation of the favorable QTL alleles for Trait 2.

Our study helps alleviate the paucity of literature with respect to the evolution of genetic variance in GS breeding schemes. A major decrease in the genetic variance and accuracy of the EBV was reported by Jannink (2010) in a stochastic simulation of GS in barley (*Hordeum vulgare*). Nielsen et al. (2011) also observed a great decrease in genetic variance after 10 generations of stochastic simulation of a fish breeding scheme based either on BLUP or genomic evaluations. However, the decrease in genetic variance was less than in the present study because of lower selection rates, larger initial number of QTL and inbreeding management to restrain the inbreeding rate per generation to 0.5 or 1%.

The evolution of genetic variance observed here during the early years after implementation of selection was consistent with what could be expected according to the Bulmer effect (Dekkers, 1992). Nevertheless, the magnitude of the long-term decrease exceeded the theoretical predictions based on the infinitesimal polygenic model. Moreover, this loss of genetic variability constituted a deviation from what might be observed in true populations selected for tens of years, where genetic diversity is still not exhausted. Nevertheless, we simulated a finite number of QTL at the beginning of the simulation, and as the number of segregating QTL decreased because of selection, our model diverged more and more from the polygenic infinitesimal model underlying the theoretical formulas.

No mutation process was implemented in our simulations. However, mutations are not expected to impact the results within the first 30 to 40 generations of selection (Hill, 1985) whereas our horizon of interest was limited to 15 yr of pig breeding (equivalent to about 10 generations of selection). As a result, the most likely hypothesis is that the large reduction in variance observed here was due to the very high intensities of selection performed on male and female pathways, and to the finite number of QTL segregating at the beginning of the simulation. Considering a much larger number of QTL initially segregating in the population as did Lillehammer et al. (2011) and Nielsen et al. (2011) who simulated 1,800 and 1,000 QTL, respectively, might be

a solution to preserve genetic variability. A dynamic modeling of gene effects, taking account of interactions between loci, might pave the way to exploring the preservation of genetic variability. Finally, considering the relationships at selection and mating steps, or even adding a constraint on the inbreeding rate as made by Nielsen et al. (2011), could also slow the decrease in genetic variance.

Our results concerning the GE-2TP and BLUP-AM scenarios confirm the theoretical expectations of Daetwyler et al. (2007) that GS should reduce the rate of inbreeding per generation when compared with selection based on traditional BLUP-AM EBV. The breeding value of an individual can be conceived as the combination of the breeding values of its parents and a Mendelian sampling term. The Mendelian sampling term is poorly estimated with the BLUP-AM methodology, the phenotypes of ancestors and relatives mostly contributing to estimating the parental component of a breeding value. In particular, full-sib candidates obtain exactly the same BLUP-AM EBV for traits exclusively recorded on relatives (i.e., Trait 2 in the present study). Consequently, the BLUP-AM scenario favored the co-selection of related animals among candidates. By contrast, under the GE-2TP scenario, the genotypes at markers enabled a more precise estimate of the Mendelian component of breeding values for Traits 1 and 2, resulting in fewer correlated EBV among sibs and more accurate EBV for each of them.

The sharp initial rise in the inbreeding rate under the GE-1TP scenario might have appeared inconsistent at first sight. Nevertheless, as shown by Habier et al. (2010), the GEBV accuracy of an animal is strongly dependent on its relationship with the TP. When genomic evaluation was implemented at TS90 under the GE-1TP scenario, the TP was limited to 1,000 animals sampled from only 10% of the litters born during the 4 previous years, which resulted in a considerably smaller average accuracy for Trait 1, when compared with the BLUP-AM scenario where all candidates were phenotyped, and compared with the GE-2TP scenario where all candidates were closely related to the TP for Trait 1. One can therefore hypothesize that the candidates belonging to the few families most closely related to the TP had a greater chance of being selected, because their more accurate GEBV were less regressed toward the mean. This would explain the greater initial inbreeding rate. Thereafter, as the TP size and the average relationship between the candidates and TP increased, the co-selection of related candidates decreased and the increase in inbreeding slowed. Further work, however, has to be done to better understand this phenomenon.

It may seem inconsistent that the GE-2TP scenario produced the smallest increase in the average inbreeding

rate while at the same time causing the most rapid reduction in genetic variance. In fact, the GE-2TP scenario produced the greatest genetic trends because of the strong accuracy of its EBV, resulting in more rapid fixing of favorable alleles of the QTL. As mentioned above, the co-selection of related animals is small in a highly accurate genomic breeding scheme. Consequently, a larger proportion of the homozygosity observed in the population results from selection of identity-by-state loci in GE-2TP compared with BLUP-AM and GE-1TP. Therefore, the pedigree-based inbreeding measurement, which only takes into account the identity-by-descent part of the population homozygosity, does not fully reflect the genetic diversity of the population.

No attention was paid in the simulations to inbreeding management in the selection and mating procedures whereas limitation of the inbreeding rate is an important concern for pig breeders. As mentioned by Lillehammer et al. (2011), adding a constraint on this parameter would probably have reduced the inbreeding rate, but might also have affected genetic trends under the BLUP-AM scenario. The superiority of GE-2TP over BLUP-AM would therefore be even greater with respect to genetic gains in the context of a constant inbreeding rate across the scenarios.

In light of these results, one might ask if it would be economically profitable to move from a classical pig male line breeding scheme to a genomic breeding scheme organized as under the GE-2TP scenario. As an initial response to this question, one may consider a very simple breeding goal combining ADG and feed efficiency, representing Traits 1 and 2 of the simulations, respectively. The GE-2TP scenario resulted in additional genetic gains of $+0.09 \sigma_g/\text{yr}$ for Trait 1 and $+0.11 \sigma_g/\text{yr}$ for Trait 2 when compared with the BLUP-AM scenario (HighQd \times MLh² case). Considering that half of the genetic improvement achieved in the selection nucleus is transmitted to the 15,000,000 pigs finished in France each year and bred by a purebred Piétrain sire, the estimated increase in profit in production herds due to GS would be approximately €1,300,000/yr. Assuming that the costs induced by creating the 2 initial TP would be written off within 4 yr, the cost of genotyping should not exceed approximately €140 per animal, so that the extra profits due to increased genetic gains in production would compensate for the annual genotyping costs of 13,770 new candidates and 270 animals increasing the TP for Trait 2. The break-even cost, however, would depend on the size of the nucleus population and on its extent at the production level. For example, the break-even cost of genotyping in the same terminal male line with only 10,000,000 finished offspring per year would be approximately €100 per animal.

A comprehensive economic study considering a more realistic breeding goal and appropriate economic weights is obviously needed to reach a firm conclusion, but this rough estimate suggests that the implementation of genomic evaluation in a pig male line is economically feasible. The porcine 60K BeadChip is currently expensive, but the price will almost certainly decrease in the future as happened in the case of cattle in recent years. Another way to reduce costs would be to genotype candidates with a less expensive lower density array and use genotype imputation techniques to preserve the efficiency of genomic evaluation, following the example of dairy cattle (Dassonneville et al., 2011; Wiggans et al., 2012).

Finally, it would also be useful to quantify the increase in genetic gain that might be achieved if the additional cost of implementing genomic evaluation were invested in improving the current BLUP-AM breeding scheme, for example, by increasing phenotyping capacity for relatives. A constant cost comparison of the results of alternative breeding schemes would aid decision making regarding the economic value of GS.

LITERATURE CITED

- Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. *J. Anim. Br. Genet.* 124:369–376.
- Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *Plos One* 3:1–8.
- Dassonneville, R., R. F. Brøndum, T. Druet, S. Fritz, F. Guillaume, B. Gulstrandtsen, M. S. Lund, V. Ducrocq, and G. Su. 2011. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *J. Dairy Sci.* 94:3679–3686.
- Dekkers, J. C. M. 1992. Asymptotic response to selection on best linear unbiased predictors of breeding values. *Anim. Prod.* 54:351–360.
- Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genet. Sel. Evol.* 42:5.
- Hayes, B., and M. E. Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* 33:209–229.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443.
- Hill, W.G. 1985. Effects of population size on response to short and long term selection. *J. Anim. Br. Genet.* 102:161–173.
- Jannink, J. L. 2010. Dynamics of long-term genomic selection. *Genet. Sel. Evol.* 42:35.
- Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011. Genomic selection for maternal traits in pigs. *J. Anim. Sci.* 89:3908–3916.
- Meuwissen, T. H. E. 2009. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41:35.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.

- Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *J. Anim. Breed. Genet.* 124:342–355.
- Nielsen, H. M., Sonesson A. K., and T. H. E. Meuwissen. 2011. Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *J. Anim. Sci.* 89:630–638.
- Quaas, R. L., and E. J. Pollak. 1980. Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.* 51:1277–1287.
- Solberg, T. R., A. K. Sonesson, J. A. Woolliams, J. Degard, and T. H. E. Meuwissen. 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genet. Sel. Evol.* 41:53.
- Wiggans, G.R., T.A. Cooper, P.M. VanRaden, K.M. Olson, and M.E. Tooker. 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. *J. Dairy Sci.* 95:1552–1558.
- Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, R. Preisinger, D. Habier, R. Fernando, D. J. Garrick, and J. C. M. Dekkers. 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genet. Sel. Evol.* 43:23.

PARTIE 4.

Article 3.

Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme.

Tribout, T., C. Larzul, and F. Phocas. 2013. *Genetics, Selection, Evolution* **45**:40.

RESEARCH

Open Access

Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme

Thierry Tribout^{1,2*}, Catherine Larzul^{1,2} and Florence Phocas^{1,2}

Abstract

Background: Replacing pedigree-based BLUP evaluations by genomic evaluations in pig breeding schemes can result in greater selection accuracy and genetic gains, especially for traits with limited phenotypes. However, this methodological change would generate additional costs. The objective of this study was to determine whether additional expenditures would be more profitably devoted to implementing genomic evaluations or to increasing phenotyping capacity while retaining traditional evaluations.

Methods: Stochastic simulation was used to simulate a population with 1050 breeding females and 50 boars that was selected for 10 years for a breeding goal with two uncorrelated traits with heritabilities of 0.4. The reference breeding scheme was based on phenotyping 13 770 candidates per year for trait 1 and 270 sibs of candidates per year for trait 2, with selection based on pedigree-based BLUP estimated breeding values. Increased expenditures were allocated to either increasing the phenotyping capacity for trait 2 while maintaining traditional evaluations, or to implementing genomic selection. The genomic scheme was based on two training populations: one for trait 2, consisting of phenotyped sibs of the candidates whose number increased from 1000 to 3430 over time, and one for trait 1, consisting of the selection candidates. Several genomic scenarios were tested, where the size of the training population for trait 1, and the number of genotyped candidates pre-selected based on their parental estimated breeding value, varied.

Results: Both approaches resulted in higher genetic trends for the population breeding goal and lower rates of inbreeding compared to the reference scheme. However, even a very marked increase in phenotyping capacity for trait 2 could not match improvements achieved with genomic selection when the number of genotyped candidates was large. Genotyping just a limited number of pre-selected candidates significantly reduced the extra costs, while preserving most of the benefits in terms of genetic trends and inbreeding. Implementing genomic evaluations was the most efficient approach when major expenditure was possible, whereas increasing phenotypes was preferable when limited resources were available.

Conclusions: Economic decisions on implementing genomic evaluations in a pig nucleus population must take account of population characteristics, phenotyping and genotyping costs, and available funds.

Background

In recent years, genomic selection (GS) [1] has been implemented with success in dairy cattle [2], which has made it possible to stop lengthy and costly progeny testing in this species. Compared with traditional selection schemes, GS in dairy cattle enables earlier selection from a larger number of candidates, while maintaining (for

males) or even increasing (for females) the accuracy of their estimated breeding values (EBV). The resulting shorter generation intervals and higher selection intensities are expected to result in markedly higher annual genetic trends (e.g. [3]). Moreover, the savings achieved from not rearing non-productive males while awaiting the phenotype of their daughters under the progeny testing system provides the funds for genotyping. Consequently, switching from traditional to more efficient genomic schemes has not generated much, if any, additional costs in dairy cattle.

In pigs, with the availability of the porcineSNP60 Illumina beadchip (Illumina, Inc., San Diego, CA) and the extent of

* Correspondence: thierry.tribout@jouy.inra.fr

¹INRA, UMR1313 Génétique Animale et Biologie Intégrative, F-78350, Jouy-en-Josas, France

²AgroParisTech, UMR1313 Génétique Animale et Biologie Intégrative, F-75231, Paris, France

linkage disequilibrium (LD) observed in the nucleus populations [4,5], it is now possible to consider the implementation of GS. Current pig breeding schemes are, however, already characterized by high selection intensities and very short generation intervals. The impact of GS on these two parameters is therefore expected to be small, in contrast to the situation in dairy cattle. The accuracy of EBV is, nevertheless, generally poor in pigs, especially for late-recorded sex-limited traits and traits that cannot be measured on candidates (e.g. meat quality) or that are too expensive to measure on a large number of animals (e.g. feed efficiency). In this context, genomic evaluations can produce more accurate EBV than the current pedigree-based BLUP (best linear unbiased prediction)-animal model (pBLUP) evaluations and increase the efficiency of breeding schemes, as reported by Lillehammer et al. [6] in dam lines for traits that are only recorded on females. For sire lines, Tribout et al. [7] estimated that replacing pBLUP evaluations by genomic evaluations in a breeding scheme based on the combined phenotyping of candidates and a limited number of sibs of the candidates could increase the annual genetic trend for the population breeding goal by approximately 30% through greater accuracy, while substantially reducing the rate of inbreeding.

Most current pig breeding schemes are based on on-farm phenotyping of candidates for a few, easy to record growth and fattening traits, and on phenotyping of a limited number of sibs or candidates in testing stations for traits that are expensive or difficult to measure. Unlike in dairy cattle, GS in pigs would not result in organisational changes to the schemes that could generate sufficient savings to compensate for the cost of genotyping. First, selection among candidates usually occurs at the end of the fattening period (at about 160 days of age), which means that there is currently no financial cost associated with maintaining selection candidates during an unproductive period of time, as opposed to a breeding scheme based on progeny testing. Moreover, even if genomic evaluations no longer require candidates to be tested on-farm, the current phenotyping of candidates (mostly weighing and ultrasonic depth measurements) is much less expensive than genotyping. Finally, phenotyping of animals in testing stations for traits that are expensive or difficult to measure would still be necessary in genomic schemes in order to constitute and update the training population (TP) for these traits. Therefore, implementation of genomic evaluations would necessarily result in additional costs for pig schemes and comparing the efficiency, in terms of selection accuracy, genetic trends and inbreeding rates, of a traditional breeding scheme based on pBLUP genetic evaluation with that of a more expensive genomic breeding scheme would be clearly unfair. Thus, it is legitimate to question whether it might not be more profitable to invest in improvements in

the efficiency of current schemes rather than in the implementation of genomic evaluation.

The objective of this study was to compare the efficiency – in terms of annual genetic trends and annual increases in inbreeding – of a traditional pig sire line breeding scheme based on pBLUP genetic evaluation and the combined phenotyping of candidates and their sibs, with that of a genomic breeding scheme involving the same overall cost. The main weakness of the traditional pig breeding scheme considered here is the limited number of sibs that are phenotyped per year, resulting in poor accuracy of EBV for traits that are not directly measured on the selection candidates. Thus, as a first step, we evaluated the benefits of gradually increasing the annual number of phenotyped sibs under a pBLUP scheme. The additional cost of a genomic scheme depends directly on the number of genotyped animals. Thus, in a second step, we estimated the efficiency of GS designs for different numbers of genotyped candidates. Finally, increases in the capacity to phenotype sibs and the different numbers of genotyped candidates were converted into extra annual costs compared to a reference situation that was assumed to be representative of a current standard pig sire line breeding scheme, in order to determine the most profitable investment from a genetic point of view.

Methods

No ethical approval was required for this study because no animals were used.

Simulated population

To compare the efficiency of different breeding schemes, we carried out a stochastic simulation of a purebred paternal pig population consisting of 1050 breeding females distributed across five herds of equal size, and 50 breeding males. The population was selected for a global breeding goal that included two genetically independent traits with an initial heritability of 0.4 and an initial genetic variance of 1, based on either traditional pBLUP EBV or genomic EBV (hereinafter referred to as PEBV and GEBV, respectively). Trait 1 was representative of a fattening trait that is inexpensive and easy to measure on-farm, and was recorded on a large number of candidates, whereas trait 2 was representative of a trait that is too difficult or too expensive to measure on a large scale (e.g. feed efficiency) or even that can not be measured on candidates (e.g. carcass composition or meat quality) and was therefore recorded on a limited number of sibs of the selection candidates. Two series of simulations were performed, considering two alternative sets of economic values in the breeding goal, i.e. W1:1 and W1:3. For W1:1, the breeding goal was $\frac{1}{\sqrt{2}}TBV_1 + \frac{1}{\sqrt{2}}TBV_2$, whereas for W1:3, the breeding goal was $\frac{1}{\sqrt{10}}TBV_1 + \frac{3}{\sqrt{10}}TBV_2$, where

TBV_1 and TBV_2 are the true breeding values for traits 1 and 2, respectively. For both W1:1 and W1:3, the initial genetic variance of the breeding goal was equal to 1. In the following, these same economic weights were applied to the PEBV of animal i for traits 1 and 2 ($PEBV_{1_i}$ and $PEBV_{2_p}$, respectively) or to its GEBV for traits 1 and 2 ($GEBV_{1_i}$ and $GEBV_{2_p}$, respectively) to compute its EBV for the breeding goal in the pBLUP ($PEBV_{gl_i}$) and genomic ($GEBV_{gl_i}$) selection schemes.

The procedures used to simulate the genome are detailed in [7]. The simulated genome consisted of 10 pairs of 100 cM chromosomes, each chromosome carrying 1500 neutral single nucleotide polymorphism (SNP) markers, 90 biallelic randomly positioned quantitative trait loci (QTL) for trait 1, and 90 biallelic randomly positioned QTL for trait 2. All markers and QTL had a minor allele frequency higher than 0.05. The SNPs were unevenly spaced, with an average distance between two consecutive SNPs of 67 kb. Short- and long-range LD in the simulated population was comparable to the LD actually observed in the major French pig nucleus populations. By convention, allele 1 of each QTL had no effect and the absolute values of the effect of QTL allele 2 were sampled from a gamma distribution with shape and scale parameters equal to 0.4 and 1/1.66 [8], and given a positive or negative sign with probability 0.5. For both traits, the effects of the QTL were rescaled to result in a genetic variance of 1 in the breeding population when selection started. The true breeding value of animal i for trait t ($t = 1$ or 2) was calculated as:

$$TBV_{t_i} = \sum_{j=1}^{10} \left(\sum_{k=1}^{90} m_{t_{jki}} * a_{t_{jk}} \right),$$

where $m_{t_{jki}}$ is the genotype (i.e. 0, 1 or 2 copies of allele 2) of animal i for the k^{th} QTL for trait t on chromosome j , and $a_{t_{jk}}$ is the effect of the k^{th} QTL for trait t on chromosome j .

A total of 100 independent replicates were performed for each of the scenarios considered. For each replicate, a new genome was simulated and 1050 base females were randomly assigned to five herds of equal size, and within each herd into seven batches of 30 females. The time step (TS) was three weeks, which corresponded to the period between the farrowing of females from two consecutive batches. Therefore, a sow farrowed every 7 TS until culling. For simplicity, we considered that 18 TS (i.e. 378 days) corresponded to one year of simulation.

The population was first selected and bred for five years (i.e. 90 TS) under a pBLUP scenario that was designed to be representative of a traditional pig sire line breeding scheme, called BL_ref hereinafter. After this boot period, selection continued for 10 more years (i.e. 180 TS), under either the same BL_ref or according to four other pBLUP scenarios with greater capacity to phenotype sibs for trait 2, or under five genomic scenarios that differed in the annual number of candidates genotyped and the size of the TP for

trait 1. The results of these alternative scenarios were compared over this 10-year period.

Breeding scenarios

The characteristics of the ten scenarios that were considered are summarized in Table 1 and presented in Figure 1. The numbers of boars and sows and the numbers of candidates selected per TS to replace the breeding animals was the same for each scenario.

Reference pBLUP breeding scenario (BL_ref)

The BL_ref scenario is identical to the so-called “BLUP-AM” scenario described in [7] and was modelled based on the average statistics of the French Piétrain nucleus population that were obtained from the data used for national genetic evaluations. Briefly, 27 and 24 litters of the 30 litters born in each batch and on each herd (i.e. 90% and 80% of the born litters, respectively) contained three female and three male selection candidates, respectively. A phenotype for trait 1 was simulated for each candidate. A total of 15 litters per batch across the five herds (representing 10% of the born litters) also had one piglet that was phenotyped for trait 2 (these animals will be referred to as the “phenotyped sibs” in the following). These 15 litters were chosen among those containing candidates, by selecting the best litters as a function of their parental PEBV for the population breeding goal, giving priority to litters of boars with the fewest offspring among the already chosen phenotyped sibs and balancing the number of litters chosen from each herd. At each TS, the PEBV were estimated independently for the two traits, using a standard BLUP-Animal Model procedure [9] and the BLUPF90 software (<http://nce.ads.uga.edu/~ignacy/newprograms.html>), considering all phenotypes and pedigrees recorded since the creation of the herds. At each TS, the best 10 of the 81 female candidates from each herd and the 9 best of the 360 male candidates from across the five herds were selected based on $PEBV_{gl}$ to partially replace the pool of male and female reproducers.

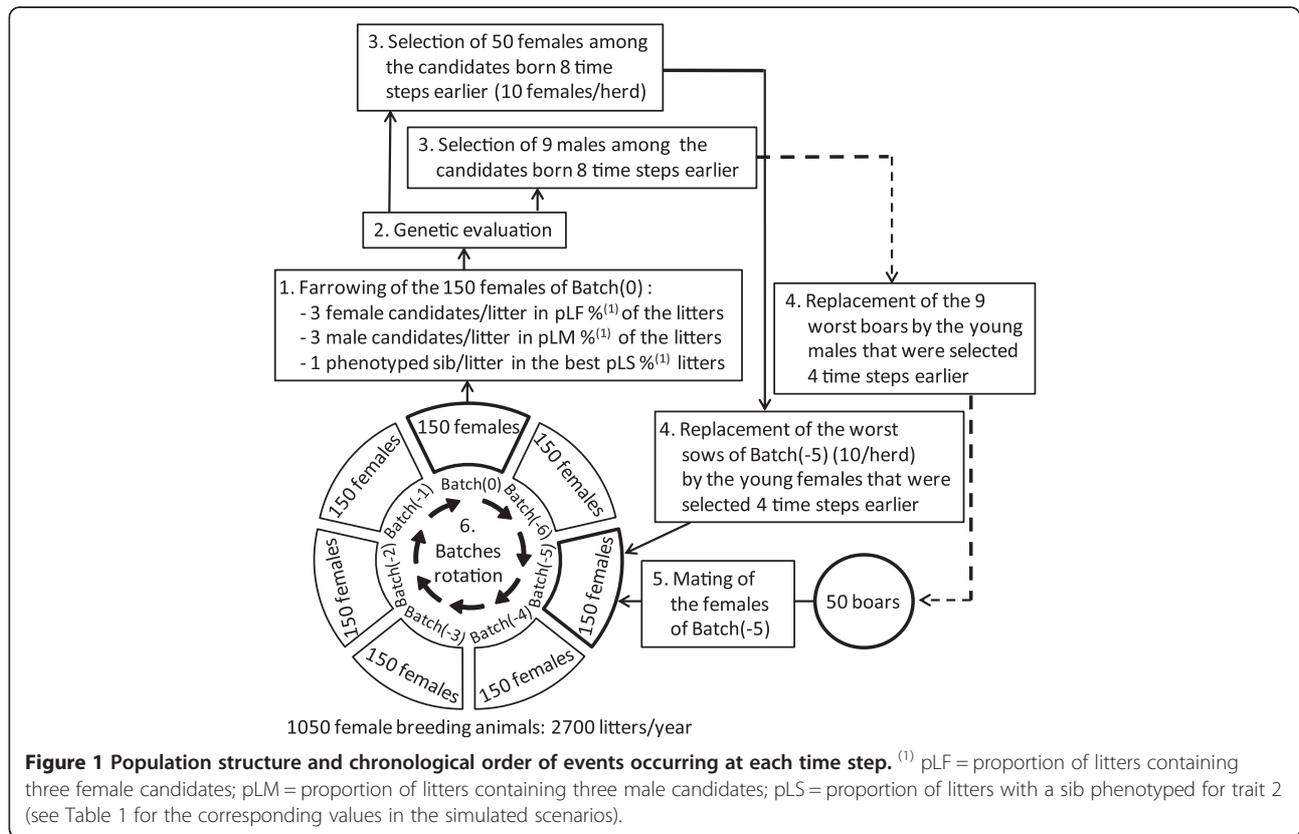
Alternative pBLUP scenarios

The main weakness of the BL_ref scheme lies in the limited number of sibs that are phenotyped for trait 2 (one animal in 10% of the litters), resulting in low accuracies of PEBV for candidates for this trait. Therefore, increasing the phenotyping capacity for trait 2 is expected to improve the efficiency of the traditional scheme. To evaluate this improvement, the BL_ref scenario was modified by increasing from the 90th TS (that will be referred to as TS90 in the following) the proportion of litters with a phenotyped sib to 30%, 50%, 70% or 90% in the alternative BL_30%, BL_50%, BL_70% and BL_90% scenarios, respectively. Similar to the BL_ref scenario, the litters were chosen based on parental PEBV for the population breeding goal, giving priority to

Table 1 Characteristics of the simulated pBLUP and genomic breeding scenarios

	Breeding scenarios ¹									
	BL_ref	BL_30%	BL_50%	BL_70%	BL_90%	GE_ref	GE_80%	GE_60%	GE_40%	GE_20%
Proportion of litters with 3 female candidates (pLF)	90%	90%	90%	90%	90%	90%	73%	53%	37%	17%
Nb of litters with female candidates/herd*batch (nLF)	27	27	27	27	27	27	22	16	11	5
Total annual ² nb of female candidates, all herds together (nFC)	7290	7290	7290	7290	7290	7290	5940	4320	2970	1350
Proportion of litters with 3 male candidates (pLM)	80%	80%	80%	80%	80%	80%	63%	47%	33%	17%
Nb of litters with male candidates/herd*batch (nLM)	24	24	24	24	24	24	19	14	10	5
Total annual ² nb of male candidates, all herds together (nMC)	6480	6480	6480	6480	6480	6480	5130	3780	2700	1350
Proportion of litters with 1 phenotyped sib (pLS)	10%	30%	50%	70%	90%	10%	10%	10%	10%	10%
Nb of litters with 1 phenotyped sib/herd*batch (nLS)	3	9	15	21	27	3	3	3	3	3
Total annual ² nb of phenotyped sibs, all herds together (nPS)	270	810	1350	1890	2430	270	270	270	270	270
Initial size of the training population for trait 1 (at year 6) (S1 ₀)						13 770	10 710	8415	5355	3060
Nb of time steps to constitute the initial training population for trait 1			not relevant			18	14	11	7	4
Annual ² increase in size of the training population for trait 1						13 770	11 070	8100	5670	2700
Initial size of the training population for trait 2 (at year 6) (S2 ₀)						1000	1000	1000	1000	1000
Nb of time steps to constitute the initial training population for trait 2			not relevant			67	67	67	67	67
Annual ² increase in size of the training population for trait 2						270	270	270	270	270

¹BL_ref, BL_30%, BL_50%, BL_70%, BL_90% = breeding schemes based on traditional pBLUP genetic evaluations using phenotypes of the candidates for trait 1 and phenotypes of sibs for trait 2 sampled in 10%, 30%, 50%, 70% or 90% of the litters, respectively; GE_ref = breeding scheme based on genomic evaluations with annual numbers of candidates and phenotyped sibs identical to BL_ref scenario; GE_80%, GE_60%, GE_40%, GE_20% = breeding schemes based on genomic evaluations with initial training population for trait 1 and annual number of candidates reduced by 20%, 40%, 60% and 80%, respectively, compared to the GE_ref scenario; ²the actual period of time corresponds to 18TS, i.e. 378 days, assimilated for simplicity to one year.



the litters of boars with the fewest offspring among the already chosen phenotyped sibs and balancing the number of litters chosen from each herd. BL_90% corresponded to the case where all litters that had female candidates (i.e. 90% of the born litters) also had one sib that was phenotyped for trait 2 (i.e. no litter selection).

Reference genomic breeding scenario (GE_ref)

The GE_ref scenario differed from the BL_ref scenario only by the procedure used to estimate breeding values; the numbers of male and female candidates, phenotyped sibs, selected candidates and breeding males and females were the same. This scenario was similar to the “GE-2TP” scenario detailed in [7].

Starting in TS90, all candidates, phenotyped sibs and breeding animals of the GE_ref scenario were genotyped, and all selection and culling decisions were based on $GEBV_{gl}$, which were estimated as follows. At TS90, the effects of all SNPs on traits 1 and 2 were estimated using the BLUP methodology described in [1] on two distinct TP, one for each trait. The initial TP for trait 2 (TP2) consisted of 1000 phenotyped sibs of the candidates born between TS16 and TS82, that were genotyped and phenotyped for trait 2. For trait 1, the initial TP (TP1) consisted in 13 770 male and female candidates that were genotyped

and phenotyped for this trait during the 5th year of the simulation (TS73 to TS90).

From a practical point of view, the genome simulated in this study was approximately three times shorter than the real pig genome (10 M instead of 30 M). Therefore, TP1 and TP2 were simulated by randomly sampling one third of the candidates and phenotyped sibs at each TS, as described and illustrated in [7], in order to produce GEBV with appropriate accuracies [10]. Thus, the initial TP1 and TP2 contained 4590 and 333 simulated animals, respectively, equivalent to 13 770 or 1000 animals, respectively, for a full genome.

The estimated SNP effects were used in the current TS and during the next 17 TS (i.e. 1 year) to calculate the GEBV of the breeding animals and candidates for trait

t as $GEBV_{t_i} = \sum_{j=1}^{10} \left(\sum_{k=1}^{1500} m_{t_{jk}} * \hat{g}_{t_{jk}} \right)$, where $\hat{g}_{t_{jk}}$ is the estimated effect of the k^{th} SNP on chromosome j for trait t . Then, $GEBV_{1_i}$ and $GEBV_{2_i}$ were combined based on their economic weights to calculate $GEBV_{gl_i}$.

Each year, the effects of the SNPs on the two traits were re-estimated using a TP that was increased by data from the 270 new phenotyped sibs and the 13 770 new candidates that were phenotyped in the past year (in

fact, 90 and 4590 randomly sampled additional animals, respectively, to account for the size of the simulated genome), and these new estimates were then used to calculate the $GEBV_i$ of breeding animals and candidates for the current and next 17 TS. Thus, the size of TP2 increased from 1000 to 3430 animals from the 6th to the 15th year of simulation. To limit computation time, the pigs of the oldest year were removed from TP1 every year, starting in year 10. Thus, the size of TP1 increased from 13 770 animals at year 6 to 55 080 at year 9, and then remained stable. Preliminary results had shown that the loss in accuracy for trait 1 due to removal of the oldest animals from TP1 was very low (approximately 1% in year 10).

Alternative genomic breeding scenarios

Under the GE_ref scenario, the number of candidates that were genotyped per year from year 5 to year 15 was identical to the annual number of candidates in the BL_ref scheme (13 770 animals/year). The extra cost due to genotyping was therefore considerable in the GE_ref scenario. A reduction in the annual number of genotyped candidates would reduce this cost but also cause a loss of technical efficiency of the GS scheme because of the resulting decreases in selection intensities, size of TP1 and accuracies of GEBV for trait 1. To evaluate this loss, four genomic schemes were derived from the GE_ref scenario. Under these schemes, the initial size of TP1 (at TS90) and the number of candidates genotyped per batch from the 6th year of simulation were reduced by approximately 20% (GE_80% scheme), 40% (GE_60%), 60% (GE_40%) and 80% (GE_20%), as described below:

1) Instead of all candidates from the 18 TS of the 5th year under the GE_ref scenario, the initial TP1 under the GE_80%, GE_60%, GE_40% and GE_20% scenarios included male and female candidates that were phenotyped in the last 14, 11, 7 and 4 TS of year 5, respectively;

2) The numbers of litters, per herd and TS, with three male and three female candidates that were phenotyped and genotyped ($n_{LM} = 24$ and $n_{LF} = 27$ litters under the GE_ref scenario, respectively) were reduced to 19 and 22 litters for GE_80%, to 14 and 16 litters for GE_60%, to 10 and 11 litters for GE_40%, and to 5 and 5 litters for GE_20%. At each TS, these litters were chosen in each herd from among the 27 litters available by selecting the best ones based on parental EBV for the breeding goal.

Technical efficiency criteria

One hundred independent replicates of five years of selection under the BL_ref scenario, followed by 10 years of selection under each of the 10 alternative scenarios, were simulated for the W1:1 and W1:3 breeding goals. For each replicate of each breeding scenario x breeding

goal combination, the following summary statistics were calculated:

the mean of the true breeding values of the piglets born in each TS for traits 1 and 2 and for the breeding goal, in order to measure the realized genetic gain;

the mean of the inbreeding coefficients of the piglets born in each TS based on pedigree information using PEDIG software [11], to measure the evolution of inbreeding over time;

the correlation between the TBV and the PEBV or GEBV of the selection candidates for each TS, in order to estimate accuracies of the selection for traits 1 and 2.

The above statistics were then used to derive the average annual realized genetic gains (ΔGa), the average annual increase in inbreeding (ΔFa), and the average accuracies from year 6 to year 15 for each replicate and averaged over the 100 replicates for each breeding scenario x breeding goal combination.

Evaluation of the extra costs of the alternative breeding schemes

To be able to compare the relative efficiencies of the genomic and pBLUP simulated schemes at constant cost, we evaluated the extra cost of the nine alternative schemes compared to the BL_ref reference scenario, as follows.

Alternative pBLUP schemes

We assumed that the increase in the number of phenotyped sibs would generate extra costs linked to the building of additional testing station facilities and the raising and phenotyping of animals. Based on the range of costs observed in testing stations in France and other countries (IFIP, French Pork and Pig Institute survey in 2005, personal communication), and assuming amortization of buildings and equipment over a 10-year period, a low (€150), medium (€350) and high (€550) cost per extra sib phenotyped for trait 2 were considered. The annual extra costs of the alternative pBLUP schemes, compared to the BL_ref scheme, were therefore calculated by multiplying the annual numbers of additional phenotyped sibs (540, 1080, 1620 and 2160 animals for BL_30%, BL_50%, BL_70% and BL_90%, respectively) by the individual additional cost under the low, medium and high cost hypotheses.

Alternative genomic schemes

Implementation of genomic selection requires genotyping of the initial TP1 and TP2, and continuous genotyping of new candidates and phenotyped sibs. To evaluate the resulting extra costs, we considered that boars and sows used for breeding were genotyped with the porcine60SNP beadchip but that selection candidates and phenotyped sibs were genotyped with a lower density and less expensive beadchip. Indeed, several studies [12-14] have shown

that it is possible to impute the genotypes of young animals for the missing SNPs with a very low error rate when both parents are genotyped with the denser beadchip. Three hypotheses were considered to take uncertainties regarding the future evolution of genotyping costs into account: the genotyping costs considered for the high density (HDpr) and the low density (LDpr) chips were €150/animal and €50/animal, respectively, under the high price hypothesis, €120 and €30 under the medium price hypothesis, and €90 and €10 under the low price hypothesis.

The numbers of animals in the initial TP1 (S_{10}) under the five genomic scenarios considered are in Table 1. The sires and dams of these animals were genotyped with the denser chip to increase the imputation accuracy; for simplicity, we considered a constant number of parents regardless of the genomic scenario, equal to the annual number of new breeding animals (900 gilts and 162 boars). The S_{20} (i.e. 1000) phenotyped sibs that constitute the initial TP2 were assumed to be genotyped with the high density chip rather than imputed. The genotyping costs for the initial TP1 and TP2 were therefore calculated as $GC_{ini} = S_{10} * LDpr + (S_{20} + 900 + 162) * HDpr$. Similar to the amortization of additional testing station capacity under the pBLUP scenarios, these initial genotyping costs were spread over a 10-year period. In addition, nFC female candidates, nMC male candidates and nPS phenotyped sibs were genotyped every year for the low density chip, and the 900 female candidates and the 162 male candidates selected each year as breeding animals were re-genotyped with the denser chip to enable high density imputation of

their future offspring. Thus, the annual extra cost of any of a genomic scheme was calculated as:

$$aGC = ((nMC + nFC + nPS) * LDpr + (900 + 162) * HDpr) + GC_{ini}/10$$

$$= ((nMC + nFC + nPS) + (S_{10}/10)) * LDpr + 1268.2 * HDpr.$$

Results

Accuracies and genetic trends

Comparison of the pBLUP and genomic reference breeding schemes

Overall, the GE_ref breeding scheme produced higher genetic gains than the BL_ref breeding scheme. When the population was selected for the W1:1 breeding goal, GE_ref resulted in 11%, 83% and 28% higher ΔGa than BL_ref for trait 1, trait 2 and the breeding goal, respectively (Table 2). This superiority was due to more accurate EBV in the genomic scheme (Table 3). Whereas the accuracy of the PEBV of candidates was equal to only 0.62 for trait 1 and 0.33 for trait 2 on average over the considered period, the average accuracy of the GEBV increased from approximately 0.76 (year 6) to 0.85 (year 9) for trait 1, and from 0.45 (year 6) to 0.55 (year 15) for trait 2, as the sizes of TP1 and TP2 increased (results not shown). On average, over the 10-year comparison, the accuracy of the GEBV was 0.83 and 0.52 for traits 1 and 2, respectively, i.e. 34% and 58% higher than the accuracy of the PEBV. The superiority of GE_ref over BL_ref in terms of ΔGa was much greater for trait 2 than for trait 1 because the increase in accuracy was

Table 2 Average annual genetic trends under different scenarios, for two breeding goals

Relative weights on traits 1 and 2 in the breeding goal ⁽²⁾	Average annual genetic trends, in genetic standard deviation units					
	Trait 1		Trait 2		Breeding goal	
	1:1	1:3	1:1	1:3	1:1	1:3
Scenarios ⁽¹⁾						
BL_ref	0.63 ^a	0.42 ^a	0.18	0.31	0.57	0.43
BL_30%	0.60	0.36 ^c	0.27	0.42	0.61 ^a	0.51
BL_50%	0.58 ^b	0.35 ^d	0.32 ^a	0.45	0.63 ^b	0.54 ^a
BL_70%	0.58 ^b	0.33 ^{de}	0.33 ^b	0.48 ^a	0.64 ^b	0.56
BL_90%	0.57	0.33 ^e	0.36	0.50 ^{bc}	0.65	0.58 ^{bc}
GE_ref	0.70	0.40 ^{ab}	0.33 ^b	0.51 ^c	0.73	0.61
GE_80%	0.68	0.39 ^b	0.34 ^b	0.50 ^{bc}	0.72	0.60 ^d
GE_60%	0.66	0.37 ^c	0.34 ^b	0.50 ^b	0.71	0.59 ^{cd}
GE_40%	0.62 ^a	0.35 ^{de}	0.34 ^b	0.49 ^b	0.68	0.57 ^b
GE_20%	0.54	0.29	0.33 ^{ab}	0.48 ^a	0.61 ^a	0.54 ^a

The results presented are means of the 100 replicates; standard deviations of the 100 replicates ranged from 0.05 to 0.07; ¹see Table 1 for a description of the different breeding scenarios;

²1:1 Breeding goal = $1/\sqrt{2}$ Breeding Value for trait 1 + $1/\sqrt{2}$ Breeding value for trait 2;

1:3 Breeding goal = $1/\sqrt{10}$ Breeding Value for trait 1 + $3/\sqrt{10}$ Breeding value for trait 2; within a column: all means differ ($p < 0.01$) except means with common superscripts.

Table 3 Average accuracy of the estimated breeding values of candidates for different scenarios, for two breeding goals

Relative weights on traits 1 and 2 in the breeding goal ⁽²⁾	Average accuracy of estimated breeding values for			
	Trait 1		Trait 2	
	1:1	1:3	1:1	1:3
Scenario ⁽¹⁾				
BL_ref	0.62	0.68	0.33	0.28
BL_30%	0.63	0.69 ^a	0.41	0.36
BL_50%	0.63 ^a	0.69 ^a	0.44	0.39
BL_70%	0.63 ^{ab}	0.69 ^a	0.46	0.41
BL_90%	0.64 ^b	0.69 ^a	0.48	0.41
GE_ref	0.83	0.88	0.52 ^a	0.47
GE_80%	0.81	0.87	0.51 ^a	0.46 ^a
GE_60%	0.78	0.85	0.51 ^a	0.46 ^a
GE_40%	0.75	0.82	0.51 ^a	0.46 ^a
GE_20%	0.68	0.76	0.52 ^a	0.46 ^a

The results presented are means of the 100 replicates; standard deviations of the 100 replicates ranged from 0.01 to 0.02 for trait 1, and from 0.02 to 0.05 for trait 2; ¹see Table 1 for a description of the different breeding scenarios; ²see Table 2 for a description of the breeding goals; within a column: all means differ ($p < 0.01$) except for values with the same superscripts.

greater for trait 2 than for trait 1. This indirectly reinforced the relative importance of trait 2 compared to trait 1 in the selection index.

When the population was selected for the W1:3 breeding goal, GE_ref also resulted in markedly higher ΔGa than BL_ref for trait 2 and for the breeding goal (64% and 41%, respectively), but produced a slightly lower ΔGa for trait 1 than the traditional scheme (-5%), despite the fact that the GEBV were more accurate than the PEBV for both traits. This result was due to the greater weighting of trait 2 in the W1:3 breeding goal than in the W1:1 breeding goal. Indeed, the larger weight on trait 2 in the W1:3 breeding goal counterbalanced its lower accuracy compared to that of trait 1 in the BL_ref scenario. Therefore, the selection pressure on the two traits was more balanced for W1:3 than for W1:1, for which the selection pressure was mainly on trait 1. Because of the greater increase in accuracy for trait 2 than for trait 1 (+68% vs. +30%) when switching from BL_ref to GE_ref, the selection pressure on trait 2 in the W1:3 selection index, which was already large in BL_ref, was further increased in GE_ref. Because of the high selection pressure on trait 2, a small loss of selection response was observed for trait 1.

Improved pBLUP scenarios with larger numbers of phenotyped sibs

The accuracy of the PEBV of candidates for trait 2 improved significantly when the proportion of litters with one sib phenotyped for trait 2 increased (Table 3). Phenotyping

one sib in 90% rather than in 10% of the litters resulted in an increase in the accuracy of the PEBV by on average 45%: 0.48 for BL_90% and 0.33 for BL_ref, for the W1:1 breeding goal. This improvement in accuracy for trait 2 was due to the larger number of half-sibs phenotyped for this trait per candidate, as well as the higher proportion of candidates that benefited from having one full-sib phenotyped. However, the average accuracy for trait 2 was lower for all pBLUP scenarios than for GE_ref. This greater accuracy for trait 2 from increasing the number of phenotyped sibs in the pBLUP schemes improved the genetic gains for this trait (Table 2) markedly. For example, testing one phenotyped sib in 50% instead of 10% of the litters resulted in a 71% greater ΔGa (+0.32 σ_g /yr vs. +0.18 σ_g /yr) with selection for the W1:1 breeding goal. In the extreme case of BL_90%, the average annual genetic trend for trait 2 (+0.50 σ_g /yr) almost equalled that of the GE_ref with the W1:3 breeding goal, and even slightly exceeded (+0.36 σ_g /yr) that of GE_ref with the W1:1 breeding goal.

As expected, accuracy of PEBV for trait 1 was the same for all pBLUP scenarios, since the number of phenotyped candidates was kept constant and the sibs were not phenotyped for trait 1. Nevertheless, it should be noted that the ΔGa for trait 1 declined with an increase in the number of sibs phenotyped for trait 2 because of the increasing selection pressure on trait 2, resulting from the increasing accuracy for trait 2. Logically, this unfavourable impact on trait 1 was greater when the population was selected for a breeding goal that placed more emphasis on trait 2. However, the genetic gain for trait 1 remained positive in all the cases.

Overall, increasing the number of phenotyped sibs significantly improved the efficiency of the pBLUP breeding scheme in terms of response in the breeding goal. The extreme BL_90% scenario, in which one animal from every litter suitable for genetic purposes was tested in station, led to a 14% increase in the ΔGa in the breeding goal with selection on W1:1, and a 35% increase with selection on W1:3. However, even in this very favourable case, the pBLUP scenario had 11% and 5% less response in the breeding goal than the GE_ref scheme with selection on W1:1 and W1:3, respectively.

Genomic scenarios with reduced numbers of candidates

Compared to GE_ref, reducing the initial size of TP1 and thereafter the number of litters containing genotyped candidates resulted in a lower ΔGa for trait 1 (Table 2). The loss of ΔGa for trait 1 was minor with small reductions in genotyping (approximately -3% and -7% compared to GE_ref for GE_80% and GE_60%, respectively) but was more marked with greater reductions in genotyping (-12% and -14% for GE_40%, and -24% and -29% for GE_20% compared to GE_ref for W1:1 and W1:3, respectively).

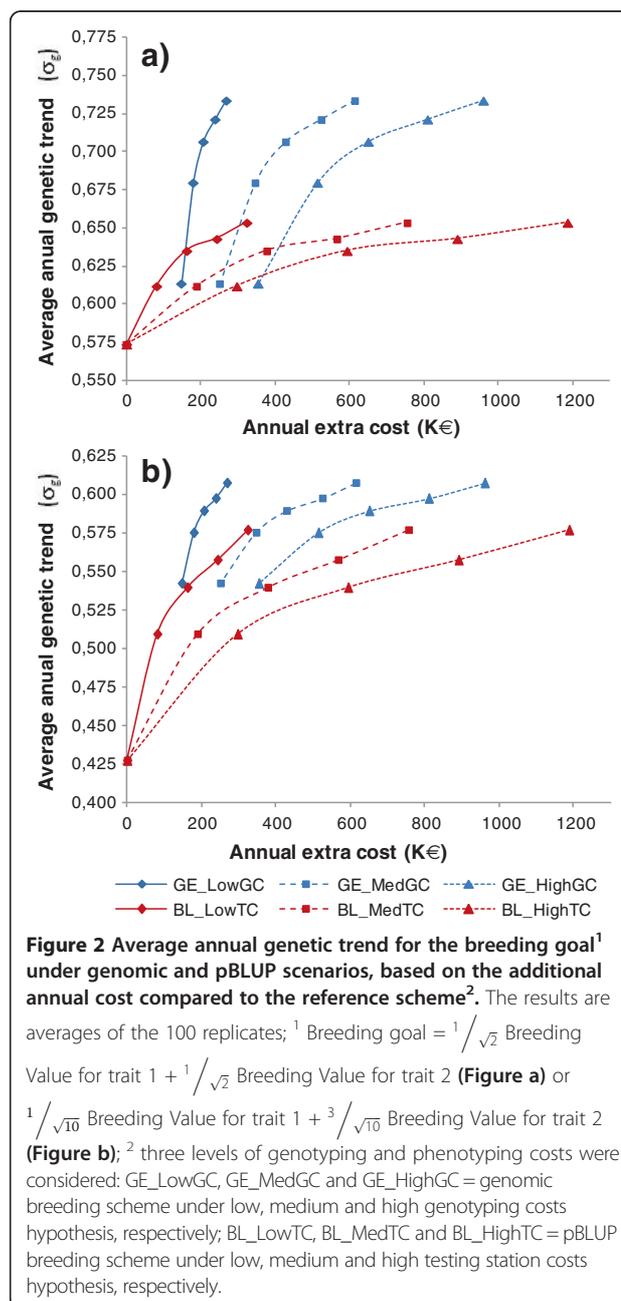
The origins of these losses were twofold. First, reducing the number of genotyped candidates reduced the initial size and growth of TP1 and, therefore, reduced the accuracy of the GEBV for trait 1 (Table 3). However, the reduction in the accuracy of the GEBV for trait 1 was only moderate. For example, the average accuracy over the 10-year period for W1:1 was only 5.5% lower for GE_{60%} than for GE_{ref}, despite a 40% reduction in the size of TP1, and only 9.4% lower for GE_{40%} than for GE_{ref}, despite a 60% reduction in the size of TP1, and remained greater than for the pBLUP schemes. The second cause for the reduction in ΔGa for trait 1 was a decrease in selection intensity, since a fixed number of young animals (10 females per TS in each herd, and nine males per TS) were selected from a smaller number of candidates.

The size and structure of TP2 was the same for the five genomic scenarios. Consequently, the accuracy of GEBV for trait 2 remained unchanged (Table 3). Thus, the decline in accuracy for trait 1 from GE_{ref} to GE_{20%} resulted indirectly in an increase in the selection emphasis on trait 2 in the selection index. This counterbalanced the negative effects of the lower selection intensity due to the smaller number of candidates on response for trait 2. Thus, ΔGa for trait 2 was not (for the W1:1 breeding goal) or only slightly (-6% from GE_{ref} to GE_{20%} for the W1:3 breeding goal) impacted by the reduction in the number of candidates (Table 2).

As a consequence of the impact on responses for traits 1 and 2, the efficiency of the genomic breeding scheme in improving the breeding goal was only moderately affected by the number of litters with candidates, and remained much greater than for BL_{ref} in all cases (Table 2). Reducing the initial size of TP1 and the number of litters with candidates by 40% and 80% only reduced the ΔGa for the breeding goal by 4% and 16%, respectively, compared to the GE_{ref} scenario, when the population was selected for the W1:1 breeding goal. This loss in efficiency was even less marked when a greater weight was given to trait 2 in the breeding goal (-3% and -11%, respectively, for GE_{60%} and GE_{20%} compared to GE_{ref}).

Economic comparison of genomic and pBLUP schemes

The ΔGa for the breeding goal (ΔGa_{bg}) produced under the five pBLUP and five genomic schemes, as a function of their additional annual costs compared to the BL_{ref} scheme, are plotted in Figure 2a and 2b for the W1:1 and W1:3 breeding goals, respectively. These comparisons were made under the three hypotheses for the cost of genotyping and under the three hypotheses for the cost of phenotyping additional sibs. Interpolations were performed to estimate the ΔGa_{bg} for numbers of litters with one phenotyped sib and for genotyping reductions that were intermediate to those simulated. The lower bound for the three pBLUP scenarios corresponded to the BL_{ref} scheme (no additional



cost), whereas the upper bound corresponded to the BL_{90%} scheme, which was the most expensive and most efficient scenario in terms of ΔGa_{bg} . Similarly, the lower bound for the three genomic schemes was observed for the least expensive and least efficient scenario (GE_{20%}), whereas the upper bound corresponded to the most expensive and most efficient scenario (GE_{ref}).

When ignoring the extra costs, the genomic breeding schemes yielded markedly higher ΔGa_{bg} than the most efficient pBLUP scheme (BL_{90%}) when candidates from at least 30% (for W1:1) or at least 40% (for W1:3) of the

best litters with candidates in the reference scheme were genotyped. The potential extra gain in ΔGa_{bg} between the best genomic scheme (GE_ref) and the best pBLUP scheme (BL_90%) was, however, greater for the W1:1 breeding goal ($\Delta Ga = 0.73\sigma_g/\text{yr}$ vs. $0.65\sigma_g/\text{yr}$, i.e. an increase of 12%) than for the W1:3 breeding goal ($\Delta Ga = 0.61\sigma_g/\text{yr}$ vs. $0.58\sigma_g/\text{yr}$, i.e. an increase of 5%). Conversely, the traditional schemes always resulted in lower ΔGa_{bg} than the least efficient genomic scheme (GE_20%) when fewer than 30% (in the W1:1 case) or 50% (in the W1:3 case) of the litters had a sib phenotyped for trait 2.

Implementing the cheapest possible genomic scheme considered in this study (GE_20%) obviously involved a minimum extra annual cost, depending on the cost of genotyping an individual. For example, under the medium genotyping cost hypothesis, this minimum investment amounted to €250 000 per year. Below this threshold for annual additional expenses, the only possibility to improve the efficiency of the breeding programme would be to increase the number of phenotyped sibs.

Comparing the respective ΔGa_{bg} resulting from the pBLUP scheme and from the genomic scheme that could be implemented at a given extra cost, revealed two areas where the most efficient of the two strategies differed. Below a certain threshold for extra annual costs, ΔGa_{bg} was greater for the improved pBLUP scheme than for the genomic scheme, and above this threshold, the genomic scheme was more efficient. The value of this threshold for each combination of genotyping costs and costs of phenotyping an additional sib is presented in Table 4. Table 4 shows that, for a given cost of phenotyping an additional sib, this threshold increased with the genotyping cost. Conversely, for a given genotyping cost, the threshold decreased as the phenotyping cost of sibs increased. To illustrate this, for W1:1, an extra annual cost of €260 000 was more effective if used to implement a genomic breeding scheme than to improve the efficiency of the pBLUP scheme when the genotyping costs were low (regardless of the phenotyping cost of sibs), or when genotyping costs were medium and phenotyping costs of sibs

were high. In all other cases, these extra funds would be more efficiently used to increase the number of phenotyped sibs in the traditional scheme.

Inbreeding

Rates of inbreeding (ΔFa) observed in the 10 scenarios simulated for the two breeding goals are in Table 5.

Comparison of the pBLUP and genomic reference breeding schemes

The GE_ref scheme resulted in a smaller increase in the average inbreeding coefficient of the population than the BL_ref scheme, for both W1:1 and W1:3 (Figure 3). For GE_ref, following implementation of genomic evaluations, the rate of increase in inbreeding initially gradually declined, reflecting the progressive replacement of breeding animals selected based on PEBV until TS90 by males and females selected based on GEBV. Thereafter, the average population inbreeding rose at a steady rate that was similar for W1:1 and W1:3 but significantly lower than for BL_ref. Considering the complete 10-year period, ΔFa was on average 57% lower for GE_ref than for BL_ref when the population was selected based on the W1:1 breeding goal (+0.36% per year vs. +0.84% per year) and 68% lower for the W1:3 breeding goal (+0.51% per year vs. +1.58% per year).

Improved pBLUP scenarios with larger numbers of phenotyped sibs

No significant relationship was found between the annual number of sibs phenotyped for trait 2 and ΔFa when the population was selected based on the W1:1 breeding goal. When greater weight was given to trait 2 in the breeding goal (W1:3), increasing the proportion of litters with a sib recorded for trait 2 from 10 to 70% significantly reduced the annual increase in inbreeding by -17%. However, ΔFa tended to be 7% higher for BL_90% than for BL_70% (P -value of the difference = 0.053).

Table 4 Thresholds for the extra annual cost (in K€/yr) that delimit areas of interest to improve the pBLUP scheme or implement the genomic scheme, for two breeding goals

Cost of phenotyping additional sibs	Relative weights on traits 1 and 2 in the breeding goal ⁽¹⁾					
	1:1			1:3		
	Genotyping cost					
	Low	Medium	High	Low	Medium	High
Low	160	300	450	145	350	530
Medium	145	270	410	145	250	350
High	145	250	370	145	250	350

¹See Table 2 for a description of the breeding goals; below the threshold: improving the pBLUP breeding scheme resulted in higher genetic trends for the population breeding goal than implementing genomic evaluations; above the threshold: implementing genomic evaluations resulted in higher genetic trends for the population breeding goal than improving the pBLUP breeding scheme.

Table 5 Average annual increase in inbreeding under the scenarios compared, for two breeding goals

Scenario ⁽¹⁾	Average annual increase in inbreeding coefficient, %	
	Relative weights on traits 1 and 2 in the breeding goal ⁽²⁾	
	1:1	1:3
BL_ref	+0.84 ^a (0.28)	+1.58 ^a (0.53)
BL_30%	+0.85 ^a (0.26)	+1.48 ^{ab} (0.43)
BL_50%	+0.85 ^a (0.26)	+1.39 ^{bc} (0.34)
BL_70%	+0.86 ^a (0.27)	+1.32 ^c (0.33)
BL_90%	+0.88 ^a (0.22)	+1.41 ^{bc} (0.40)
GE_ref	+0.36 ^b (0.13)	+0.51 ^d (0.23)
GE_80%	+0.39 ^b (0.13)	+0.52 ^d (0.20)
GE_60%	+0.46 (0.16)	+0.58 (0.21)
GE_40%	+0.56 (0.16)	+0.65 (0.20)
GE_20%	+0.76 (0.22)	+0.84 (0.25)

The results presented are means and (standard deviations) of the 100 replicates; ¹see Table 1 for a description of the different breeding scenarios; ²see Table 2 for a description of the breeding goals; within a column: all means differ ($p < 0.01$) except for values with the same superscripts.

Genomic scenarios with reduced numbers of candidates

Increasing restrictions on the number of litters with candidates logically resulted in selecting males and females that were more related to each other. This significantly accelerated the increase in inbreeding in the population.

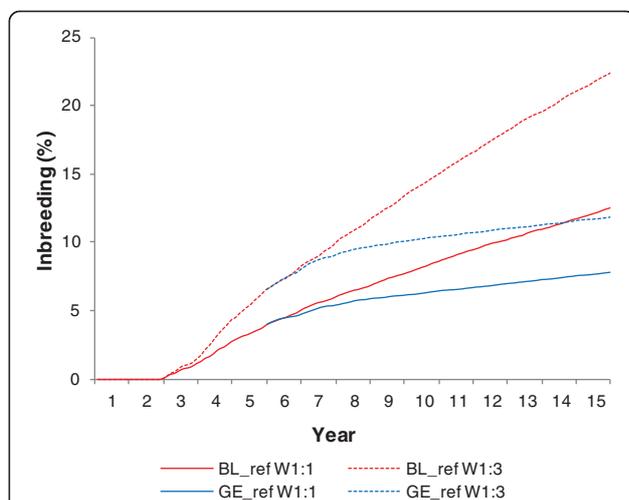


Figure 3 Evolution of population inbreeding under the pBLUP and genomic reference scenarios¹, for two breeding goals².

Averages of the 100 replicates; ¹simulated scenarios: BL_ref = breeding scheme based on traditional pBLUP genetic evaluations using phenotypes of the candidates for trait 1 and phenotypes of sibs for trait 2 sampled in 10% of the litters; GE_ref = breeding scheme based on genomic evaluations with annual numbers of candidates and phenotyped sibs identical to the BL_ref scenario; ²W1:1: Breeding goal = $1/\sqrt{2}$ Breeding Value for trait 1 + $1/\sqrt{2}$ Breeding Value for trait 2; W1:3: Breeding goal = $1/\sqrt{10}$ Breeding Value for trait 1 + $3/\sqrt{10}$ Breeding Value for trait 2.

For example, the average inbreeding coefficient increased 8, 28, 55 and 111% faster for GE_80%, GE_60%, GE_40% and GE_20%, respectively, than for GE_ref, when the population was selected based on the W1:1 breeding goal. However, the global increase in inbreeding over the 10-year period was smaller for all genomic schemes than for any of the pBLUP schemes. Moreover, the advantage of the genomic over the pBLUP schemes was even more pronounced when the relative weight of trait 2 in the breeding goal increased: ΔFa values were 10 and 36% lower for the least favourable genomic scenario (GE_20%) compared to the most favourable pBLUP scheme for W1:1 and W1:3, respectively.

Discussion

Simulation hypothesis

The results presented here were obtained by stochastic simulation of a selected population under several scenarios. Particular attention was paid to modelling a realistic pig population structure with overlapping generations, as well as a genome presenting appropriate linkage disequilibrium and adequate marker density. However, several assumptions and simplifications were necessary, as in any simulation study, mainly to ensure that computation time remained reasonable. For instance, management of inbreeding was not considered in the selection and mating procedures, while this is a major concern for pig breeders. This resulted in markedly higher inbreeding rates and probably stronger selection intensities than what is observed in practice. However, the same selection and mating procedures were applied to both the genomic and the pBLUP scenarios. One can therefore assume that the comparisons were relevant to assess the relative merits of the breeding schemes considered. Nevertheless, the economic results should not be considered as precise predictions but rather as orders of magnitude.

For simplicity, only two unrelated traits were considered in the breeding goal. However, in practice, most traits recorded on phenotyped sibs and candidates are genetically correlated, meaning that measurements on one trait carry information on other traits. On the one hand, multiple-trait pBLUP genetic evaluations are routinely used in current pig breeding schemes to take advantage of this additional information, providing increased accuracies compared to single-trait PEBV. On the other hand, multiple-trait genomic evaluation models have been little explored and evaluated to date, although several models have been proposed. Calus and Veerkamp [15] and Jia and Jannink [16] showed by simulations that multiple-trait genomic evaluations can generate higher accuracies of GEBV than single-trait models. Similar to traditional pBLUP evaluations, greater improvements in accuracy were observed for low-heritable traits that are highly correlated with high-heritable traits. Moreover, according to [15], using multiple-

trait genomic evaluations would be particularly advantageous for traits for which the phenotype is not directly observed on candidates, like trait 2 in the present study. However, for such traits, the relative increase in accuracy generated by multiple-trait compared to single-trait models tended to be lower with genomic evaluations than with pBLUP evaluations [15]. This suggests that the superiority of the pig genomic breeding schemes over traditional breeding schemes could be somewhat lower than estimated here for independent traits when traits are correlated and multiple-trait evaluations are used. In addition, genomic selection generates greater genetic gain than pedigree-based genetic evaluations, and could therefore result in fixing favourable alleles of the QTL more rapidly. One can thus assume that genomic selection could change genetic correlations between traits more markedly than traditional selection, modifying the long-term efficiency of the breeding scheme.

Several studies [17-19] have shown that the accuracy of genomic predictions may depend significantly on the number of QTL affecting the traits and on the distribution of their effects. It would therefore be interesting to test the sensitivity of our results by repeating our simulations with different genetic architectures. However, similar relative improvements in terms of average annual genetic trends were observed in [7] for three different QTL densities (10 QTL/M, 30 QTL/M and 60 QTL/M) when changing from BL_ref to GE_ref. Moreover, Daetwyler *et al* [18] observed that the accuracy of the GEBV calculated with the GBLUP method, as in our study, seemed not to be affected by the number of QTL or by the distribution of QTL variance. These findings suggest that our results should be relatively conservative to the genetic architecture hypothesis, but further simulations are needed to confirm this.

Accuracy and genetic trends

Our results tended to show that replacing traditional pBLUP genetic evaluations by genomic evaluations in a breeding scheme organized around the combined phenotyping of candidates and sibs could result in more improvement in terms of accuracy of EBV, annual genetic trends and inbreeding than increasing the number of sibs phenotyped for trait 2 while retaining traditional evaluations. This superiority of genomic selection in terms of annual response in the breeding goal was due to the fact that the accuracy of EBV increased significantly for both traits compared to the reference traditional scheme. By contrast, under the pBLUP scheme, increasing the number of phenotyped sibs only improved the accuracy of the EBV for trait 2.

In the nucleus population simulated here, only six animals per litter (seven in litters that had a phenotyped sib) were used for genetic purposes, whereas the average number of weaned piglets per litter in the French Piétrain

population that was modeled is approximately eight. Therefore, another way to further improve the efficiency of the pBLUP scheme would be to phenotype these unused animals for trait 1, resulting in a higher accuracy of the PEBV for that trait. However, each candidate in the pBLUP scenarios already had an own phenotype for trait 1, along with phenotypes on its two parents, five full-sibs, and several half-sibs. Therefore, the benefits of these additional animals on the accuracy and genetic gain for trait 1 would be very low, and this option was not considered here.

We compared the accuracy of the pBLUP and genomic evaluations through the correlations between the TBV and the PEBV or GEBV of the evaluated animals. This criterion has emerged as the most commonly used metric to assess the prediction accuracy of a genetic evaluation method, since it has a linear relationship with response to selection. Another criterion to evaluate the merits of different breeding value prediction methods is their ability to produce unbiased estimates of true breeding values. Unbiasedness is a desirable property in genetic evaluation because it enables, in particular, a proper comparison of the genetic merit of animals present in different environments or times, and correct estimates of realized genetic trends. In simulations, unbiasedness of a given breeding value prediction method can be assessed by comparing the estimated and the true genetic trends in the simulated population over time, or by estimating the regression coefficient of true on estimated breeding values at each genetic evaluation time. Unfortunately, these criteria were not calculated in this study.

The potential extra gain for the breeding goal produced by the most efficient genomic scheme (GE_ref) compared to the most efficient pBLUP breeding scheme (BL_90%) depended on the relative economic weights of the traits in the breeding goal. When the trait recorded on candidates and the trait recorded only on sibs had comparable economic weights (W1:1), the potential extra gain of a genomic scheme was considerable, since it improved selection for both traits. By contrast, when a large proportion of the breeding goal concerns the trait recorded on phenotyped sibs (as in the W1:3 case), improving accuracy for the trait recorded on candidates was less crucial, and the potential extra gain of a genomic scheme over a traditional scheme was less.

The accuracy of a candidate's PEBV depends mainly on the availability of an own phenotype and on the number of its closest relatives that have a phenotype. In this study, this number was on average fixed over time for a given pBLUP scenario, as was the corresponding average accuracy of PEBV. In contrast, the accuracy of GEBV increased with the size of the TP. Thus, it could be assumed, if the simulation had been run for more years and if TP1 had grown continuously instead of being limited to candidates from the last four years after year 10

(for computation reasons), that the accuracy of GEBV for both traits would have tended to increase steadily until the oldest animals in the TP had become uninformative. The superiority of the GE_ref over the BL_ref scheme would therefore have increased further over time. Furthermore, in the simulations, the TP and SNP effects were only updated once a year in order to limit computation time. As a result, the accuracy of the GEBV of candidates, and therefore the selection efficiency in the 17 batches between two updates were lower than if the SNP effects had been re-estimated each TS using TP that would have been continuously updated with animals that are closely related to the candidates (e.g. [20-22]). In addition, selection based on a combination of the candidates' GEBV and their own phenotypes, which were routinely recorded, would also have slightly increased the accuracy of genomic selection. Thus, our results probably underestimated the potential technical superiority of genomic versus pBLUP schemes.

Inbreeding

Our results also showed that replacing pBLUP by genomic evaluations in a pig nucleus would have a very beneficial impact on the evolution of population inbreeding, confirming theoretical expectations [23]. The reason is that the breeding value of an individual can be considered as a combination of the breeding value of its parents and a Mendelian sampling term. The Mendelian sampling term is poorly estimated using pBLUP methodology for animals without phenotyped offspring, since the phenotypes of ancestors and relatives mostly contribute to the estimation of the parental component of an animal's breeding value. As a consequence, the PEBV of related animals are correlated, especially for traits for which the candidates do not have an own phenotype. Consequently, the pBLUP scenario favoured the co-selection of related animals from among the candidates. This phenomenon was even more pronounced for breeding goal W1:3 than for W1:1, because of the greater weight on trait 2, for which full-sib candidates had the same relatives phenotyped for this trait and therefore the same PEBV. In contrast, genomic evaluation results in a more precise estimate of the Mendelian component of the breeding values, producing less correlated GEBV for related animals, less co-selection of related animals, and therefore a smaller increase in inbreeding.

Increasing the number of litters with one sib phenotyped for trait 2 from 10% to 70% also reduced the increase in inbreeding under the pBLUP scenarios when a large weight was given to trait 2 in the breeding goal (W1:3). However, this favourable effect seemed to disappear, or even reverse, when this number exceeded 70%. The favourable trend observed up to 70% was not due to a reduction in the within-litter correlation between the EBV

of candidates, as in the genomic schemes, but the result of a change in the heterogeneity of the accuracy of the PEBV for trait 2 among the candidates. In schemes with the smallest numbers of phenotyped sibs, only a few candidates benefited from one full-sib being phenotyped for trait 2. These animals had more accurate, and therefore less regressed, PEBV for this trait compared to the majority of candidates, which had only more distant phenotyped relatives. The animals born in these few litters consequently had a greater chance of being selected to renew the nucleus, especially since their birth litter had a high parental average PEBV. Moreover, the PEBV of litter-mate candidates were highly correlated, since they had the same PEBV for trait 2 (based on the data of the same phenotyped sibs), which had a greater weight than trait 1 in the W1:3 breeding goal. This resulted in high inbreeding rates. Thus, increasing the proportion of litters with a phenotyped sib allowed candidates from more litters to have accurate PEBV for trait 2, which led to the selection of animals from a larger number of families and a deceleration in the increase in inbreeding. However, the procedure applied to choose litters with a phenotyped sib gave priority to litters produced by boars with the fewest offspring phenotyped for trait 2. In other words, the additional 20% phenotyped sibs in the BL_90% scenario compared to the BL_70% scenario mostly came from boars that had had the longest productive life and already a large number of offspring. One can assume that the litter-mates of these phenotyped sibs had consequently more accurate PEBV for trait 2 than the other candidates, and a greater chance to be selected, re-accelerating the increase in inbreeding in the population. In contrast, greater selection pressure was focused on trait 1 in the W1:1 goal, since the accuracy of the PEBV for this trait was higher than for trait 2. Because each candidate had its own phenotype for trait 1, the within-litter correlation between the PEBV was weaker for goal W1:1 than for W1:3, and the co-selection of full-sib candidates was less important. Thus, the level of inbreeding was lower for the W1:1 breeding goal, and the impact of the number of phenotyped sibs on this parameter was not noticeable.

Extra costs

The benefits of GS in terms of greater annual gain for the breeding goal and lower rates of inbreeding had a cost, since the animals making up the initial TP, as well as 13 770 candidates and 270 phenotyped sibs per year under the reference scheme had to be genotyped. Genotyping all these animals with the porcineSNP60 beadchip would generate very high extra costs, exceeding €2 million per year (considering an approximate current price of €150/animal) for the simulated population considered under the reference scheme. Thus, we assumed that the use of imputation techniques was possible, enabling a

dramatic reduction in the cost of implementing GS. Imputation is already routinely used in dairy cattle, with excellent results [12-14]. In pigs, recent studies [24-26] have also shown promising results. For simplicity, we assumed here that imputation of the candidates' missing genotypes did not cause a loss in selection efficiency, which is not fully realistic and tends to favour the genomic scenarios in the economic comparisons. However, the results of Hickey et al. [12] in pigs suggest that, assuming that candidates are genotyped for 5000 markers or more and have both their parents genotyped with the porcineSNP60 beadchip, the imputation error rate and therefore the loss of selection efficiency would indeed be very low, in which case our results remain valid.

The second solution investigated here was to reduce the implementation costs of genomic evaluation by reducing the number of animals genotyped. This was achieved by downsizing the initial TP1 and reducing the annual number of candidates. In order to limit the loss in selection efficiency, the parental EBV of the litters were used as prior information to detect the most useful animals to genotype. Our results showed that this strategy was effective. The loss of efficiency in terms of response in the breeding goal and rate of inbreeding compared to the best potential genomic scheme (GE_ref) was indeed limited with reductions up to approximately 50% in the initial size of TP1 and in the annual number of genotyped candidates. The limited degree of loss in accuracy of GEBV despite significant reductions in the number of genotyped animals was due to the characteristics of TP1. First, the accuracy of GS is positively related to the size of the TP (e.g. [10,27]) but the marginal increase in the predictive ability of a TP due to the inclusion of additional individuals decreases for large TP [10,28]. In our simulations, TP1 varied from 13 770 in year 6 to 55 080 after year 9 under GE_ref. With such numbers, and given the heritability of trait 1 considered here ($h^2 = 0.4$), the impact of a 20% or 40% reduction in size of the TP on accuracy was limited. It is however likely that the loss of accuracy would have been more marked for a less heritable trait. In addition, the accuracy of GEBV is also related to the degree of relationships between the candidates and the TP [29], which was high in the present case since TP1 consisted of candidates, some of which became parents of the next generation. Thus, despite a reduction in size, the candidates still had closely related animals in TP1, which probably limited the loss in accuracy.

The reduction in the number of candidates genotyped also resulted in a reduction in response for trait 1 because of a reduction in selection intensity, since a fixed number of young animals (10 females per TS for each herd, and nine males per TS) were selected from a smaller number of genotyped candidates. However, the loss in selection intensity was lower than might have been expected

considering only the decrease in the number of candidates, since the discarded litters were selected based on their parental EBV. Thus, the chances of detecting one of the best candidates from among the animals born in the X% worst litters would have been low anyway, especially for low values of X. To verify this, we repeated the GE_40% scenario simulations for the W1:1 goal but discarding 60% of the litters at random. As expected, the loss in response for trait 1 compared to GE_ref was greater (-24%) than when the litters' parental EBV was used (-12%). The moderate loss in selection response despite a marked reduction in the number of genotyped candidates observed in this study was consistent with the conclusions reached by Henryon et al. [30] who also showed by simulation that most of the benefits from GS in a pig nucleus breeding scheme, in terms of genetic gain, can be achieved by genotyping a small proportion of the available candidates selected based on their PEBV, including their own phenotype.

In our study, the prior information considered to target candidates to be genotyped was the parental EBV at birth, and piglets from the discarded litters were excluded from the pool of candidates, resulting in lower selection intensity. Another strategy would have been to phenotype for trait 1 also the piglets born in discarded litters and then jointly evaluate the genotyped and non-genotyped candidates using a 1-step GBLUP evaluation method [31,32]. This strategy is similar to that considered in [30] but with different prior information, and might have been even more efficient than the genomic scenarios that we applied, particularly in cases with a small number of genotyped candidates. First, the total number of candidates would have remained unchanged across the compared genomic scenarios, thus limiting the unfavourable impact on selection intensity of genotyping only a proportion of the candidates. Second, the genotyped animals would have had slightly more accurate GEBV for trait 1, resulting from additional records on non-genotyped related candidates. In that context, optimizing the proportion of male-to-female genotyping would be one way to further improve the efficiency of the genomic scenario [30]. The 1-step GBLUP strategy was, however, computationally too costly to be implemented in our simulations, and its additional benefits compared to the genomic scenarios simulated were therefore not evaluated.

At present, the cost of genotyping breeding animals with the porcine60SNP beadchip and of genotyping young animals for sufficient SNPs to minimize the imputation error rate (5000 or more [12]) is similar to the most expensive of our three genotyping costs scenarios. In that context, an additional annual expense of €350 000 to €530 000 (depending on the cost of phenotyping additional sibs and on the relative weights of the traits in the breeding goal) would be necessary to render it more efficient, in terms of genetic gains, to implement genomic evaluation rather

than improving the current pBLUP design. Thus, the threshold for implementation of GS is high, but probable reductions in genotyping costs in the future will gradually lower it. In addition, the results obtained by Huang *et al.* [24] suggest that it might be possible to genotype candidates and phenotyped sibs for only a few hundred markers, while preserving most of the efficiency of genomic evaluations and reducing individual genotyping costs to approximately \$25. The use of such optimized imputation strategies would lower the threshold for implementation of GS.

Extension to other populations

Our results were obtained by simulating a pig nucleus of 1050 breeding females and 50 boars, selected to improve a combination of two unrelated traits. The average accuracy of the PEBV of candidates is expected to be similar in populations of different sizes but with the same structure as simulated here, since a candidate would have approximately the same number of closely related phenotyped animals, regardless of population size. As a consequence, ignoring the effects of population size on selection intensity and the effects of inbreeding, a given pBLUP scenario should provide comparable genetic gains in either a smaller or larger nucleus. In contrast, the sizes of the TP in a genomic scheme with the same structure as here, which are composed of candidates or phenotyped sibs, would increase with size of the nucleus population. Therefore, larger or smaller selection nuclei than simulated here can be expected to result in higher or lower average accuracies of GEBV and realized genetic gains. The relative benefits of the different pBLUP and genomic scenarios considered here therefore depend, to some extent, on the size of the nucleus population. Our proposed methodology must be repeated in order to produce results relevant to the different populations being considered for implementation of genomic evaluations. Further work is also necessary to examine the benefits of GS with respect to more realistic breeding goals that combine several correlated traits.

Conclusions

Our results show that genomic selection could substantially increase genetic gains achieved in a purebred pig sire line breeding scheme organized around combined phenotyping of candidates and sibs with pBLUP evaluations, while significantly reducing the annual increase in inbreeding in the population. The potential improvements provided by implementing genomic evaluations significantly exceed the possible benefits of increasing phenotyping capacities for the most limited traits while retaining traditional pBLUP evaluations. Implementing a genomic breeding scheme would generate large additional costs. However, substantial savings could be achieved by pre-selecting the candidates to be genotyped based on

their parental EBV, and still maintain an advantage over the traditional pBLUP scheme. From an economic point of view, implementing genomic evaluations is not always the most efficient investment to improve the efficiency of a breeding scheme, but a threshold for additional annual expenses can be determined, below which increasing the number of phenotyped sibs while maintaining pBLUP evaluations is preferred, and above which genotyping animals to implement genomic evaluations is more efficient. This threshold depends on the cost of phenotyping additional sibs, on genotyping costs in combination with the imputation strategy, on the breeding goal and on the size of the nucleus population. Therefore, a breeding company must consider the relevance of implementing genomic evaluations based on the characteristics of the population and its possible annual financial commitment.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TT designed the simulation study, wrote the computer programs, ran the simulations, interpreted the results and wrote the article. FPH approved the methods used and the simulated scenarios, contributed to the interpretation of the results and critically revised the manuscript. CL approved the methods used and the simulated scenarios, and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank A. Legarra for supplying the program used for GBLUP evaluations, and I. Misztal for supplying the BLUPF90 software. The authors would also like to thank S. Lemarié for his valuable advice, V. Hawken for the linguistic review of the manuscript, and two anonymous reviewers whose comments contributed to improve the manuscript. The authors also wish to thank the ANR for its financial support (ANR_10_GENOM_BT_V_015).

Received: 7 March 2013 Accepted: 5 September 2013

Published: 15 October 2013

References

1. Meuwissen THE, Hayes BJ, Goddard ME: Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 2001, **157**:1819–1829.
2. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 2009, **92**:433–443.
3. Schaeffer LR: Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet* 2006, **123**:218–223.
4. Du FX, Clutter AC, Lohuis MM: Characterizing linkage disequilibrium in pig populations. *Int J Biol Sci* 2007, **3**:166–178.
5. Uimari P, Tapio M: Extent of linkage disequilibrium and effective population size in finnish landrace and finnish yorkshire pig breeds. *J Anim Sci* 2011, **89**:609–614.
6. Lillehammer M, Meuwissen THE, Sonesson AK: Genomic selection for maternal traits in pigs. *J Anim Sci* 2011, **89**:3908–3916.
7. Tribout T, Larzul C, Phocas F: Efficiency of genomic selection in a purebred pig male line. *J Anim Sci* 2012, **90**:4164–4176.
8. Hayes B, Goddard ME: The distribution of the effects of genes affecting quantitative traits in livestock. *Genet Sel Evol* 2001, **33**:209–229.
9. Quaas RL, Pollak EJ: Mixed model methodology for farm and ranch beef cattle testing programs. *J Anim Sci* 1980, **51**:1277–1287.
10. Meuwissen THE: Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet Sel Evol* 2009, **41**:35.
11. Boichard D: PEDIG: a FORTRAN package for pedigree analysis suited for large populations. Montpellier: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production: 19–23 August 2002; 2002:28–13.

12. Hickey JM, Kinghorn BP, Tier B, van der Werf JH, Cleveland MA: **A phasing and imputation method for pedigree populations that results in a single-stage genomic evaluation.** *Genet Sel Evol* 2012, **44**:9.
13. Wiggans GR, Cooper TA, VanRaden PM, Olson KM, Tooker ME: **Use of the illumina bovine3K beadchip in dairy genomic evaluation.** *J Dairy Sci* 2012, **95**:1552–1558.
14. Dassonneville R, Brondum RF, Druet T, Fritz S, Guillaume F, Gulbrandsen B, Lund MS, Ducrocq V, Su G: **Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations.** *J Dairy Sci* 2011, **94**:3679–3686.
15. Calus MPL, Veerkamp RF: **Accuracy of multi-trait genomic selection using different methods.** *Genet Sel Evol* 2011, **43**:26.
16. Jia Y, Jannink JL: **Multiple-trait genomic selection methods increase genetic value prediction accuracy.** *Genetics* 2012, **192**:1513–1522.
17. Coster A, Bastiaansen JWM, Calus MPL, van Arendonk JAM, Bovenhuis H: **Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance.** *Genet Sel Evol* 2010, **42**:9.
18. Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA: **The impact of genetic architecture on genome-wide evaluation methods.** *Genetics* 2010, **185**:1021–1031.
19. Meuwissen THE, Goddard ME: **Accurate prediction of genetic values for complex traits by whole-genome resequencing.** *Genetics* 2010, **185**:623–631.
20. Habier D, Tetens J, Seefried FR, Lichtner P, Thaller G: **The impact of genetic relationship information on genomic breeding values in German Holstein cattle.** *Genet Sel Evol* 2010, **42**:5.
21. Solberg TR, Sonesson AK, Woolliams JA, Odegard J, Meuwissen THE: **Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect.** *Genet Sel Evol* 2009, **41**:53.
22. Wolc A, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Habier D, Fernando R, Garrick DJ, Dekkers JCM: **Persistence of accuracy of genomic estimated breeding values over generations in layer chickens.** *Genet Sel Evol* 2011, **43**:23.
23. Daetwyler HD, Villanueva B, Bijma P, Woolliams JA: **Inbreeding in genome-wide selection.** *J Anim Breed Genet* 2007, **124**:369–376.
24. Huang Y, Hickey JM, Cleveland MA, Maltecca C: **Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost.** *Genet Sel Evol* 2012, **44**:25.
25. Yu X, Woolliams JA, Meuwissen THE: **Which animals are to be densely genotyped in order to impute the missing genotypes of sparsely genotyped animals.** Edinburgh: Proceedings of the 4th International Conference of Quantitative Genetics: 17–22 June 2012; 2012:165.
26. Kinsman A, Sargolzaei M, Schenkel F, Vender Voort G, Jafarikia M, Robinson A: **Accuracy of genotype imputation in Canadian Yorkshire pigs using FIMPUTE software.** Edinburgh: Proceedings of the 4th International Conference of Quantitative Genetics: 17–22 June 2012; 2012:192. <http://cdn.f1000.com/posters/docs/251407277>.
27. Goddard M: **Genomic selection: prediction of accuracy and maximisation of long term response.** *Genetica* 2009, **136**:245–257.
28. Van Grevenhof EM, Van Arendonk JA, Bijma P: **Response to genomic selection: the bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting.** *Genet Sel Evol* 2012, **44**:26.
29. Pszczola M, Strabel T, Mulder HA, Calus MP: **Reliability of direct genomic values for animals with different relationships within and to the reference population.** *J Dairy Sci* 2012, **95**:389–400.
30. Henryon M, Berg P, Ostensen T, Nielsen B, Sørensen AC: **Most of the benefits from genomic selection can be realized by genotyping a small proportion of available selection candidates.** *J Anim Sci* 2012, **90**:4681–4689.
31. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ: **Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score.** *J Dairy Sci* 2010, **93**:743–752.
32. Christensen OF, Lund MS: **Genomic prediction when some animals are not genotyped.** *Genet Sel Evol* 2010, **42**:2.

doi:10.1186/1297-9686-45-40

Cite this article as: Tribout *et al.*: Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. *Genetics Selection Evolution* 2013 **45**:40.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit



DISCUSSION GENERALE ET PERSPECTIVES

1. RESUME DES PRINCIPAUX RESULTATS

L'objectif de ce travail de thèse était d'évaluer l'intérêt sur un plan technique (précision, progrès génétique annuel attendu, rythme d'augmentation de la consanguinité dans la population sélectionnée) de mettre en place des évaluations génomiques dans les schémas de sélection porcine. Des éléments de réponse à cette question ont été apportés au moyen de simulations stochastiques, dans le cadre d'un programme de sélection d'une lignée porcine de type mâle de grande taille, organisé autour d'un contrôle combiné de candidats et de collatéraux.

Dans une première étude (Tribout et al., 2012), nous avons montré que la mise en place d'évaluations génomiques dans un tel programme de sélection en remplacement d'évaluations génétiques conventionnelles BLUP-MA pouvait permettre un gain important dans la précision des valeurs génétiques estimées (**VGE**), non seulement pour un caractère dont le phénotypage est limité car mesuré sur un petit nombre de collatéraux, mais également pour un caractère dont la mesure est réalisée directement sur l'ensemble des candidats à la sélection. Dans le cas d'un caractère mesuré sur les candidats, obtenir un gain de précision par rapport à une évaluation BLUP-MA nécessite toutefois de disposer d'une population de référence (**PR**) de grande taille, ce qui concrètement signifie que la mise en place d'évaluations génomiques efficaces ne permet pas de mettre fin au phénotypage des candidats. La sélection génomique a également un effet favorable sur le rythme d'augmentation de la consanguinité dans la population, car la valeur génomique estimée d'un jeune candidat contient une estimation plus précise de la part d'aléa de méiose de sa valeur génétique, et est donc moins dépendante de l'information familiale qu'une VGE BLUP-MA. Une évaluation génomique permet donc de mieux valoriser la variabilité génétique intra portée qu'une évaluation BLUP-MA. Toutefois, cette amélioration d'efficacité de la sélection se fait au prix d'un surcoût important du programme, compte tenu du grand nombre d'individus à génotyper pour constituer les populations de référence initiales et pour assurer le fonctionnement en routine du programme.

Dans une seconde étude (Tribout et al., 2013), nous avons montré que le surcoût lié à la mise en place d'évaluations génomiques dans le programme de sélection de notre lignée mâle de référence pouvait être largement maîtrisé en présélectionnant les candidats à génotyper sur la base de leur valeur génomique estimée sur ascendance. Nos résultats indiquent qu'il serait ainsi possible de réduire de manière significative le nombre de candidats à génotyper, tout en préservant une grande partie de l'avantage du programme de sélection génomique en termes de progrès génétique réalisé et de rythme d'augmentation de la consanguinité dans la

population pas rapport au programme de sélection BLUP-MA. Nous avons également montré qu'au-delà d'un seuil, une dépense supplémentaire pour améliorer l'efficacité du programme de sélection était plus efficacement investie dans la mise en place d'évaluations génomiques que dans l'augmentation de la capacité de phénotypage des collatéraux dans le dispositif conventionnel. Ce seuil d'intérêt de mise en place d'un programme génomique est d'autant plus bas que le coût individuel du génotypage est faible et que le coût de phénotypage des collatéraux est élevé. Les coûts de génotypage étant en diminution constante, il est certain que la solution génomique verra son intérêt se renforcer encore dans les années à venir. A ce jour, plusieurs entreprises de sélection ont d'ailleurs d'ores et déjà engagé des programmes de sélection génomique dans certaines de leurs lignées, parmi lesquelles les entreprises Danbred (http://www.danbred.com/News/~~/media/danbred/Collateral/Genomisk%20selektion%20UK_final.ashx), TOPIGS (<http://topigsusa.com/Insider/TOPIGS%20Insider%20Spring%202013.pdf>), et PIC (Cleveland et al., 2012).

2. LIMITES DE NOTRE MODELE STOCHASTIQUE

Nous avons vu dans la première partie de ce mémoire que la précision des valeurs génomiques estimées dépend d'un grand nombre de paramètres, dont certains ne sont pas pris en compte au travers des formules déterministes disponibles : déséquilibre de liaison entre marqueurs et QTL, structure de la PR, relations de parenté entre candidats et PR. Or, la précision est un paramètre clé de l'efficacité d'un programme de sélection. Afin d'estimer au mieux la précision des valeurs génomiques prédites dans un programme de sélection porcin, nous avons tenté, autant que possible, de simuler une population de structure réaliste conduite en bandes et en générations chevauchantes, avec des capacités de phénotypage des candidats et des collatéraux cohérentes, un génome présentant un déséquilibre de liaison comparable à celui observé dans des populations porcines réelles, et une densité de marqueurs SNP comparable à celle de la puce porcine SNP60 commercialisée par la société Illumina. Toutefois, des hypothèses et simplifications ont dû être faites, comme dans toute modélisation de programme de sélection, et celles-ci ont pu avoir un impact sur nos résultats et sur leur pertinence vis-à-vis des populations porcines réelles.

2.1. Longueur du génome simulé.

En premier lieu, il ne nous a pas été possible de simuler un génome de complet de 30M. En effet, en raison du grand nombre d'individus générés pendant les 15 années de

sélection simulées, de la densité de marqueurs considérée (1 500 marqueurs / M) et de la taille de la population de référence pour le caractère mesuré sur les candidats (jusqu'à 13 770 individus supplémentaires chaque année), le temps de calcul et les besoins en mémoire auraient été trop importants pour répliquer des simulations de plusieurs scénarios compte tenu des moyens informatiques dont nous disposons. Nous avons donc dû nous limiter à un génome de 10 M. Daetwyler et al. (2008) ont cependant montré que la précision des valeurs génomiques estimées en tirant profit du DL entre marqueurs et QTL est constante, pour une valeur d'héritabilité donnée, lorsque le ratio entre la taille de la PR et la longueur du génome est fixé. Afin de ne pas surestimer la précision de la sélection génomique en raison de ce génome trois fois trop court, nous avons donc simulé des PR trois fois plus petites que leur taille théorique, comme décrit dans Tribout et al. (2012).

Cependant, l'évaluation génomique tire non seulement parti du DL entre marqueurs et QTL, mais également des relations de parenté entre candidats et PR. En simulant des PR constituées d'un tiers des candidats et des collatéraux simulés afin de préserver la cohérence de la précision issue du DL, le nombre d'apparentés dont disposait chaque candidat dans les PR a également été réduit, ce qui a probablement eu pour conséquence de sous estimer la part de précision issue des relations de parenté, et donc la précision de la sélection dans les scénarios génomiques. Des simulations préliminaires avaient cependant été réalisées sur des génomes de 5 M et de 10 M en simulant des PR contenant, respectivement, un sixième et un tiers de leur taille théorique (Tribout et al., 2012). Une bonne cohérence de la précision moyenne des valeurs génomiques des candidats pour les deux caractères considérés dans l'étude avait été observée pour ces deux longueurs de génomes simulés. Par ailleurs, Habier et al. (2013) ont montré par simulation que la part de la précision des valeurs génomiques estimées issue de la parenté entre candidats et PR tend à diminuer lorsque la taille de la PR augmente. Dans nos simulations, les PR considérées étaient de grande taille (en particulier pour le caractère mesuré sur les candidats), et on peut donc supposer que la sous estimation de la précision due au nombre plus faible qu'attendu d'apparentés proches des candidats dans la PR est limitée.

2.2. Génération du DL dans la population de base.

Plusieurs types de méthodes ont été mises en œuvre dans la littérature pour simuler le génome des individus d'une population d'élevage, comme présenté dans la synthèse de Daetwyler et al. (2013). La pratique la plus couramment utilisée consiste à simuler des accouplements aléatoires entre des reproducteurs pendant un grand nombre de générations, de manière à générer un déséquilibre de liaison entre loci. Le nombre de générations réalisées, le

nombre de reproducteurs impliqués au cours des générations ainsi que le taux de mutation appliqué varient cependant fortement entre les études, sans que les raisons du choix des valeurs de ces paramètres soient clairement explicitées. Dans notre cas, notre objectif était de générer un profil de déséquilibre de liaison comparable à celui observé dans les populations porcines. Pour cela, nous avons utilisé un modèle très simple, s'appuyant uniquement sur la dérive, sans mutation, en déterminant empiriquement le nombre de reproducteurs et le nombre de générations nécessaires à la génération du profil de DL recherché. En particulier, les populations porcines se distinguent, par rapport aux populations bovines par exemple, par une persistance élevée du DL (r^2 de l'ordre de 0,15 à 0,20 à une distance de 1 Mb), que nous avons reproduite en réalisant un goulot d'étranglement dans le nombre de reproducteurs 35 générations avant le début de la sélection.

Il est évident que le modèle utilisé pour générer les chromosomes des fondateurs de l'élevage de notre étude ne rend pas compte de la véritable histoire d'une population porcine. Sa seule finalité était de générer une population de départ dont le génome présentait une structure de DL ressemblant à celle d'une population en cours de sélection, et cet objectif a été atteint. La littérature concernant l'impact de la méthode employée pour générer des génomes sur les résultats de simulations de populations sélectionnées est cependant rare. Afin de tester la sensibilité de nos résultats à ce facteur, il serait intéressant de reproduire nos simulations en générant les chromosomes des fondateurs sous un modèle différent, voire même en partant non pas de chromosomes simulés mais des véritables génotypes des reproducteurs en activité d'un noyau de sélection.

2.3. Diminution de la variabilité génétique dans la population.

Comme discuté dans Tribout et al. (2012), une très forte diminution de la variance génétique a été observée dans nos simulations pour les deux caractères sélectionnés, quelle que soit la modalité d'évaluation génétique de la population (BLUP-MA ou GBLUP). Cette chute de variance génétique était d'autant plus forte que le nombre de QTL simulés était faible. L'ampleur de ce phénomène constitue un écart à ce qui peut être observé dans les noyaux de sélection réels sélectionnés depuis plusieurs décennies souvent de manière intense et dans lesquels la variabilité génétique reste pourtant relativement stable. Cette diminution de variance génétique s'est traduite dans nos simulations par une diminution de l'héritabilité des caractères au cours du temps, qui a progressivement réduit la précision d'estimation des VGE et donc le progrès génétique réalisé dans les différents scénarios par rapport à une situation idéale dans laquelle la variabilité génétique aurait été mieux préservée.

Cette diminution de la variabilité génétique a une double origine. Tout d'abord, le travail de sélection conduit évidemment à sélectionner des individus porteurs d'allèles favorables aux QTL et à les faire se reproduire. La fixation progressive des allèles favorables aux QTL entraîne nécessairement une diminution de la variance génétique dans la population. De fortes diminutions de variabilité génétique ont également été rapportées par Jannink (2010) et Nielsen et al. (2011) qui ont respectivement simulé 24 générations de sélection sur de l'orge et 10 générations de sélection dans une population de poissons. Par ailleurs, nous n'avons pas tenu compte de l'apparentement entre individus lors de la sélection des candidats et lors des accouplements entre reproducteurs. En raison de la sévérité des taux de sélection pratiqués et du nombre limité de reproducteurs (en particulier de reproducteurs mâles), cette simplification dans la gestion de la population a eu pour conséquence de faire augmenter la consanguinité dans la population beaucoup plus rapidement que dans les noyaux de sélection porcins réels, et d'accélérer la diminution de la variance génétique des caractères.

Plusieurs voies d'amélioration de notre modèle sont envisageables. Concernant les règles de gestion de la population, il serait tout d'abord possible d'introduire une contrainte sur l'augmentation de la consanguinité dans la population au moment de la sélection des jeunes reproducteurs mâles et femelles de renouvellement par la méthode des contributions optimales (Meuwissen, 1997), comme cela a été réalisé dans les simulations de Nielsen et al. (2011). Notre modélisation du déterminisme génétique des caractères pourrait également être améliorée, en simulant par exemple un génome plus long portant un plus grand nombre de QTL dont la fixation serait plus lente, et en faisant intervenir des mutations ainsi que des effets de dominance et d'épistasie qui semblent jouer un rôle dans le maintien de la variabilité génétique (Caballero et Keightley, 1994 ; Noguera et al., 2006).

2.4. Objectif de sélection et caractères considérés.

L'objectif de ce travail était d'évaluer les conséquences de la mise en place d'évaluations génomiques dans une lignée mâle organisée autour d'un contrôle combiné. Pour cela, nous avons simulé une population sélectionnée sur une combinaison de deux caractères, l'un mesuré sur un grand nombre de candidats, l'autre mesuré sur un nombre limité de collatéraux, ce qui, à notre connaissance, n'avait pas été réalisé de manière stochastique jusqu'à présent. Toutefois, notre modèle ne peut prétendre pouvoir représenter la complexité des populations porcines réelles, dont l'objectif de sélection est généralement composé d'un plus grand nombre de caractères plus ou moins héréditaires et corrélés entre eux. Nos simulations ne peuvent donc fournir que de grandes tendances. L'obtention de résultats plus

précis, dans le but par exemple d'optimiser l'organisation d'un programme de sélection génomique dans une population donnée, nécessiterait un modèle d'un niveau de complexité qu'il semble difficile à atteindre. D'une part, le déterminisme génétique réel des caractères d'intérêt chez le porc est actuellement mal connu. Modéliser plus de deux caractères présentant une structure de variance-covariance imposée nécessiterait donc de faire un grand nombre d'hypothèses (nombre et distribution des effets des QTL affectant chaque caractère, proportions de QTL pléiotropiques, ...) dont les conséquences sur les résultats des simulations devraient être étudiées, ce qui nécessiterait de simuler un grand nombre de cas. D'autre part, le temps nécessaire à la réalisation d'évaluations génomiques dans chaque bande et pour chaque caractère serait un facteur très limitant.

2.5. Modèle d'évaluation génétique.

Dans nos simulations, les évaluations génomiques ont été réalisées en estimant les effets des SNP par la méthode du GBLUP (Meuwissen et al., 2001), puis en utilisant ces estimations pour calculer la valeur génomique des candidats et des reproducteurs génotypés. Cette méthode a été choisie pour sa simplicité de mise en œuvre et sa rapidité d'exécution bien adaptée à des simulations. Afin de limiter les temps de calcul, nous avons cependant dû faire plusieurs simplifications par rapport à un programme de sélection réel. Tout d'abord, les effets des SNP n'ont été estimés qu'une fois par an. En réalité, on aurait intérêt à ré-estimer ces effets plus fréquemment, dans l'idéal dans chaque bande, afin de tirer profit des phénotypes des nouveaux candidats et collatéraux génotypés et d'accroître la précision de la sélection tout au long de l'année. Par ailleurs, les scénarios GE_X% se sont traduits par une diminution du nombre de candidats par rapport au scénario de référence. En réalité, il serait probablement plus avantageux, en particulier lorsque la proportion de candidats génotypée est faible, de continuer à phénotyper les individus non génotypés pour disposer d'une estimation de leur valeur génétique BLUP-MA, afin de réaliser la sélection parmi l'ensemble des candidats génotypés et non génotypés.

En pratique, parmi les différents modèles d'évaluation génomique disponibles, la méthode du "single-step GBLUP" (**SSGBLUP**) (Legarra et al., 2009 ; Aguilar et al., 2010 ; Christensen et Lund, 2010) semble être particulièrement adaptée à la situation d'un programme de sélection porcin. Cette méthode est déjà appliquée dans certaines populations porcines (Cleveland et al., 2012). Le SSGBLUP est une extension de la méthode du BLUP génomique à une situation où une partie des individus n'est pas génotypée. Elle consiste à estimer simultanément la valeur génétique de l'ensemble des individus de la population en

résolvant un système analogue au système des équations du modèle mixte obtenu lorsque les performances sont décrites avec un modèle animal individuel. Dans ce système, la matrice de parenté entre individus est une matrice « hybride » combinant les coefficients de parenté génomique entre individus génotypés aux coefficients de parenté calculés à partir du pedigree pour l'ensemble des individus.

Le premier avantage de cette méthode est qu'elle permet d'estimer des valeurs génétiques directement comparables pour l'ensemble des individus de la population, qu'ils soient génotypés ou non, et donc d'effectuer la sélection sur l'ensemble des candidats, préservant l'intensité de sélection. Christensen et al. (2012) ont appliqué ce modèle dans une population porcine, et ont observé une augmentation de la précision des VGE des individus non génotypés par rapport à une évaluation BLUP-MA traditionnelle. La mise en œuvre d'évaluations s'appuyant sur un SSGBLUP permettrait donc d'augmenter la précision de la sélection pour l'ensemble des candidats par rapport à la situation actuelle, contrairement aux autres modèles d'évaluation génomique avec lesquels seuls les animaux génotypés bénéficieraient d'un gain de précision.

Le SSGBLUP peut être appliqué aux mêmes modèles qu'un BLUP-MA classique. En particulier, la mise en œuvre d'évaluations génomiques multicaractères est possible. Calus et Veerkamp (2011) et Jia et Jannink (2012) ont montré par simulation que des évaluations génomiques multi-caractères pouvaient apporter un gain de précision par rapport à des évaluations mono-caractères, en particulier pour des caractères peu héritables et fortement corrélés à des caractères ayant une héritabilité élevée. Dans le cas d'un programme porcin organisé autour d'un contrôle combiné, disposer d'évaluations génomiques multicaractères serait particulièrement avantageux pour augmenter la précision des VGE pour les caractères mesurés sur collatéraux dont la capacité de phénotypage est limitée, mais qui sont corrélés aux phénotypes mesurés sur l'ensemble des candidats (Calus et Veerkamp, 2011).

Le point faible de la méthode SSGBLUP est qu'elle fait l'hypothèse que les effets de tous les segments chromosomiques suivent une loi normale de même variance. Or, une partie des caractères d'intérêt en sélection porcine, comme la qualité de la viande et la quantité de muscle de la carcasse dans certaines populations Piétrain, sont sous l'effet d'un gène majeur de sensibilité à l'halothane. Pour ces caractères, les méthodes bayésiennes, permettant les écarts à la normalité des effets, seraient mieux adaptées et produiraient des VGE plus précises. Il serait toutefois possible d'inclure en tant qu'effet fixe dans le modèle d'évaluation d'un SSGBLUP le génotype au gène de sensibilité à l'halothane des individus évalués pour les

populations concernées, ainsi que le génotype pour d'éventuels autres QTL à effet fort qui pourraient être mis en évidence à l'avenir. Ainsi, les effets de ces SNP ne seraient plus soumis à la contrainte de normalité inhérente à cette méthode, permettant une meilleure précision d'estimation des valeurs génomiques.

La réalisation d'une évaluation génomique multicaractère s'appuyant sur le SSGBLUP est cependant nettement plus coûteuse en temps de calcul qu'une évaluation BLUP-MA traditionnelle. Le point limitant de cette méthode est le calcul et le stockage de l'inverse de la matrice de parenté génomique et d'une partie de la matrice de parenté établie à partir des pedigrees, dont la taille dépend du nombre d'individus génotypés pris en compte dans l'évaluation. Tsuruta et al. (2011) ont déjà mis en œuvre cette méthode sur une population bovine laitière américaine comptant 17 000 individus génotypés en évaluant simultanément 18 caractères. Cependant, le nombre d'individus génotypés dans une population porcine de grande taille comme les populations Piétrain, Large White type femelle et Landrace collectives françaises dépasserait rapidement plusieurs dizaines de milliers d'individus, dépassant peut-être les limites actuelles des logiciels s'appuyant sur ce modèle. Plusieurs stratégies ne nécessitant pas le calcul explicite des matrices de parenté sont toutefois envisageables (Legarra et Ducrocq, 2012), et des travaux sont actuellement en cours pour rendre ce modèle applicable à des évaluations en routines de populations contenant un grand nombre d'individus génotypés (Meyer et al., 2013).

3. RENTABILITE ECONOMIQUE DE LA MISE EN PLACE D'EVALUATIONS GENOMIQUES DANS LA POPULATION PIETRAIN COLLECTIVE FRANCAISE

Le verrat terminal actuellement le plus utilisé en France est le verrat Piétrain dit « collectif » (PC). Le programme actuel de sélection de la population PC ressemble au scénario BL_ref qui a été modélisé dans notre étude, bien que la population PC soit toutefois environ 45% plus grande que celle de nos simulations (1 500 femelles reproductrices au lieu de 1 050, et 20 000 candidats par an au lieu de 13 770). Au vu de nos résultats, on peut donc penser que la mise en place d'évaluations génomiques dans cette population permettrait d'y augmenter significativement le progrès génétique annuel réalisé sur l'objectif global de sélection (ΔGa_{bg}), tout en facilitant la maîtrise de l'augmentation de la consanguinité.

Un prélèvement systématique de tissus est réalisé depuis 2011 sur tous les animaux contrôlés dans les stations publiques françaises. Parmi les échantillons déjà prélevés, près d'un millier correspond à des individus PC, ce qui permettrait de constituer une population de référence de départ pour les caractères mesurés sur les collatéraux. Par ailleurs, deux puces porcines, l'une de 60 000 marqueurs et l'autre de 8 000 marqueurs SNP, sont commercialisées. Cette dernière puce pourrait être utilisée pour génotyper à moindre coût les candidats du contrôle en ferme et faire de l'imputation de génotype. La mise en place d'un programme de sélection génomique dans la population PC semble donc envisageable.

Une question importante n'a cependant pas été abordée dans nos simulations : l'investissement réalisé en sélection pour mettre en place des évaluations génomiques serait-il économiquement rentable ? En effet, augmenter l'efficacité d'un programme de sélection n'a d'intérêt que si les profits supplémentaires générés excèdent le surcoût lié à cette amélioration. Des études économiques approfondies sont évidemment nécessaires pour étudier la rentabilité d'un tel investissement. Quelques éléments de réflexion peuvent cependant être avancés.

Nous avons tenté d'évaluer la rentabilité de la mise en place dans la population PC d'un programme de sélection génomique organisé sur le modèle des scénarios GE_X% simulés. Pour cela, nous avons estimé l'augmentation minimum de ΔGa_{bg} en sélection nécessaire pour que l'accroissement de marge générée dans la filière cumulée sur les N premières années de fonctionnement du programme génomique compense le coût de constitution des populations de référence initiales et le coût du génotypage en routine des candidats et des reproducteurs au cours de ces N années. A partir de la N+1^{ème} année, le programme de sélection génomique deviendrait alors rentable, ce moment correspondant au « point mort » du programme.

Surcoût du programme génomique.

Le surcoût généré par un programme d'évaluation génomique dans la population PC a été estimé comme décrit dans Tribout et al. (2013), en considérant que le coût de constitution des deux populations de référence initiales (CPR_{ini}) est amorti sur N années. Le surcoût cumulé sur les N premières années de fonctionnement du programme (SC_N) est donc égal à la somme de CPR_{ini} et du coût de génotypage des candidats, des collatéraux et des reproducteurs de renouvellement pendant les N années considérées. Comme dans nos simulations, nous avons considéré que les jeunes individus étaient génotypés pour une puce de faible densité, et

que les candidats sélectionnés pour renouveler le noyau de sélection étaient re-génotypés pour la puce porcine SNP60 afin de maximiser la qualité de l'imputation des génotypes des jeunes individus.

Sous ces hypothèses,

$$SC_N = HD_{pr} * [TP2_0 + (N + 1) * (NbFR_a + NbMR_a)] + LD_{pr} * [TP1_0 + N * (NbCand_a + NbColl_a)]$$

où

- HD_{pr} et LD_{pr} sont, respectivement, les coûts individuels de génotypage pour les puces haute et basse densités ;

- $TP1_0$ et $NbCand_a$ sont, respectivement, la taille initiale de la population de référence pour les caractères mesurés sur candidats et le nombre annuel de candidats génotypés ;

- $TP2_0$ et $NbColl_a$ sont, respectivement, la taille initiale de la population de référence pour les caractères mesurés sur collatéraux et le nombre annuel de collatéraux génotypés ;

- $NbFR_a$ et $NbMR_a$ sont, respectivement, les nombres de jeunes candidats femelles et mâles sélectionnés chaque année pour le renouvellement du noyau de sélection.

Après ces N années, le coût de constitution des populations de références initiales ayant été amorti, le surcoût annuel (S_a) lié au fonctionnement en routine du programme génomique se limiterait au coût de génotypage des candidats, des collatéraux et des reproducteurs de renouvellement de l'année :

$$S_a = HD_{pr} * (NbFR_a + NbMR_a) + LD_{pr} * (NbCand_a + NbColl_a)$$

Seuil de rentabilité du programme génomique.

En France, environ 15 millions de porcs charcutiers par an sont issus de verrats PC. Compte tenu de la disproportion entre les effectifs annuels de porcs PC de l'étage de sélection abattus et de porcs charcutiers issus d'un verroat PC, seul le gain économique réalisé dans les élevages de production a été considéré dans nos calculs.

Le progrès génétique réalisé dans la population PC grâce au programme de sélection actuel génère une augmentation de marge en élevage de production (ΔM_{a_prod}) estimée à 0,367 €/porc charcutier/an (voir détails en annexe 1). En supposant :

- que la mise en place d'évaluations génomiques augmente de K% le ΔGa_bg dans le noyau de sélection PC ;

- que cette amélioration d'efficacité soit intégralement convertie en augmentation de ΔM_{a_prod} ;
- que l'efficacité de la sélection et la valeur économique des caractères soient constantes au cours du temps,

et en tenant compte du fait que les premiers porcs charcutiers dont le père a été sélectionné sur la base de sa valeur génomique seraient abattus un an après la mise en place des évaluations génomiques, alors le supplément d'augmentation de marge dans les élevages de production par rapport à la situation actuelle, cumulé sur les N premières années de fonctionnement du programme génomique, serait donc égal à :

$$15\,000\,000 * (N - 1) * \frac{(N-1)*K*\Delta M_{a_prod}}{2} \text{ € , comme illustré sur la figure 1.}$$

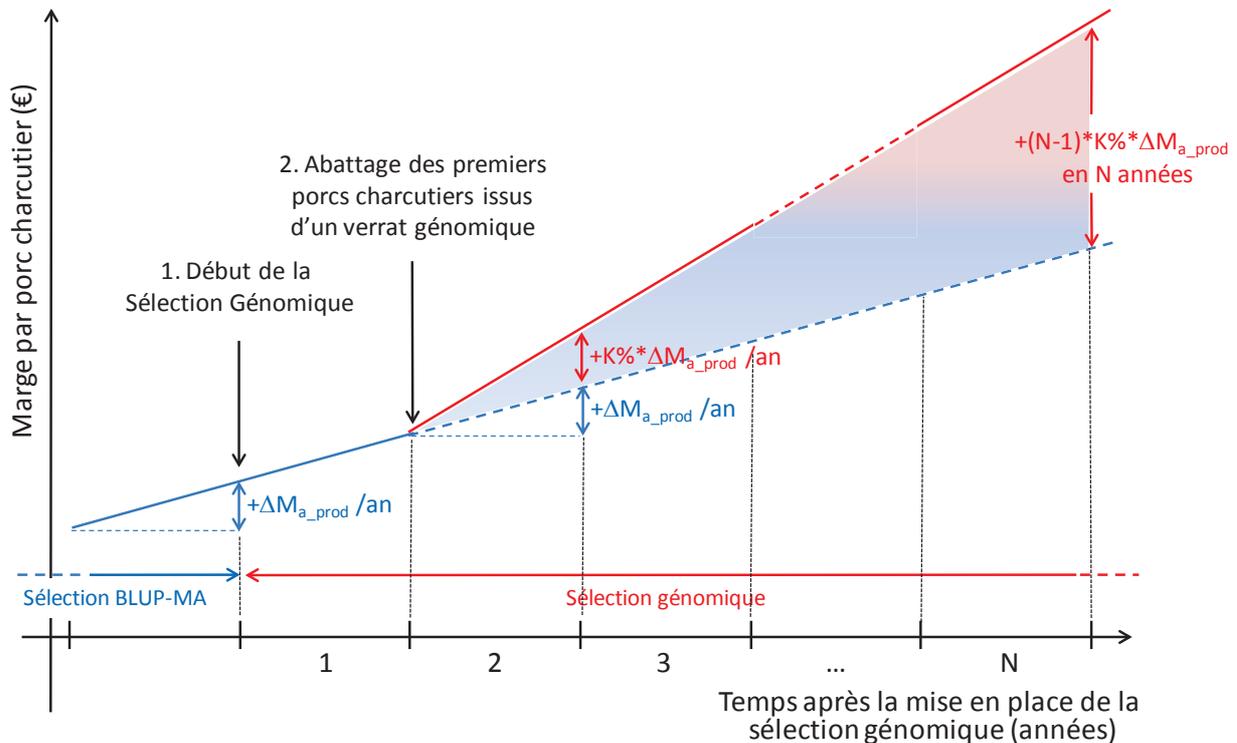


Figure 1 : Augmentation au cours du temps de la marge par porc charcutier en élevage de production, lorsque le programme de sélection de la population dont est issu le verrat terminal s'appuie sur des évaluations BLUP-MA (avant l'année 1) ou sur des évaluations génomiques (à partir de l'année 1). L'augmentation cumulée de marge au cours des N premières années de fonctionnement du programme de sélection génomique correspond à la surface de la zone colorée.

L'investissement en sélection serait donc compensé en N années par une augmentation

supplémentaire de marge en production si $K * (15\,000\,000 * \Delta M_{a_prod} * (N - 1)^2 / 2) \geq SC_N$, c'est-à-dire, si $K \geq K_{min}$,

$$\text{où } K_{min} = \frac{HD_{pr} * [TP2_0 + (N+1) * (NbFR_a + NbMR_a)] + LD_{pr} * [TP1_0 + N * (NbCand_a + NbColl_a)]}{15\,000\,000 * \Delta M_{a_prod} * (N-1)^2 / 2} \quad (1)$$

Pour la population PC, $NbFR_a = 1\,250$ femelles de renouvellement/an, $NbMR_a = 200$ nouveaux verrats/an et $NbColl_a = 400$ collatéraux/an. Nous avons fait l'hypothèse que la population de référence initiale pour les caractères mesurés en station ($TP2_0$) compte 1 000 individus. Comme dans les scénarios génomiques GE_X% simulés dans Tribout et al. (2013), $TP1_0$ et $NbCand_a$ ont été supposés égaux, les $NbCand_a$ candidats génotypés correspondant à une fraction variable des 20 000 candidats annuels de l'actuel programme de sélection PC, présélectionnés sur la base de leur valeur sur ascendance. Nous avons considéré que $HD_{pr} = 110\text{€}$ et $LD_{pr} = 40\text{€}$. Le calcul de K_{min} a été réalisé en faisant varier le nombre d'années N nécessaires à la rentabilité du programme génomique. Par simplicité, et compte tenu du caractère approximatif du raisonnement, nous n'avons pas appliqué de taux d'actualisation aux coûts.

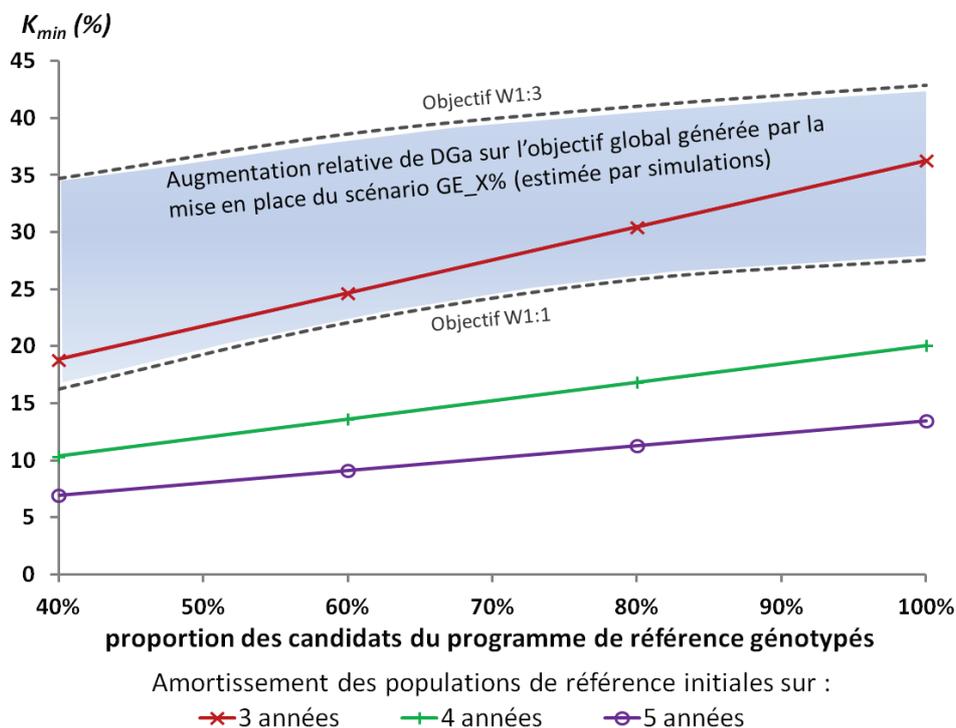


Figure 2 : Accroissement de ΔGa_{bg} par rapport au programme de sélection conventionnel (K_{min} , en %) nécessaire pour rentabiliser en 3, 4 ou 5 années la mise en place d'une sélection génomique dans la population Piétrain collective française (lignes continues), en fonction de la proportion de candidats génotypés annuellement.

La figure 2 ci-dessus indique l'accroissement K_{min} de ΔGa_{bg} nécessaire pour que le programme de sélection génomique génère un bénéfice supplémentaire par rapport au programme de sélection conventionnel dès sa N+1^{ème} année de fonctionnement.

L'objectif de sélection réel de la population PC n'intègre pas directement de caractère mesuré sur les candidats ; il est en fait composé de caractères mesurés sur les collatéraux, dont certains sont cependant fortement corrélés aux caractères mesurés sur les candidats. On peut toutefois estimer que cet objectif représente une situation intermédiaire à celle des objectifs W1:1 et W1:3 considérés dans nos simulations (dans lesquels le caractère mesuré sur collatéraux avait, respectivement, le même poids économique ou un poids économique trois fois plus élevé que celui du caractère mesuré sur les candidats à la sélection), et que l'accroissement de ΔGa_{bg} attendu pour la population PC se situerait vraisemblablement dans la zone grisée présentée sur la figure 2, délimitée par les augmentations de ΔGa_{bg} estimées par simulation pour ces deux objectifs. Au vu de ces résultats, et en supposant que seulement 50% des 20 000 candidats annuels soient génotypés, un accroissement du ΔGa_{bg} de 22% par rapport à celui obtenu dans le programme actuel suffirait à assurer en trois ans la rentabilité de la mise en place du programme de sélection génomique, tandis qu'un accroissement du ΔGa_{bg} de 12% seulement assurerait sa rentabilité en quatre ans. Au vu des résultats de nos simulations, l'obtention de telles améliorations de l'efficacité de la sélection semble possible. Il est donc probable qu'un programme de sélection génomique dans la population PC serait économiquement rentable en moins de quatre ans.

En restant sous ces hypothèses du génotypage de 10 000 candidats par an et d'un amortissement sur trois ans des coûts de mise en place des évaluations génomiques, le surcoût du programme de sélection génomique de la population PC s'élèverait à 800 k€/an pendant trois ans, puis se limiterait à 575 k€ à partir de la quatrième année de fonctionnement. Ce surcoût initial de 800 k€/an pourrait être financé en augmentant par exemple le prix de la semence vendue aux producteurs de 0,25€ par dose (soit une augmentation d'environ 4% du prix moyen actuel de la dose de semence), ou en prélevant une somme de 0,053€ par porc charcutier commercialisé. Une étude économique plus poussée et impliquant les partenaires de la filière est bien évidemment nécessaire pour déterminer les modalités de financement de l'investissement en sélection pouvant être effectivement envisagées sans pénaliser la semence de verrats PC par rapport à celle des types génétiques concurrents.

Dans les calculs précédents, nous avons considéré que le progrès génétique réalisé en production est égal au progrès génétique théoriquement attendu, ce qui est toutefois peu vraisemblable (Merks, 2001). En effet, les conditions d'élevage diffèrent sensiblement entre élevages de sélection et de production, et l'existence d'interactions génotype x milieu est probable. De plus, les porcs charcutiers sont des animaux croisés chez qui des effets génétiques non additifs (dominance, épistasie) pourraient limiter l'expression du progrès génétique réalisé dans la population PC. Par ailleurs, nous n'avons pas non plus pris en compte le fait que le remplacement des verrats PC terminaux par des reproducteurs mâles issus de sélection génomique se ferait progressivement (la durée de carrière d'un verrat terminal est d'environ 18 mois). Le retour sur investissement d'un programme génomique dans la population PC pourrait donc être moins rapide que suggéré par les estimations précédentes. Toutefois, nos hypothèses concernant le prix des puces (110€ et 40€ pour les puces 60K et faible densité, respectivement) sont prudentes. En effet, le tarif de la puce porcineSNP60 a baissé d'environ 25% depuis 2012, et il est très probable que les coûts de génotypage diminueront encore nettement à l'avenir. Par ailleurs, la création d'un consortium regroupant les entreprises de sélection impliquées dans des programmes de sélection génomique porcins par l'intermédiaire duquel seraient passées les commandes de puces (sur le modèle de la structure VALOGENE chez les bovins) permettrait de réduire sensiblement les coûts de génotypage et donc de réduire le seuil de rentabilité économique d'un programme génomique.

4. INTERET DE LA SELECTION GENOMIQUE DANS LES POPULATIONS PORCINES MATERNELLES

Dans nos simulations, nous avons uniquement considéré le cas d'une population sélectionnée de type mâle. Une partie des programmes de sélection concerne cependant des populations porcines maternelles, qui sont sélectionnées pour améliorer une combinaison de caractères d'engraissement, de carcasse et d'aptitudes maternelles. Dans la plupart des cas, ces derniers caractères ont un poids économique important, pouvant représenter jusqu'à 50% de l'objectif de sélection de la population. Les phénotypes d'aptitudes maternelles, comme la taille de portée à la naissance ou au sevrage, ne sont observables que sur les femelles reproductrices. Les candidats n'ont donc pas de performance propre pour ces caractères au moment de la sélection et disposent uniquement dans les programmes actuels d'une valeur génétique estimée sur ascendance, identique pour tous les candidats d'une même portée. De

plus, les aptitudes maternelles sont en général peu héritables ($h^2 \sim 0,10$), si bien que la précision des VGE des candidats est faible. La sélection des candidates femelles après un premier cycle de reproduction et l'organisation d'un testage sur descendance pour les candidats mâles permettraient certes d'améliorer la précision des VGE et d'estimer la part d'aléa de méiose des candidats pour les caractères d'aptitude maternelle. Cependant, l'amélioration obtenue sur la précision des VGE serait contrebalancée par une forte augmentation de l'intervalle entre générations, rendant cette solution sans intérêt en pratique (Lillehammer et al., 2011b).

On retrouve donc les mêmes problématiques pour les caractères d'aptitudes maternelles que pour les caractères dont le phénotype est mesuré uniquement sur des collatéraux, à savoir une faible précision d'évaluation entraînant une réponse à la sélection limitée, et une mauvaise gestion de la variabilité intra-famille qui conduit à sélectionner des candidats apparentés et tend à accélérer l'augmentation de la consanguinité dans la population. Les bénéfices attendus par la mise en place d'évaluations génomiques pour les caractères d'aptitudes maternelles sont donc potentiellement importants, en termes de précision d'évaluation, de réponse à la sélection et de gestion de la variabilité génétique, comme montré par Lillehammer et al. (2011b).

Les noyaux de sélection des populations porcines Large White type femelle (LWC) et Landrace (LC) collectives françaises comptent respectivement 2 700 et 1 300 truies reproductrices. Chaque année, le nombre de femelles de renouvellement, de candidats à la sélection et de collatéraux s'élèvent approximativement à 2 000 femelles, 32 000 candidats et 740 collatéraux, respectivement, dans la population LWC, et à 1 000 femelles, 16 700 candidats et 480 collatéraux, respectivement, dans la population LC. Avec de tels effectifs, il est vraisemblable que les populations de référence constituées seraient suffisamment grandes pour augmenter de manière importante la précision d'estimation des VGE pour les caractères d'aptitude maternelle et de production par rapport aux programmes de sélection actuels. Un programme de sélection génomique permettrait donc probablement d'accroître de manière importante le progrès génétique réalisé dans ces populations. Des simulations complémentaires seraient nécessaires pour évaluer l'ampleur de ce gain.

Contrairement à la population PC dont les mâles sont utilisés comme verrat terminal, les populations LWC et LC sont utilisées en croisement pour produire une truie parentale qui bénéficie de 50% du progrès génétique réalisé dans chacun des deux noyaux de sélection. En conséquence, 25% seulement du progrès génétique créé sur les caractères d'engraissement et de carcasse dans les noyaux de sélection LWC et LC sont théoriquement transmis aux porcs

charcutiers des élevages de production, contre 50% pour la population PC. De plus, le nombre annuel de porcs charcutiers produits à partir de reproducteurs issus des noyaux de sélection LWC et LC est plus faible que le nombre de porcs charcutiers nés d'un verrat terminal PC. En conséquence, on peut penser que le temps nécessaire pour rentabiliser la mise en place d'un programme de sélection génomique dans les populations LWC et LC serait plus long que dans le cas de la population PC.

5. INTERET DE LA SELECTION GENOMIQUE DANS LES POPULATIONS PORCINES DE FAIBLE EFFECTIF

Dans nos simulations, nous avons considéré le cas d'une population sélectionnée de grande taille. De nombreuses lignées et variétés de race paternelles et maternelles comptent cependant moins de 200 femelles reproductrices, et l'intérêt de mettre en place des évaluations génomiques dans ces populations doit être spécifiquement évalué. En effet, on peut craindre que la taille des populations de référence qu'il serait possible de constituer dans une population de petit effectif serait trop faible pour générer un gain de progrès génétique suffisant pour compenser le surcoût lié aux génotypages. Des simulations complémentaires permettraient d'apporter des éléments de réponse à cette question.

Une solution envisageable en théorie pour accroître la précision des valeurs génomiques estimées dans des populations d'effectifs limités pourrait être de regrouper les individus phénotypés et génotypés de plusieurs types génétiques, afin de constituer une population de référence multiraciale de plus grande taille qu'en travaillant intra race. Cependant, les diverses tentatives d'utilisation de populations de référence multiraciales se sont jusqu'à présent révélées relativement décevantes. Par exemple, Erbe et al. (2012) n'ont obtenu qu'un accroissement très faible - voire même une diminution pour certains caractères - de la précision des valeurs génomiques estimées d'individus de race Jersey en ajoutant 2 257 taureaux Holstein testés sur descendance à une population de référence constituée de 540 taureaux Jersey. De même la réunion de deux populations de référence ovines Lacaune françaises n'a pas permis d'améliorer la précision des valeurs génomiques estimées dans chacune de ces deux sous-populations (Astruc, communication personnelle).

En fait, l'évaluation génomique exploite deux sources d'information : le déséquilibre de liaison entre marqueurs et QTL, et la parenté entre les candidats évalués et les individus de la population de référence. Une population de référence d'un type génétique différent de celui

des candidats évalués n'apporte évidemment aucun nouvel apparenté proche pour les candidats. Un gain de précision des valeurs génomiques estimées ne peut donc être obtenu que si des associations communes entre allèles aux QTL et allèles aux marqueurs sont retrouvées dans les différentes populations réunies. Il faut donc que les segments conservés entre populations soient nettement plus longs que l'intervalle entre marqueurs. On peut donc penser qu'une population de référence multiraciale ne serait efficace que pour des populations génétiquement proches comme des variétés d'une même race ayant divergé depuis peu de temps, ainsi que dans les lignées synthétiques récentes. Une hypothèse supplémentaire est bien sûr la ségrégation de QTL communs entre les populations. Les QTL fixés dans une population, ou bien les QTL issus de mutations récentes ne peuvent bien sûr pas contribuer à l'évaluation entre populations et ne peuvent être utilisés qu'intra population. Pour des populations ayant divergé depuis plus longtemps, ou pour des lignées synthétiques de compositions différentes, une densité de marqueurs supérieure à la densité de la puce porcineSNP60 serait nécessaire, ce qui augmenterait toutefois les coûts de génotypage.

6. SELECTION GENOMIQUE POUR LA VALEUR EN CROISEMENT

Une des principales faiblesses des dispositifs d'amélioration génétique porcins actuels tient au fait que la sélection y est le plus souvent réalisée en race pure à partir des phénotypes d'individus élevés dans un milieu très maîtrisé, alors que l'objectif de la sélection est d'améliorer les performances d'individus croisés élevés dans des conditions de production beaucoup moins contrôlées. Les fortes différences d'environnement pouvant exister entre élevages de production et élevages de sélection ou stations de contrôle, ainsi que les différences génétiques entre individus croisés et individus de race pure font que la transmission du progrès génétique créé en sélection vers l'étage de production n'est que partielle.

Comme déjà discuté dans Tribout (2011), des évaluations génomiques s'appuyant sur des populations de référence constituées de truies parentales et de porcs charcutiers phénotypés en élevage de production permettraient, en théorie, d'augmenter fortement la corrélation entre la valeur génétique estimée des candidats et la performance moyenne de leurs descendants croisés en élevage de production. Grâce à une telle sélection en race pure pour la valeur en croisement, l'amélioration des performances des populations commerciale serait plus rapide.

Plusieurs modèles d'évaluation génomique ont été proposés pour estimer la valeur en croisement d'individus de race pure à partir d'une population de référence constituée d'animaux croisés. Les premiers modèles incluent uniquement des effets additifs (Ibanez-Escriche et al., 2009), pouvant être communs à l'ensemble des races impliquées dans le croisement, ou bien spécifiques à chacune des races (modèles additifs « moyen » et « racial », respectivement). Plus récemment, Zeng et al. (2013) ont proposé un modèle incluant des effets de dominance en plus des effets additifs (modèle à « dominance »). L'efficacité de ces modèles pour sélectionner des individus de race pure pour leur valeur en croisement a été comparée sur des données simulées. Il semble ressortir des résultats de Zeng et al. (2013) que le modèle générant la réponse à la sélection la plus importante dans la population croisée est le modèle à dominance. Selon ces auteurs, la supériorité de ce modèle s'expliquerait par son aptitude à fixer dans les populations parentales des allèles différents pour les QTL présentant un effet de dominance favorable, ce qui maximiserait l'hétérosis dans la population croisée. Leurs résultats suggèrent que l'avantage du modèle à dominance sur les deux modèles additifs augmente avec la part de la variance génétique expliquée par la dominance et l'hétérosis, mais tend à s'amenuiser lorsque le nombre de QTL impliqués dans le déterminisme du caractère sélectionné s'accroît. Contre toute attente, le modèle additif racial s'est révélé en général moins efficace que le modèle additif moyen, y compris dans le cas où les populations impliquées dans le croisement étaient génétiquement très distantes (Ibanez-Escriche et al., 2009 ; Zeng et al., 2013). Ce résultat s'explique probablement par le plus grand nombre de paramètres à estimer dans le modèle additif racial (nombre de marqueurs * nombre de races impliquées dans le croisement considéré) par rapport au modèle additif moyen.

Le principe de la sélection génomique d'individus de race pure pour leur valeur en croisement semble donc séduisant en théorie. Toutefois, l'efficacité de cette approche n'a, à notre connaissance, pas encore été évaluée sur des populations animales réelles. Un projet expérimental impliquant l'INRA et des entreprises de sélection porcines françaises est actuellement en cours (projet UTOPIGE, ANR_10_GENOM_BTV_015), dont l'un des objectifs sera justement de comparer l'efficacité d'évaluations génomiques s'appuyant sur une population de référence croisée ou sur une population de référence de race pure pour prédire la valeur en croisement de reproducteurs Piétrain. Les résultats de ce projet sont attendus pour 2014.

Plusieurs éléments suggèrent toutefois que la mise en œuvre d'évaluations génomiques d'individus de race pure pour leur valeur en croisement serait plus complexe que celle d'un

programme de sélection génomique strictement conduit en race pure tel qu'envisagé dans nos simulations :

- Les associations entre marqueurs et QTL peuvent varier entre les différentes races ou lignées constituant le croisement considéré, ces différences étant *a priori* d'autant plus fortes que les populations sont génétiquement distantes comme c'est le cas entre populations paternelles et populations maternelles par exemple. Dans ces conditions, un modèle génétique permettant d'estimer des effets de substitution propres à chaque race ou lignée sélectionnée semble s'imposer, qui nécessiterait de constituer et d'entretenir une population de référence croisée de plus grande taille, donc plus onéreuse, que pour une évaluation génomique en race pure.
- Si on considère un croisement trois voies entre 2 populations maternelles (A et B) et une population paternelle (C), l'apparentement entre les candidats et les individus croisés (AxB)xC de la population de référence serait nettement plus faible que dans le cas d'une évaluation en race pure, comme illustré sur la figure 3. On constate en effet que les candidats à la sélection des races pures A et B auraient au mieux un coefficient de parenté de $0,5^5$ (1 grand-parent en commun) avec les individus croisés (AB)C de la population de référence pour les caractères de production, et au mieux un coefficient de parenté de $0,5^4$ avec les femelles AB constituant la population de référence pour les caractères d'aptitude maternelle. Or, la précision des valeurs génomiques estimées dépend non seulement du DL entre marqueurs et QTL, mais également de la parenté entre les candidats évalués et la population de référence. Ce faible apparentement aurait donc probablement pour conséquence de faire chuter fortement la précision des valeurs génomiques estimées par rapport à une évaluation conduite en race pure, même en utilisant une méthode bayésienne, tirant plus profit du DL que de la parenté. La situation serait toutefois beaucoup plus favorable pour les candidats à la sélection de la race pure C, qui pourraient disposer de demi-germains dans la population de référence croisée pour les caractères de production.
- Une sélection génomique pour la valeur en croisement nécessiterait *a priori* de génotyper non seulement les candidats à la sélection afin de pouvoir les évaluer (comme pour une sélection en race pure), mais également les individus croisés constituant la population de référence. A ce surcoût viendrait s'ajouter celui de la mise en place d'un phénotypage en élevage de production. On pourrait toutefois envisager de constituer des populations de référence composées de reproducteurs des noyaux de sélection génotypés, et dont le phénotype serait la performance moyenne de leurs descendants croisés. On pourrait ainsi s'abstenir de génotyper les porcs charcutiers et les truies parentales et réduire

considérablement le surcoût du programme de sélection. Cette solution imposerait toutefois de connaître le pedigree des animaux croisés, et permettrait *a priori* une moins bonne valorisation de l'hétérosis dans les populations commerciales.

- La constitution de populations de référence croisées nécessiterait de modifier lourdement l'organisation actuelle des programmes d'amélioration porcins, puisqu'elle requerrait de mettre en place un contrôle de performances et un circuit de remontée de l'information zootechnique individuelle dans les élevages de production.

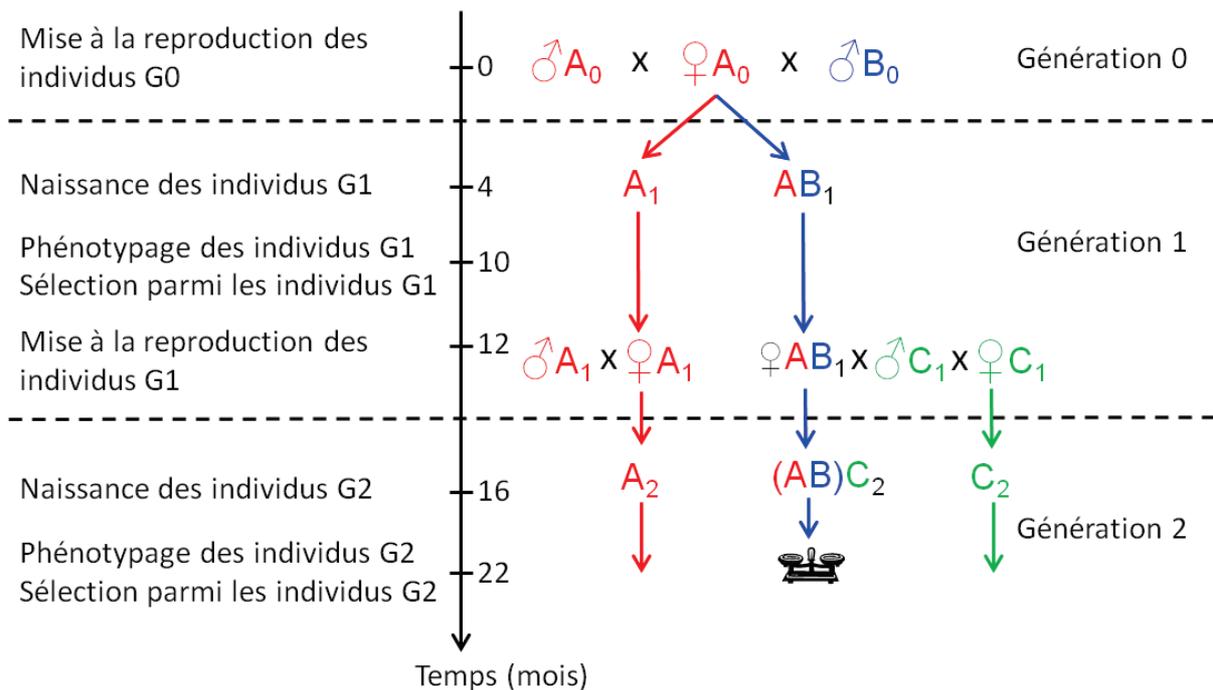


Figure 3 : Délai d'obtention des phénotypes de production de porcs charcutiers et d'aptitudes maternelles de truies parentales dans un système de production porcine fondé sur un croisement trois voies (AB)C : A t=0, accouplement des reproducteurs de races pures de la génération 0. Les performances d'aptitudes maternelles des premières femelles parentales AB et de production des premiers porcs charcutiers (AB)C issus des reproducteurs des races A et B accouplés à t=0 sont disponibles, respectivement, à t1=16 mois et à t2=22 mois, au moment de la sélection des candidats des races A et B de la génération 2.

Compte tenu de ces difficultés organisationnelles et surtout méthodologiques, on peut donc penser que la mise en place de programmes performants de sélection génomique pour la valeur en croisement nécessiterait une amélioration préalable de l'efficacité des méthodes d'évaluations génomiques disponibles actuellement.

On pourrait donc imaginer mettre en place dans un premier temps des évaluations génomiques en race pure dans les noyaux de sélection, ce qui ne nécessiterait pas de modifier l'organisation actuelle des programmes de sélection. Les dizaines de milliers d'individus phénotypés et génotypés pour constituer les populations de référence pourraient alors être mis à profit pour constituer de grands dispositifs de recherche, génotypés et imputés jusqu'à la séquence, destinés à cartographier très finement les mutations causales sous-jacentes aux QTL en ségrégation dans l'ensemble des populations sélectionnées. La connaissance de ces mutations causales permettrait alors d'améliorer l'efficacité des modèles d'évaluation génomique, ce qui ouvrirait la voie à la mise en place de programmes performants de sélection pour la valeur en croisement.

ANNEXES

ANNEXE 1

Estimation de l'augmentation annuelle de marge par porc charcutier issu d'un verrat terminal Piétrain collectif français (PCF).

Nous avons considéré que la marge par porc engraisé est égale à la différence entre la valeur commerciale de la carcasse et le coût de production de la carcasse.

La valeur commerciale de la carcasse dépend du taux de muscle des pièces (TMP) et du poids de la carcasse. Les coûts de production dépendent des performances de vitesse de croissance et d'efficacité alimentaire de l'animal en période d'engraissement, ainsi que d'autres coûts considérés ici comme non impactés par le travail de sélection pour simplifier (coûts alimentaires en période de post-sevrage, prix de revient du porcelet, frais vétérinaires, ...).

Les estimations de progrès génétique réalisé dans la population PCF entre 2008 et 2012 sont présentées sur la figure 1.

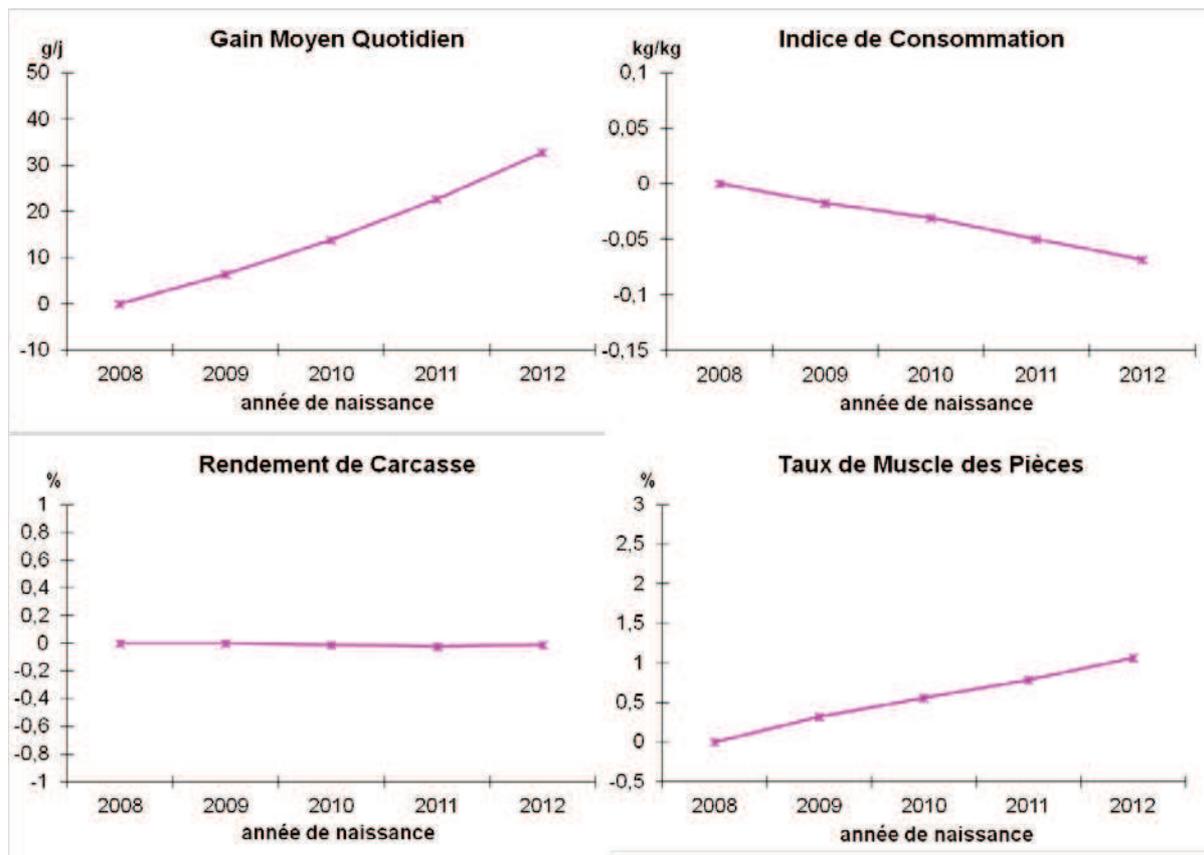


Figure 1 : Evolutions génétiques estimées entre 2008 et 2012 dans la population Piétrain LGPC pour le gain moyen quotidien et l'indice de consommation en période

d'engraissement, le rendement de carcasse et le taux de muscle des pièces (IFIP, évaluation génétique des populations collectives françaises).

Les verrats PCF étant utilisés majoritairement comme verrats terminaux, la moitié du progrès génétique créé dans le noyau de sélection des lignées mâles est théoriquement transmise à l'étage de production. En l'absence de progrès génétique sur le rendement de la carcasse, l'augmentation annuelle de marge ΔM_{a_prod} générée en production par l'amélioration génétique dans la population PCF est donc égale à :

$$\Delta M_{a_prod} = \left(\frac{\Delta G_{a_GMQ}}{2} * ValEco_{GMQ} \right) + \left(\frac{\Delta G_{a_IC}}{2} * ValEco_{IC} \right) + \left(\frac{\Delta G_{a_TMP}}{2} * ValEco_{TMP} \right)$$

où

- ΔG_{a_GMQ} , ΔG_{a_IC} et ΔG_{a_TMP} sont, respectivement, le progrès génétique réalisé annuellement dans le noyau de sélection PCF pour le gain moyen quotidien en période d'engraissement (**GMQ**), l'indice de consommation en période d'engraissement (**IC**) et le TMP de la carcasse, estimés à +8,20 g/j /an pour le GMQ, -0,0171 point/an pour l'IC et +0,267 point/an de TMP en moyenne sur la période 2008-2012;
- $ValEco_{GMQ}$, $ValEco_{IC}$ et $ValEco_{TMP}$ sont, respectivement, la valeur économique par unité de caractère du GMQ, de l'IC et du TMP de la carcasse.

La valeur économique du GMQ a été estimée à 0,02128 €/porc/g en 2012 (IFIP, communication personnelle).

La valeur économique de l'IC en période d'engraissement a été estimée à -23,21 €/porc/point d'IC, en considérant un gain de poids moyen par porc de 84,4 kg sur cette période (référence naisseurs-engraisseurs GTE 2010) et un prix moyen de l'aliment de 275 €/t (prix moyen de l'aliment engraissement en 2012, IFIP).

L'impact d'une augmentation de TMP sur la valeur de la carcasse n'est pas linéaire, comme indiqué sur la figure 2. Le prix du kg de carcasse augmente avec la quantité de muscle de la carcasse jusqu'à un TMP de 61%, mais n'augmente plus au-delà. Nous avons considéré que les carcasses des 20 millions de porcs charcutiers issus d'un verrot terminal Piétrain (PCF et autres variétés de Piétrain) représentaient les 80% de carcasses ayant les TMP les plus élevés dans la distribution présentée sur la figure 2 (TMP compris entre 58% et 68%).

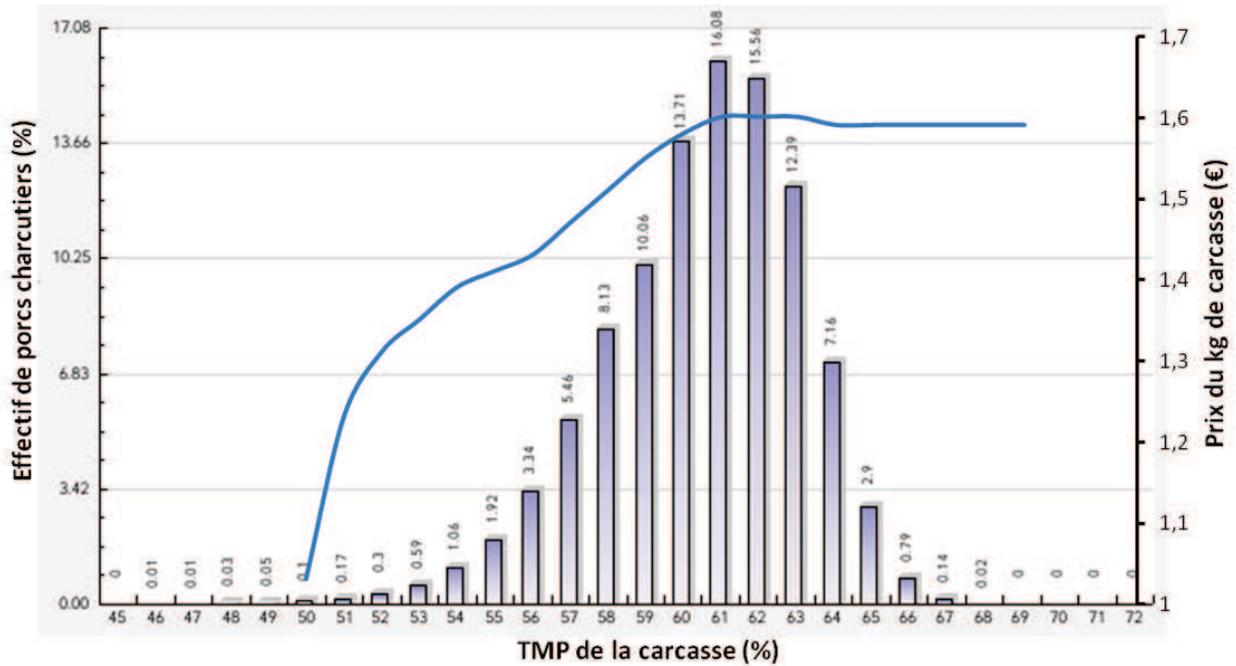


Figure 2 : Distribution du TMP des carcasses des porcs charcutiers (tous types génétiques confondus) et prix moyen du kg de carcasse en fonction du TMP en 2012 sur la zone Uniporc-Ouest (source : MPB Uniporc-Ouest).

Par conséquent, la valeur économique d'un point de TMP pour les porcs charcutiers issus d'un verret terminal PCF a été estimée en calculant la moyenne des augmentations de valeur commerciale d'une carcasse de 92 kg (poids de carcasse moyen en 2012, MPB Uniporc-Ouest) ΔVC_{TMPi} sur chaque catégorie de TMP entre 58% et 68%, en pondérant chaque augmentation de valeur par la proportion p_{TMPi} de porcs dans la catégorie considérée :

$$ValEco_{TMP} = \sum_{i=58}^{68} p_{TMPi} * \Delta VC_{TMPi} = 0,61 \text{ €/porc charcutier.}$$

L'augmentation annuelle de marge générée en production par l'amélioration génétique dans la population PCF ΔM_{a_prod} a donc été évaluée à :

$$\begin{aligned} \Delta M_{a_prod} &= \left(\frac{\Delta G_{a_GMQ}}{2} * ValEco_{GMQ} \right) + \left(\frac{\Delta G_{a_IC}}{2} * ValEco_{IC} \right) + \left(\frac{\Delta G_{a_TMP}}{2} * ValEco_{TMP} \right) \\ &= \left(\frac{8,20}{2} * 0,02128 \right) - \left(\frac{0,0171}{2} * (-23,21) \right) + \left(\frac{0,267}{2} * 0,61 \right) \\ &= \mathbf{0,367 \text{ €/porc charcutier/an.}} \end{aligned}$$

REFERENCES BIBLIOGRAPHIQUES

- Aguilar, I., I. Misztal, D. Johnson, A. Legarra, S. Tsuruta, and T. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93(2):743-752.
- Archibald, A. L., C. S. Haley, J. F. Brown, S. Couperwhite, H. A. McQueen, D. Nicholson, W. Coppeters, A. Van de Weghe, A. Stratil, A. K. Winterø, M. Fredholm, N. J. Larsen, V. H. Nielsen, D. Milan, N. Woloszyn, A. Robic, M. Dalens, J. Riquet, J. Gellin, J. C. Caritez, G. Burgaud, L. Ollivier, J. P. Bidanel, M. Vaiman, C. Renard, H. Geldermann, R. Davoli, D. Ruyter, E. J. M. Verstege, M. A. M. Groenen, W. Davies, B. Høyheim, A. Keiserud, L. Andersson, H. Ellegren, M. Johansson, L. Marklund, J. R. Miller, D. V. Anderson Dear, E. Signer, A. J. Jeffreys, C. Moran, P. Le Tissier, Muladno, M. F. Rothschild, C. K. Tuggle, D. Vaske, J. Helm, H. C. Liu, A. Rahman, T. P. Yu, R. G. Larson, and C. B. Schmitz. 1995. The PIGMaP consortium linkage map of the pig (*Sus scrofa*). *Mammalian Genome* 6:157-175.
- Bastiaansen, J., A. Coster, M. Calus, J. van Arendonk, and H. Bovenhuis. 2012. Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures. *Genetics Selection Evolution* 44(3).
- Boichard, D. 2002. PEDIG: a FORTRAN package for pedigree analysis suited for large populations. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production: 19-23 August 2002; Montpellier. 2002: 28-13.*
- Brito, F. V., J. Braccini Neto, M. Sargolzaei, J. A. Cobuci, and F. S. Schenkel. 2011. Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle. *BMC Genetics* 12(80):(20 September 2011).
- Buch, L. H., M. K. Sørensen, P. Berg, L. D. Pedersen, and A. C. Sørensen. 2011. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. *Journal of Animal Breeding and Genetics* 128(2):138-151.
- Caballero, A. and P. D. Keightley. 1994. A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* 138:883-900.
- Calus, M. P. L., Y. de Haas, M. Pszczola, and R. F. Veerkamp. 2013. Predicted accuracy of and response to genomic selection for new traits in dairy cattle. *Animal* 7(2):183-191.
- Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamp. 2008. Accuracy of Genomic Selection Using Different Methods to Define Haplotypes. *Genetics* 178:553-561.
- Calus, M. P. L. and R. F. Veerkamp. 2011. Accuracy of multi-trait genomic selection using different methods. *Genetics Selection Evolution* 43(26).
- Christensen, O. and M. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42(2).
- Christensen, O. F., P. Madsen, B. Nielsen, T. Ostensen, and G. Su. 2012. Single-step methods for genomic evaluation in pigs. *Animal* 6(10):1565-1571.
- Clark, S., J. Hickey, H. Daetwyler, and J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genetics Selection Evolution* 44(4).
- Cleveland, M. A., J. M. Hickey, and S. Forni. 2012. A Common Dataset for Genomic Analysis of Livestock Populations. *G3: Genes|Genomes|Genetics* 2(4):429-435.

- Coster, A., J. W. M. Bastiaansen, M. P. L. Calus, J. A. M. van Arendonk, and H. Bovenhuis. 2010. Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance. *Genetics Selection Evolution* 42(9).
- Daetwyler, H. D., M. P. L. Calus, R. Pong-Wong, G. de los Campos, and J. M. Hickey. 2013. Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics* 193:347-365.
- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185:1021-1031.
- Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics* 124(6):369-376.
- Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of Prediction of the Genetic Risk of Disease Using a Genome-Wide Approach. *Plos One* 3(10):1-8.
- Dassonneville, R., R. Brondum, T. Druet, S. Fritz, F. Guillaume, B. Guldbbrandtsen, M. Lund, V. Ducrocq, and G. Su. 2011. Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *Journal of Dairy Science* 94(7):3679-3686.
- de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193.
- de Roos, A. P. W., B. J. Hayes, R. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. *Genetics*, 179, 1503-1512.
- de Roos, A. P. W., B. J. Hayes, and M. E. Goddard. 2009. Reliability of genomic predictions across multiple populations. *Genetics* 183:1545-1553.
- de Roos, A. P. W., C. Schrooten, R. F. Veerkamp, and J. A. M. van Arendonk. 2010. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. *Journal of Dairy Science* 94(3):1559-1571.
- Dekkers, J. C. M. 1992. Asymptotic response to selection on best linear unbiased predictors of breeding values. *Animal Production* 54(3):351-360.
- Dekkers, J. C. M. 2007a. Marker assisted selection for commercial crossbred performance. *Journal of Animal Science* 85(9):2104-2114.
- Dekkers, J. C. M. 2007b. Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* 124:331-341.
- Druet, T. and M. Georges. 2010. A hidden Markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* 184(3):789-798.
- Du, F. X., A. C. Clutter, and M. M. Lohuis. 2007. Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences* 3(3):166-178.
- Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science* 95(7):4114-4129.

- Falconer, D. S. 1981. Introduction to quantitative genetics. 2nd Ed. Longmans Green, London/New York.
- Fritz, S., J. J. Colleau, T. Druet, M. Y. Boscher, M. N. Rossignol, A. Malafosse, and D. Boichard. 2003. Mise en place d'une sélection assistée par marqueurs dans les trios principales races bovines laitières françaises. *Rencontres Recherches Ruminants* 10:53-56.
- Gallais, A. 1990. Théorie de la sélection en amélioration des plantes. Ed. Masson, Paris.
- Gianola, D. 2013. Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. *Genetics* 194(3):573-596.
- Goddard, M. E. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136(2):245-257.
- Guillaume, F., D. Boichard, V. Ducrocq, and S. Fritz. 2011. Utilisation de la sélection génomique chez les bovins laitiers. *INRA Productions Animales* 24(4):363-368.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007a. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389-2397.
- Habier, D., R. L. Fernando, and D. J. Garrick. 2013. Genomic-BLUP decoded: a look into the black box of genomic prediction. *Genetics* 194:597-607.
- Habier, D., R. L. Fernando, K. Kizilkay, and D. J. Garrick. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186.
- Habier, D., K. U. Gotz, and L. Dempfle. 2007b. Estimation of genetic parameters on test stations using purebred and crossbred progeny of sires of the Bavarian Pietrain. *Livestock Science* 107(2/3):142-151.
- Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution* 42(1):5.
- Haley, C. S. and P. M. Visscher. 1998. Strategies to utilize marker-quantitative trait loci associations. *Journal of Dairy Science* 81:85-97.
- Harris, B. L., D. L. Johnson, and R. J. Spelman. 2008. Genomic selection in New Zealand and the implications for national genetic evaluations. in *Proceedings of the Interbull meeting, Niagara Falls, Canada*.
- Hayes, B. and M. E. Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution* 33(3):209-229.
- Hayes, B. J. 2011. QTL Mapping, MAS and Genomic Selection, Course Notes. 12-16 Septembre 2011, Toulouse, France.
- Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009a. Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92(2):433-443.
- Hayes, B. J., H. D. Daetwyler, P. Bowman, G. Moser, B. Tier, R. Crump, M. Khatkar, H. W. Raadsma, and M. E. Goddard. 2009b. Accuracy of genomic selection: comparing theory and results. Pages 352-355 in *Proceedings of the Association for the Advancement of Animal Breeding and Genetics, Barossa Valley, Australia*.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009c. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetical Research* 91:47-60.

- Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* 13:635-643.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447.
- Henryon, M., P. Berg, T. Ostersen, B. Nielsen, and A. C. Sørensen. 2012. Most of the benefits from genomic selection can be realized by genotyping a small proportion of available selection candidates. *Journal of Animal Science* 90:4681-4689.
- Hickey, J. M., B. P. Kinghorn, B. Tier, J. H. van der Werf, and M. A. Cleveland. 2012. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genet Sel Evol* 44:9.
- Hill, W. G. and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38:226-231.
- Hill, W.G. 1985. Effects of population size on response to short and long term selection. *J. Anim. Br. Genet.* 102:161-173.
- Hoerl, A. E. and R. W. Kennard. 1970. Ridge regression: biased estimation for non orthogonal problems. *Technometrics* 12:55-67.
- Hu, Z. L., C. A. Park, X. L. Wu, and J. M. Reecy. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research* 41(D1):D871-D879.
- Hu, Z. L. and J. M. Reecy. 2007. Animal QTLdb: beyond a repository: a public platform for QTL comparisons and integration with diverse types of structural genomic information. *Mammalian Genome* 18(1):1-4.
- Huang, Y., J. M. Hickey, M. A. Cleveland, and C. Maltecca. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. *Genet Sel Evol* 44:25.
- Ibanez-Escriche, N., R. L. Fernando, A. Toosi, and J. C. M. Dekkers. 2009. Genomic selection of purebreds for crossbred performance. *Genetic Selection Evolution* 41:12-21.
- IFIP. 2009. *Le porc par les chiffres*. Ed. IFIP.
- Jannink, J. L. 2010. Dynamics of long-term genomic selection. *Genetics Selection Evolution* 42(35).
- Jia, Y. and J. L. Jannink. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192:1513-1522.
- Kinghorn, B. P., J. M. Hickey, and J. H. J. van der Werf. 2010. Reciprocal recurrent genomic selection for total genetic merit in crossbred individuals. In *Proceedings of the 9th WCGALP, Leipzig, Germany, August 1-6, 2010*. Comm 36.
- Kinsman, A., M. Sargolzaei, F. Schenkel, G. Vender Voort, M. Jafarikia, and A. Robinson. 2012. Accuracy of genotype imputation in Canadian Yorkshire pigs using FIMPUTE software. In *Proceedings of the 4th International Conference of Quantitative Genetics: 17-22 June 2012; Edinburgh. 2012: P-192*. [<http://cdn.fl1000.com/posters/docs/251407277>]
- König, S., H. Simianer, and A. Willam. 2009. Economic evaluation of genomic breeding programs. *Journal of Dairy Science* 92(1):382-391.

- Lande, R. and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124(3):743-756.
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92(9):4656-4663.
- Legarra, A. and V. Ducrocq. 2012. Computational strategies for national integration of phenotypic, genomic and pedigree data in a single-step best linear unbiased prediction. *Journal of Dairy Science* 95(8):4629-4645.
- Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011a. A comparison of dairy cattle breeding designs that use genomic selection. *Journal of Dairy Science* 94(1):493-500.
- Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011b. Genomic selection for maternal traits in pigs. *Journal of Animal Science* 89:3908-3916.
- Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbadtsten, Z. Liu, R. Reents, C. Schrooten, M. Seefried, and G. Su. 2010. Improving genomic prediction by EuroGenomics collaboration. Paper 880 in Proceedings of the 9th WCGALP, 1-6 August 2010, Leipzig, Germany.
- Lutaaya, E., I. Misztal, J. W. Mabry, T. Short, H. H. Timm, and R. Holzbauer. 2001. Genetic parameter estimates from joint evaluation of purebreds and crossbreds in swine using the crossbred model. *Journal of Animal Science* 79:3002-3007.
- Luther, H., P. Vögeli, and A. Hofer. 2009. Increasing genetic E. coli F18 resistance in Swiss pigs. Pages Comm. 18-03. in Proceedings of the 60th EAAP, 24-27 August 2009, Barcelona, Spain.
- Maignel, L., T. Tribout, D. Boichard, J. P. Bidanel, and R. Guéblez. 1998. Analyse de la variabilité génétique des races porcines Large White, Landrace Français et Piétrain, sur la base de l'information généalogique. *Journées de la Recherche Porcine* 30:109-116.
- Merks, J. M. W. 2001. Genetic improvement at the commercial level compared at the genetic progress at the nucleus level. in Proceedings of the National Swine Improvement Federation, 6-7 Decembre 2001, St Louis, Missouri, USA.
- Merks, J. W. M. 1989. Genotype x environment interactions in pig breeding programmes. VI. Genetic relations between performances in central test, on-farm test and commercial fattening. *Livestock Production Science* 22:325-339.
- Meuwissen, T. H. E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* 75:934-940.
- Meuwissen, T. H. E. 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution* 41(35).
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4):1819-1829.
- Meuwissen, T. H. E. and M. E. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 185:623-631.
- Meyer, K., B. Tier, and H. U. Graser. 2013. Technical note: Updating the inverse of the genomic relationship matrix. *Journal of Animal Science* 91:2583-2586.
- Minvielle, F. 1990. Principes d'amélioration génétique des animaux domestiques. Les Presses de l'Université Laval, Québec, Canada.

- Moser, G., B. Tier, R. E. Crump, M. S. Khatkar, and H. W. Raadsma. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genet Sel Evol* 41(56).
- Muir, W. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics* 124(6):342-355.
- Nielsen, H., A. Sonesson, and T. Meuwissen. 2011. Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *Journal of Animal Science* 89(3):630-638.
- Noguera, J. L., M. C. Rodríguez, L. Varona, A. Tomas, G. Muñoz, O. Ramirez, C. Barragan, M. Arqué, J. P. Bidanel, M. Amills, C. Ovilo, and A. Sánchez. 2006. Epistasis is a fundamental component of the genetic architecture of prolificacy in pigs. in *Proceedings of the 8th WCGALP, 13-18 August 2006, Belo Horizonte, MH, Brasil.*
- Ostensen, T., O. F. Christensen, M. Henryon, B. Nielsen, G. Su, and P. Madsen. 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution* 43(38).
- Pimentel, E. C. G. and S. König. 2012. Genomic selection for the improvement of meat quality in beef. *Journal of Animal Science* 90:3418-3426.
- Pryce, J. E., M. E. Goddard, H. W. Raadsma, and B. J. Hayes. 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. *Journal of Dairy Science* 93(11):5455-5466.
- Pszczola, M., T. Strabel, H. Mulder, and M. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *Journal of Dairy Science* 95(1):389-400.
- Quaas, R. L. and E. J. Pollak. 1980. Mixed Model Methodology for Farm and Ranch Beef Cattle Testing Programs. *Journal of Animal Science* 51:1277-1287.
- Rincint, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V. M. Rodríguez, J. Moreno-Gonzalez, A. Melchinger, E. Bauer, C. C. Schoen, N. Meyer, C. Giauffret, C. Bauland, P. Jamin, J. Laborde, H. Monod, P. Flament, A. Charcosset, and L. Moreau. 2012. Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (*Zea mays* L.). *Genetics* 192(2):715-728.
- Robert-Granié, C., A. Legarra, and V. Ducrocq. 2011. Principes de base de la sélection génomique. *INRA Productions Animales* 24(4):331-340.
- Roos, A. P. W. d., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics* 179(3):1503-1512.
- Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123:218-223.
- Schwob, S., J. Riquet, T. Bellec, L. Kernaleguen, T. Tribout, and J. P. Bidanel. 2009. Mise en place d'un programme de sélection assistée par marqueurs dans la population sino-européenne Duochan. *Journées de la Recherche Porcine* 41:29-30.
- Sidani, C., J. M. Astruc, M. Baelden, F. Barillet, B. Bibé, A. Bonnot, M. Y. Boscher, D. Bouchel, B. Bouffartigue, J. Bouix, M. Brochard, F. Dion, D. Francois, E. Jouhet, E.

- Jullien, C. R. Leymarie, C. Moreno, M. Orlianges, I. Palhière, G. Perret, J. Raoul, A. Raynal, L. Tiphine, and P. Tribon. 2010. The french ovine scrapie plan: results and prospects. in Proceedings of the 9th WCGALP, 1-6 August 2010, Leipzig, Germany.
- Solberg, T. R., A. K. Sonesson, J. A. Woolliams, J. Odegard, and T. H. E. Meuwissen. 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genetics Selection Evolution* 41(53).
- Su, G., R. F. Brøndum, P. Ma, B. Guldbbrandtsen, G. P. Aamand, and M. S. Lund. 2013. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *Journal of Dairy Science* 95(8):4657-4665.
- Sved, J. A. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite population. *Theoretical Population Biology* 2:125-141.
- Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 58:267-288.
- Tribout, T. 2011. Perspectives d'application de la sélection génomique dans les schémas d'amélioration génétique porcins. *INRA Productions Animales* 24(4):369-376.
- Tribout, T., J. C. Caritez, J. Gogué, J. Gruand, Y. Billon, M. Bouffaud, H. Lagant, J. Le Dividich, F. Thomas, H. Quesnel, R. Guéblez, and J. P. Bidanel. 2003. Estimation, par utilisation de semence congelée, du progrès génétique réalisé en France entre 1977 et 1998 dans la race porcine Large White : résultats pour quelques caractères de reproduction femelle. *Journées de la Recherche Porcine* 35:285-292.
- Tribout, T., J. C. Caritez, J. Gogué, J. Gruand, M. Bouffaud, Y. Billon, C. Péry, H. Griffon, S. Brenot, M. H. Le Tiran, F. Bussièrès, P. Le Roy, and J. P. Bidanel. 2004. Estimation, par utilisation de semence congelée, du progrès génétique réalisé en France entre 1977 et 1998 dans la race porcine Large White : résultats pour quelques caractères de production et de qualité des tissus gras et maigres. *Journées de la Recherche Porcine* 36:275-282.
- Tribout, T., C. Larzul, and F. Phocas. 2012. Efficiency of genomic selection in a purebred pig male line. *Journal of Animal Science* 90(12):4164-4176.
- Tribout, T., C. Larzul, and F. Phocas. 2013. Economic interest of implementing genomic evaluations in a pig male line breeding scheme. *Genetics Selection Evolution* (en cours de re-revision).
- Tsuruta, S., I. Misztal, I. Aguilar, and T. Lawlor. 2011. Multiple-trait genomic evaluations of linear type traits using genomic and phenotypic data in US Holsteins. *Journal of Dairy Science* 94:4198-4204.
- Uimari, P. and M. Tapio. 2011. Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. *Journal Of Animal Science* 89(3):609-614.
- Van der Steen, H. A. M., G. F. W. Prall, and G. S. Plastow. 2005. Application of genomics to the pork industry. *Journal of Animal Science* 83(E. Suppl.): E1-E8.
- Van Grevenhof, E. M., J. A. Van Arendonk, and P. Bijma. 2012. Response to genomic selection: the Bulmer effect and the potential of genomic selection when the number of phenotypic records is limiting. *Genet Sel Evol* 44:26.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91(11):4414-4423.

- VanRaden, P. M., D. J. Null, M. Sargolzaei, G. R. Wiggans, M. E. Tooker, J. B. Cole, T. S. Sonstegard, E. E. Connor, M. Winters, J. B. C. H. M. van Kaam, A. Valentini, B. J. Van Doormaal, M. A. Faust, and G. A. Doak. 2013. Genomic imputation and evaluation using high-density Holstein genotypes. *Journal of Dairy Science* 96(1):668-678.
- VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92:16-24.
- Verbyla, K., P. Bowman, B. Hayes, H. Raadsma, M. Khatkar, and M. E. Goddard. 2009. Comparison of Bayesian methods for genomic selection using real dairy data. In *Proceedings of the 60th EAAP, 24-27 August 2009, Barcelona, Spain*.
- Wei, M. and H. A. M. Van der Steen. 1991. Comparison of reciprocal recurrent selection with pure-line selection systems in animal breeding (a review). *Animal Breeding Abstract* 59:281-298.
- Wensch-Dorendorf, M., T. Yin, H. H. Swalve, and S. König. 2011. Optimal strategies for the use of genomic selection in dairy cattle breeding programs. *Journal of Dairy Science* 94(8):4140-4151.
- Wiggans, G., T. Cooper, P. VanRaden, K. Olson, and M. Tooker. 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. *Journal of Dairy Science* 95(3):1552-1558.
- Wolc, A., J. Arango, P. Settar, J. Fulton, N. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D. Garrick, and J. Dekkers. 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. *Genetics Selection Evolution* 43(1):23.
- Wright, S. 1931. Evolution in mendelian populations. *Genetics* 16(2):97-159.
- Xu, S. 2003. Theoretical Basis of the Beavis Effect. *Genetics* 165(4):2259-2268.
- Yu, X., J. A. Woolliams, and T. H. E. Meuwissen. 2012. Which animals are to be densely genotyped in order to impute the missing genotypes of sparsely genotyped animals. In *Proceedings of the 4th International Conference of Quantitative Genetics: 17-22 June 2012; Edinburgh. 2012: P-165*.
- Zeng, J., A. Toosi, R. L. Fernando, J. C. M. Dekkers, and D. J. Garrick. 2013. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genetics Selection Evolution* 45(11).
- Zhou, H. and T. Hastie. 2005. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society, Serie B*, 67:301-320.