Utilisations d'un Panel SNPs très basse densité dans les populations en sélection de petits ruminants

Jérôme Raoul

Université Fédérale

Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Présentée et soutenue par :
Jérôme Raoul

le mardi 28 novembre 2017

Titre :

Utilisations d'un panel SNPs très basse densité dans les populations en sélection de petits ruminants

---

École doctorale et discipline ou spécialité :
ED SEVAB : Pathologie, Toxicologie, Génétique et Nutrition

Unité de recherche :
UMR1388 - Génétique, Physiologie et Systèmes d'Elevage (GenPhySE) - INRA

Directeur/trice(s) de Thèse :

Jean-Michel Elsen

Jury :
Jean-Michel Elsen, directeur de Thèse
Frédéric Farnir, rapporteur
Rodolpho Juan Carlos Cantet, rapporteur
Catherine Larzul, examinatrice
Florence Phocas, examinatrice
Mickaël Brochard, invité

Avant-Propos

Ces travaux de thèse ont nécessité un certain investissement individuel (paraît-il, c'est normal) mais ils n'ont pas été réalisés seul. Aussi je tiens à remercier l'ensemble des personnes et institutions qui ont contribué à leur réalisation.

••••••••••••••••••••••••••••••

Merci à l'institut de l'élevage, l'INRA et Apis-gène pour leur contribution financière.

Merci Jean-Michel pour ta bienveillance à mon égard.

Merci aux membres du jury, Frédéric Farnir, Rodolpho J. C. Cantet, Catherine Larzul, Florence Phocas et Mickaël Brochard pour avoir accepté d'évaluer ces travaux.

Merci aux membres de mon comité de thèse, Isabelle Palhière, Flavie Tortereau, Jean-Michel Astruc, Loys Bodin, Albin Bouquet et Andrew Swan pour le suivi et la contribution à ces travaux.

Merci à l'ensemble des collègues INRA et idele et en particulier les assistantes et les équipes GESPER et MG² pour leur soutien, les échanges, les bons moments partagés.

Thanks to Andrew and Julius. I really appreciated to discuss my results with you.

Merci à mes co-auteurs. Me lire n'est pas toujours un plaisir ! Pour preuve, je vous ressers un des commentaires (le plus beau forcément) de maître Elsen sur la première version du second article : « *Franchement, n'importe quel être normal, arrivé ici, soit se suicide, soit va boire une bière, soit s'endort* ».

Je tiens à remercier également l'ensemble des personnes qui m'ont accompagné depuis mon arrivée à l'institut de l'élevage en 2007 et notamment Gilles Perret, Jacques Bouix, Eric Jullien, Aline Bonnot et Loys Bodin.

••••••••••••••••••••••••••••••

La réalisation de cette thèse a été l'occasion de partir un an en Australie, une aventure dans l'aventure.

Merci au métaprogramme SELGEN de l'INRA, au département de génétique de l'INRA, à l'INPT, à GenPhySe et à l'AGBU pour leur participation financière à cette mission.

All my thanks to AGBU staff, especially Kathy, Susanne and Rob, to make our trip fantastic! A great thanks to Kim and Andrew for all things you did for us. Life is so easy when you take care of us!

Thanks Sarrita to explain me the Lazy Ossie Speaking Trick (LOST)… I was often lost! I can't mention everybody but a special thought to the AGBU running team (headed by Aaron and Sam), the running mates group, the soccer group, my second roommate, Tukky, and Naomie, the official supplier of Dutch cheese (at least the flag!!)

••••••••••••••••••••••••••••••

Merci à ma famille et amis de m'apprécier pour ce que je suis. Dites-vous que je pourrais être pire !

Merci à toi Aurélie de partager mes moments de bonheur mais aussi pour ton soutien face à mes moments d'insatisfaction (permanents), de doutes (récurrents) et de tristesse (beaucoup plus rares mais ça arrive).

Merci à mes parents pour l'amour qu'ils ont su m'apporter, à mes sœurs pour leur fraternité.

Merci à mes enfants, Prosper, Auguste et Fantine pour leur habilité à me faire déconnecter.

Charles, j'ai une pensée émue aux moments passés ensemble. Merci de m'avoir fait découvrir beaucoup de choses et en particulier le métier d'éleveur dont tu parlais avec passion et fierté. J'associe bien sûr à ces remerciements Badette, Denis et l'ensemble des Pitel pour leur amitié et leur accueil toujours chaleureux (et le calva, il faut garder les pieds sur terre !).

••••••••••••••••••••••••••••••

Je me suis souvent dit qu'*on ne devrait jamais quitter Montaudran** ! J'ai accepté ce nouveau départ malgré sa part de risque. J'en suis heureux.

*l'auteur a habilement modifié le nom de la ville par souci d'échapper aux droits d'auteur (principe d'optimisation fiscale).

Résumé

Les programmes de sélection visent à produire des reproducteurs de bonnes valeurs génétiques pour la filière. La connaissance de marqueurs moléculaires du génome des individus et de mutations d'intérêt ouvrent des perspectives en termes d'organisation de la sélection. A l'aide de simulations déterministes et stochastiques, l'intérêt technique et économique de l'utilisation d'un panel de marqueurs moléculaires très basse densité a été évalué dans les populations ovines et caprines en sélection et permis d'obtenir les résultats suivants : i) utiliser un tel panel pour accroître, quand elle est limitée, la quantité de filiations paternelles n'est pas toujours rentable, ii) la stratégie de gestion des gènes d'ovulation qui maximise la rentabilité économique du plan de sélection a été déterminée par optimisation et des stratégies simples à implémenter, qui donnent des rentabilités proches de la rentabilité maximale, ont été proposées, iii) un programme de sélection génomique basé sur un panel très basse densité, permet à coût constant une efficacité supérieure aux programmes basés actuellement sur le testage sur descendance des mâles.

Abstract

Breeding programs aim to transfer high genetic value breeding stock to the industry. The knowledge of molecular markers of individual's genome and causal mutations allow to conceive new breeding program designs. Based on deterministic and stochastic simulations, the technical and economic benefits of using a very low density molecular markers panel were assessed in sheep and goat populations. Following results were obtained: i) using such a panel to increase female paternal filiations in case of incomplete pedigree is not always profitable, ii) a method of optimization has been used to derive the maximal profits of managing ovulation genes, and practical management giving profits close to the maximal profits have been determined, iii) at similar cost, a genomic design based on a very low density panel is more efficient than the current design based on progeny testing.

Table des matières

Liste des abréviations

aAGG: gain génétique annuel asymptotique (asymptotic annual genetic gain)

AGG : gain génétique annuel (annual genetic gain)

AI_NoPT : insémination artificielle sans testage sur descendance (artificial insemination without progeny testing)

AI_PT : insémination artificielle avec testage sur descendance (artificial insemination with progeny testing)

BLUP : meilleur prédicteur linéaire non biaisé (best linear unbiaised prediction)

EBV : valeur génétique estimée (estimated breeding value)

GBLUP : meilleur prédicteur linéaire non biaisé génomique (genomic best linear unbiaised prediction)

GC : coût du génotypage (genotyping cost)

GE : effet du génotype (genotype effect)

GEBV : valeur génomique estimée (genomic estimated breeding value)

HD : haute densité (high density)

IA: insemination artificielle

Init.KPR : taux de paternités connues initial (initial known paternity rate)

KPR : taux de paternités connues (known paternity rate)

LD: basse densité (low density)

MD: densité moyenne (medium density)

M_ear : précision moyenne modélisée des valeurs génétiques estimées (model average accuracy of estimated breeding values)

MME : équations du modèle mixte (mixed model equations)

NM : monte naturelle (natural mating)

No_AI : aucune insémination artificielle (No artificial insémination)

OPT : design optimisé

QTL : locus de caractères quantitatifs (quantitative trait locus)

R_ear : précision moyenne réelle des valeurs génétiques estimées (real average accuracy of estimated breeding values)

SD : écart-type (standard deviation)

SGG : gain génétique supplémentaire (suplementary genetic gain)

SNP : polymorphisme d'une seule base nucléotidique (single nucleotide polylmorphism)

ssGBLUP : meilleur prédicteur linéaire non biaisé génomique en une étape (single step genomic best linear unbiaised prediction)

VLD : très basse densité (very low density)

Introduction générale

Les programmes d'amélioration génétiques des petits ruminants visent à garantir à leurs filières, notamment aux élevages de production, l'approvisionnement en reproducteurs de qualité. Basés sur une organisation collective regroupant les éleveurs sélectionneurs, leur encadrement administratif, technique et scientifique, ces programmes sont fondés sur trois processus interconnectés : i) le recueil de données en élevages ou sites dédiés (*e.g.* stations de contrôle des jeunes mâles, centre d'insémination) incluant les performances zootechniques et la généalogie, ii) l'évaluation du potentiel génétique des animaux et iii) la gestion des populations en sélection (*e.g.* choix du renouvellement mâle et femelle, organisation des accouplements). Une des fonctions essentielles des programmes de sélection est d'améliorer le potentiel génétique des reproducteurs pour un ou plusieurs caractères zootechniques tout en limitant la perte de diversité génétique de la population en sélection. Le gain génétique (*i.e.* variation du niveau génétique moyen des reproducteurs au cours de générations successives) dépend des moyens mis en œuvre (*e.g.* nombre de femelles et males dont on enregistre les performances) et des paramètres génétiques des caractères sélectionnés.

Les moyens mis en œuvre sont très variables au sein des programmes de sélection petits ruminants français. L'insémination artificielle (IA), qui induit en le facilitant le contrôle des paternités, est apparue comme un outil très utile à la planification du testage sur descendance, permettant à la fois de contrôler la taille des lots de testage et la connexion statistique des effets troupeaux par une répartition de ces lots entre élevages. Mais la mise en œuvre de l'insémination animale en petits ruminants souffre de contraintes techniques importantes et représente un coût qui limite son développement. Dans de nombreuses races,

une part importante voire prépondérante des femelles est accouplée par lutte naturelle, mode de reproduction pour lequel l'obtention des paternités est contraignante et constitue une activité coûteuse en temps et parfois difficile à réaliser dans certains systèmes d'élevage (pâturage, transhumance). En conséquence, les taux de filiations paternelles des femelles sont variables allant de situations où les généalogies sont quasi complètes (la plupart des races ovines laitières et bouchères) à très incomplètes (races rustiques de montagne) en passant par l'ensemble des situations intermédiaires (48% de chèvres inscrites au contrôle de performances officiel ont une filiation paternelle). Les évolutions récentes des productions ovines et caprines, avec un accroissement de la taille des élevages et la nécessité d'optimiser la main d'œuvre, renforce cette contrainte : l'obtention des filiations constitue pour de nombreuses races un frein important au maintien, à la création et au renouvellement des élevages de sélection. Au-delà des conséquences en termes d'attractivité du métier de sélectionneurs en petits ruminants, les filiations incomplètes induisent une perte d'efficacité du schéma via notamment une perte de précision au niveau de l'évaluation génétique.

L'ADN, molécule transmise à la descendance, est le principal support physique de l'hérédité, c'est-à-dire de la transmission de caractéristiques parentales à la descendance. Le balisage du génome par des marqueurs de plus en plus denses a permis de localiser des portions de ce génome qui ont un effet sur des caractères d'intérêt (QTL pour Quantitative Trait Locus lorsque le caractère est quantitatif), notamment d'anomalies transmissibles, de phénomènes de résistance à des maladies infectieuses ou de phénotypes extrêmes tels que l'hypertrophie musculaire ou les tailles de portées extrêmes. Ces détections de QTL constituent les premières

étapes dans la recherche d'une mutation causale. L'identification de QTL avec une précision croissante permet d'affiner la localisation de la mutation voire de l'identifier précisément.

Lorsque la mutation est identifiée, il est possible d'opérer une sélection directe des individus porteurs de la forme favorable en génotypant (*i.e.* lecture de l'ADN) les candidats à la sélection au niveau de la mutation. Sous l'impulsion des efforts de recherche déployés en génétique humaine pour le séquençage du génome humain dans les années 2000, les développements technologiques ont permis d'élaborer de nouvelles méthodes de séquençage du génome dites à haut débit et l'utilisation d'une nouvelle génération de marqueurs appelés SNP (Single Nucléotide Polymorphism). Bien que la très grande majorité de la séquence d'ADN (quelques milliards de base) soit identique d'un individu à l'autre, quelques différences, de l'ordre du pourcent, existent entre les individus. Certaines de ces différences sont dues au changement d'un seul nucléotide au sein de la séquence et peuvent être exploitées comme marqueur (SNP) du génome. Le développement spectaculaire des outils d'analyse du génome s'est étendu aux espèces ovines (2009) et caprines (2011) avec la mise à disposition de la communauté scientifique et de l'industrie des premières versions d'outils haut débit d'analyse de l'ADN permettant de génotyper simultanément de l'ordre de 50 000 SNPs par individu.

Les outils d'analyse du génome et l'accumulation des connaissances sur son fonctionnement remettent en question les modalités traditionnelles de sélection des animaux en ferme. Ces progrès se sont matérialisés dans les programmes de sélection des petits ruminants par la mise en place de différentes stratégies de prise en compte de l'information moléculaire, réduite à certaines mutations d'intérêt (gènes affectant la masse musculaire, la taille de portée et la résistance à la tremblante en ovin) ou étendue à l'ensemble du génome dans le cadre de la sélection génomique d'abord en race ovine Lacaune lait suivi par les races

caprines Saanen et Alpine et races ovines laitières des Pyrénées. Les stratégies d'application sont soumises à des contraintes techniques et économiques et varient avec la taille des unités de sélection, les parts relatives de l'insémination et de la monte naturelle et les objectifs de sélection. Les coûts actuels de génotypages, rapportés à la valeur moyenne des reproducteurs, constituent un frein majeur au développement généralisé de ces stratégies dans les autres populations ovines et caprines en sélection.

A l'aide d'un soutien financier des acteurs de la sélection, un panel de marqueurs SNP visant à contrôler les filiations et assigner les paternités a récemment été sélectionné en ovin et caprin. Ce panel inclut des marqueurs SNPs dédiés à l'assignation de parenté et au contrôle des filiations et des marqueurs SNPs correspondant aux mutations en ségrégation connues dans chaque espèce. Ainsi ce sont 1432 SNPs candidats qui ont été identifiés en ovin et 1179 en caprins. Au préalable, les contrôles de filiation étaient effectués sur la base de marqueurs ancienne génération de type microsatellites. L'évolution des technologies, *i.e.* le passage des microsatellites aux SNPs, permet une automatisation des génotypages et le génotypage simultané de marqueurs dédiés à plusieurs objets (assignation de parenté/contrôles de filiations et mutations). Le passage aux SNPs offre donc potentiellement une réduction de coût et ouvre de nouvelles perspectives de déploiement de l'assignation de paternité dans les élevages en sélection, d'autant plus que les marqueurs SNP dédiées sont choisis parmi ceux utilisés pour l'analyse de génome (panels de plus hautes densités).

Les bénéfices techniques et économiques attendus de l'utilisation du panel de SNPs varient selon les populations en raison de la diversité des programmes de sélection : moyens mis en œuvre, objectifs, mutations d'intérêt en ségrégation, etc. L'accroissement de la quantité de filiations enregistrées pour l'évaluation génétique est un enjeu prépondérant pour les

populations dont la connaissance de l'information généalogique est actuellement incomplète. La gestion des mutations d'intérêt, a fait l'objet de travaux scientifiques dans les années 90 avec la recherche d'une gestion optimale, notamment en matière de vitesse de fixation, compte-tenu des répercussions sur les gains à court et long termes. Cependant, la gestion de mutations intéressantes à l'état hétérozygote mais indésirables à l'état homozygote, correspondant au cas des mutations conférant une hyper-ovulation en ovin, a été peu étudiée et exclusivement pour des mutations localisées sur le chromosome sexuel X (McEwan, 1995 ; Amer *et al.*, 1998). Enfin, l'utilité des génotypages à très basse densité dans le cadre d'un programme de sélection génomique reste à évaluer pour les populations ovines d'effectifs limités et basées au moins partiellement sur l'utilisation de la monte naturelle. L'évaluation des bénéfices techniques et économiques de ces différents usages du panel est une demande forte des acteurs de la sélection ovine et caprine en France qui ont apporté leur soutien à ces travaux de thèse à travers Apis-Gène.

Compte tenu de ces enjeux techniques et économiques liés à l'utilisation d'un panel SNP très basse densité dans les populations ovines et caprines en sélection, les objectifs de mes travaux de thèse sont de i) mesurer le gain génétique et la rentabilité économique d'une meilleure connaissance de la filiation paternelle des femelles via l'assignation de parenté dans les programmes de petits ruminants, ii) déterminer les stratégies optimales d'accouplements pour gérer une mutation avantageuse à l'état hétérozygote en ovin, iii) mesurer le gain génétique et l'intérêt économique de l'utilisation d'un panel de SNPs très basse densité pour la sélection génomique des ovins.

Le chapitre 1 donne une brève présentation des concepts de la sélection animale, des outils disponibles pour leur modélisation et utilisés pour ces travaux de thèse incluant la

théorie des indices de sélection, l'évaluation génétique, les modèles déterministes et stochastiques utilisés pour décrire l'évolution des valeurs génétiques ou encore apprécier la rentabilité économique d'un programme de sélection. Enfin quelques caractéristiques des programmes de sélection ovins et caprins français sont présentées. Les chapitres 2, 3 et 4, articulés autour des articles publiés sont respectivement consacrés à chacun des objectifs de la thèse. Le dernier chapitre est consacré à la discussion générale des résultats, à leur appropriation par les acteurs de la sélection ainsi qu'aux perspectives en termes de travaux complémentaires.

Chapitre 1 - Bibliographie.

Les programmes de sélection visent à créer et cumuler le progrès génétique pour un ensemble de caractères zootechniques au sein de la population en sélection et diffuser ce progrès à la population commerciale. Nous décrirons premièrement les concepts de la sélection animale puis ferons une présentation de méthodes de modélisation des plans de sélection et d'estimation du progrès génétique. Nous terminerons par la description des enjeux liés à l'assignation de parenté, la gestion des gènes majeurs et la sélection génomique pour les programmes de sélection ovins et caprins.

Concepts de la sélection animale

Le gain génétique annuel

Plusieurs éléments sont pris en compte pour choisir les animaux retenus pour la reproduction, le standard (*i.e.* la conformité aux critères phénotypiques raciaux), la qualité des aptitudes fonctionnelles et la valeur de la descendance pour un ou des caractères d'intérêt. En supposant l'absence d'effet d'interaction entre génome et milieu, le modèle génétique décompose la variance des performances en une variance d'origine génétique et une variance d'origine « environnementale » (effets de milieu). La variance d'origine génétique est elle-même décomposée en une variance génétique additive liée aux effets des allèles (sur la variabilité des performances), une variance de dominance liée aux effets d'interaction entre allèles de chaque

locus et une variance liée aux effets d'interaction entre allèles de différents loci (dénommé épistasie).

Les parents transmettent la moitié de leur génome (*i.e.* allèles) à leur descendance. Sous l'hypothèse du modèle génétique infinitésimal supposant un contrôle des caractères quantitatifs par un grand nombre de gènes ayant chacun un effet faible, les parents transmettent en espérance la moitié de leur valeur génétique additive. En s'appuyant sur le modèle génétique, on peut donc estimer la valeur génétique additive des candidats à la sélection et les sélectionner sur cette base.

Le gain génétique est une mesure de l'évolution de la valeur génétique additive moyenne de la population. Dans des populations sélectionnées en générations séparées, le gain génétique par génération se mesure par la différence entre la valeur génétique additive moyenne des reproducteurs sélectionnés et la valeur génétique additive moyenne de leur cohorte de naissance. Le gain génétique annuel correspond au gain par génération divisé par la durée de l'intervalle de génération exprimé en années. La notion de gain génétique annuel est identique dans les populations sélectionnées en générations chevauchantes mais son estimation prend en compte les intervalles de générations, alors définies comme l'âge moyen des reproducteurs à la naissance de leurs descendants, entre parents (père ou mère) et descendants (fils ou filles).

L'objectif de sélection

Le ou les caractères à améliorer, composant l'objectif de sélection, sont choisis parmi des caractères héritables. Idéalement, ils sont déterminés collectivement par l'ensemble des utilisateurs de la population : élevage en sélection, élevages commerciaux, transformateurs,

consommateurs, etc. Le poids relatif de chaque caractère dans l'objectif de sélection dépend principalement de son importance économique, de ses paramètres génétiques et de son nombre d'expressions.

L'importance économique d'un caractère est généralement appréciée au niveau des unités de production, les élevages, en tenant compte des systèmes d'élevages représentatifs de la race. A l'aide de modèles bioéconomiques, prenant en compte le nombre d'expressions de chaque caractère, les pertes et profits dus à la variation du niveau zootechnique des caractères sont évalués. La connaissance des poids économiques des caractères dans l'objectif de sélection et de leurs paramètres génétiques sont ensuite pris en compte pour déterminer la combinaison qui maximise le gain économique.

Au-delà des éléments d'appréciation strictement économiques, les objectifs de sélection tendent pour certaines espèces à prendre en compte de nouvelles considérations, difficilement chiffrables, telles que le temps de travail ou l'impact environnemental.

L'estimation des valeurs génétiques

Théorie des indices de sélection

La valeur génétique (additive) estimée pour un caractère correspond à l'estimation de l'espérance de la somme des effets des allèles des individus, exprimée en écart à la moyenne phénotypique de la population (μ). Basé sur le principe de la régression linéaire, les valeurs génétiques $\hat{a}$ des individus au sein de la population sont prédites à partir des phénotypes observés y.

$$\hat{a} = cov\,(a, y) * var(y)^{-1} * (y - \mu)$$

Les structures de variance covariances entre valeurs génétiques et performances et entre performances des individus peuvent être exprimées en fonction des coefficients de parentés entre individus (*i.e.* la probabilité d'identité par descendance entre un allèle tiré au hasard chez ces individus) et des paramètres génétiques, notamment l'héritabilité du caractère c'est-à-dire le rapport entre variances génétique additive et phénotypique. Ce principe est la base de la théorie des indices (Hazel, 1943) formalisée pour un objectif de sélection multi-caractères. Dans le cas d'un objectif de sélection $H = \mathbf{w}'\mathbf{a}$ où $\mathbf{a}$ désigne le vecteur des valeurs génétiques de chaque caractère et $\mathbf{w}$ le vecteur de leur pondération dans l'objectif de sélection, l'indice de sélection utilisé est égal à $I = \mathbf{v}'\mathbf{y}$ où $\mathbf{v}$ désigne le vecteur des pondérations des performances et $\mathbf{y}$ le vecteur des performances préalablement corrigées des effets de milieu et centrées. Le vecteur des pondérations qui maximise la corrélation entre H et I est $\mathbf{v} = \mathbf{P^{-1}Gw}$ où $\mathbf{P}$ désigne la matrice de variance-covariance des performances et $\mathbf{G}$ la matrice de variance-covariance entre les valeurs génétiques et les performances. La nécessité de pré-corriger les effets de milieu peut être une source de biais due à la confusion entre effet génétique et effet des niveaux des facteurs de correction (*e.g.* dans une population en sélection, on s'attend à un niveau génétique moyen variable en fonction de l'âge des individus).

Le BLUP

Le développement de la méthode BLUP (Best Linear Unbiased Predictor) basée sur un modèle linéaire mixte a permis une évaluation conjointe des effets de milieu et des effets génétiques (Henderson, 1975). Pour une évaluation mono-caractère avec un seul effet aléatoire, le modèle suivant est utilisé : $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$ où $\mathbf{y}$ (n,1) est le vecteur des n performances, $\mathbf{b}$ (p,1) le vecteur des effets fixes (p niveaux), $\mathbf{u}$ (q,1) le vecteur des effets génétiques des q individus

évalués, **e** (n, 1) le vecteur des résiduelles et **X** (n,p) et **Z** (n,q) des matrices d'incidence reliant performances et effets. **u** et **e** sont des vecteurs aléatoires indépendants qui suivent respectivement des lois normales de moyennes nulles et de variances $\mathbf{G} = \mathbf{A}\sigma_a^2$ (modèle animal) et $\mathbf{R} = \mathbf{I}\sigma_e^2$ où **A** est la matrice de parenté entre les animaux évalués (ses éléments sont égaux à 2 x le coefficient de parentés entre individus), σ_a^2 la variance génétique additive, **I** la matrice identité et σ_e^2 la variance résiduelle. Sous certaines hypothèse (**R** et **G** sont inversibles), les équations du modèle mixte (MME pour mixed model equations) permettant d'estimer **b** et de prédire **u** sont sous la forme : $\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + G^{-1}} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \end{bmatrix}$. Quand les variances résiduelles sont supposées indépendantes et identiques ($\mathbf{R} = \mathbf{I}\sigma_e^2$), la simplification par $\mathbf{R^{-1}}$ donne : $\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z + A^{-1}}\alpha \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{b}} \\ \hat{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X'y} \\ \boldsymbol{Z'y} \end{bmatrix}$ où $\alpha = \sigma_e^2/\sigma_a^2$. Le BLUP a ensuite été étendu à des modèles plus complexes (effets d'environnement permanent, effets maternels, multi-caractères,…) et a été universellement utilisé pour l'évaluation génétique jusqu'à l'arrivée de l'information génomique.

La prise en compte de l'information génomique

La prise en compte de l'information génomique dans l'évaluation génétique a débuté par la valorisation de l'information à un ou quelques locus à effet majeur (Smith et Webb, 1980 ; Fernando et Grossman, 1989 ; Dekkers et van Arendonk, 1998) puis a été généralisé à un nombre de marqueurs plus conséquent, de quelques dizaines à quelques centaines de milliers, répartis uniformément sur le génome (Lande et Thompson, 1990 ; Meuwissen *et al.*, 2001).

La présence d'une mutation avec un effet fort sur les performances est un cas particulier et constitue un écart aux hypothèses du modèle infinitésimal. Préalablement à l'utilisation de marqueurs moléculaires, différentes méthodes, rapportées par Le Roy (1989), ont été proposées pour la détection des gènes majeurs : i) les méthodes basées sur l'observation de la distribution du caractère dans l'ensemble de la population (*e.g.* analyse des mélanges de distributions du caractère pour des populations - F2, backcross- issues de lignées parentales), ii) l'analyse de données relatives à des fratries et la mise en évidence d'une hétérogénéité des types de distributions intra-familles, iii) l'analyse conjointe des performances des parents et de leurs descendants (*i.e.* ressemblance entre la performance d'un individu et la performance moyenne de ses parents) et iv) les analyses de ségrégation qui reposent sur des modèles permettant de tester des hypothèses de transmission génétique d'un caractère (comparaison des vraisemblances des données sachant les modèles testés). La disponibilité des informations apportées par les marqueurs génétiques a ensuite été utilisée pour la détection de gènes majeurs et plus généralement de QTL (locus de caractères quantitatifs - quantitative trait locus) (Soller et Genizi, 1978).

Dès lors que l'information génotypique est connue pour l'ensemble des animaux, une méthode simple à mettre en place pour prendre en compte un tel effet, consiste à inclure le génotype des individus à ce locus particulier comme effet fixe dans un modèle BLUP (Kennedy *et al.*, 1992). Dans ce cas les valeurs génétiques estimées pour les individus sont corrigées pour l'effet de l'allèle au gène majeur. D'autres modélisations, par exemple le « gene content » (Lynch et Walsh, 2008), peuvent être utilisées pour prendre en compte un tel effet. En pratique, seulement une partie des individus utilisés pour l'évaluation génétique dispose d'un génotype. Des méthodes basées sur le principe du « gene content » ont été développées pour le cas de populations partiellement génotypées (Gengler *et al.*, 2008 ; Legarra et Vitezica, 2015).

Concernant la valorisation des données de puce haut débit, il s'agit d'utiliser les marqueurs SNP couvrant le génome pour estimer la valeur des segments chromosomiques auxquels ils sont associés et prendre en compte cette information dans l'estimation de la valeur génétique des individus. Les effets des marqueurs peuvent être estimés sur une population dite de référence comprenant des animaux à la fois génotypés et phénotypés (et/ou apparentés).

Différentes méthodes appartenant aux méthodes de régression pénalisée dont le SNP-BLUP ou méthodes Bayésiennes ont été développées pour estimer ces effets (Robert-Granié *et al.*, 2011 ; Tribout, 2013a). Le premier groupe de méthodes considère une distribution normale des effets des marqueurs tandis que le second autorise une distribution *a priori* non normale des effets. Le BLUP génomique permet d'évaluer directement la valeur génétique d'un individu (équivalence avec le SNP-BLUP) en remplaçant dans les équations du BLUP la matrice de parenté **A** par la matrice de parenté génomique **G** qui caractérise la parenté entre individus sur la base de la ressemblance entre leurs génotypes aux marqueurs. Ces méthodes permettent d'obtenir la valeur génétique des candidats à la sélection dès lors que leur génotypage est disponible soit potentiellement très précocement. Elles nécessitent cependant de réaliser plusieurs étapes : i) obtention des phénotypes des individus appartenant à la population de référence, souvent des pseudo-performances obtenues à partir des évaluations génétiques classiques pour des animaux qui leur sont apparentés, ii) estimation des effets des marqueurs et des valeurs génomiques des individus génotypés, iii) intégration des données génomiques et classiques pour l'ensemble de la population.

Le single step GBLUP (ssGBLUP) permet la prise en compte en une étape de toutes les informations phénotypiques, généalogiques et génotypiques connues pour la population (Legarra *et al.*, 2009). Les apparentements basés sur l'information généalogique (***A***) et

l'information génomique ($\boldsymbol{G}$) sont combinées dans une matrice de parenté $\boldsymbol{H}$ en distinguant les individus non génotypés indicés 1 de ceux génotypés indicés 2

$\mathbf{H} = \begin{bmatrix} \mathbf{A_{11}} + \mathbf{A_{12}A_{22}^{-1}(G - A_{22})A_{22}^{-1}A_{21}} & \mathbf{A_{12}A_{22}^{-1}G} \\ \mathbf{GA_{22}^{-1}A_{21}} & \mathbf{G} \end{bmatrix}$ dont l'inverse, nécessaire pour la résolution des équations du BLUP a une forme relativement simple : $\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G^{-1} - A_{22}^{-1}} \end{bmatrix}$.

Imputation et conséquences sur la précision des évaluations génomiques

Pour des raisons de coût, les candidats à la sélection peuvent être génotypés sur des puces de plus basse densité (LD pour « Low density ») que les puces moyennes densités (≈50 000 SNPs, MD pour « medium density ») et/ou haute densité (≈600 000 SNPs, HD pour « high density ») utilisées pour génotyper la population de référence. Ces puces LD comportent un sous-ensemble des marqueurs SNPs, généralement 5000-15 000 SNPs, présents sur les puces MD.

Les méthodes d'imputation visent à inférer les marqueurs manquants des individus génotypés avec des puces de densités inférieures ou les marqueurs manquants liés à la qualité du génotypage. Les logiciels d'imputation sont basées soit sur des méthodes « populationnelles », par exemple Beagle (Browning et Browning, 2007), qui exploitent le déséquilibre de liaison entre marqueurs avoisinant ou bien des méthodes « familiales », par exemple Fimpute (Sargolzaei, 2014) qui exploitent, généralement en plus du déséquilibre de liaison, la longueur des fragments chromosomiques partagés entre individus en fonction de leur degré d'apparentement : les apparentés proches partagent des haplotypes plus longs tandis que les individus distants partagent des haplotypes plus courts.

La précision de l'imputation peut être mesurée par la corrélation entre génotype imputé et génotype vrai ou par le taux de concordance égal au rapport du nombre de SNPs

correctement imputés sur le nombre de SNPs imputés (Wang, 2013). Cette précision est modulée par des caractéristiques propres à la puce LD utilisée, densité et méthodes de sélection des marqueurs (espacement uniforme, déséquilibre de liaison), ainsi qu'à la taille de la population d'étalonnage (*i.e.* individus génotypés sur les puces de densité supérieure) et à son degré d'apparentement avec la population cible (*i.e.* individus génotypés sur les puces de densité inférieure). La précision est d'autant plus élevée que le nombre de marqueurs LD augmente (Wellmann *et al.*, 2013 ; Cleveland et Hickey, 2013). Néanmoins, le nombre de marqueurs a un effet moindre sur la précision lorsque l'apparentement entre populations d'étalonnage et cible est important et que la méthode d'imputation utilisée valorise cette information généalogique (Toghiani *et al.*, 2016). Un déséquilibre de liaison fort entre marqueurs adjacents et une fréquence élevée de l'allèle mineur moyenne ont un effet favorable sur la précision (Zhang et Druet, 2010, Huang *et al.*, 2012, Wellman *et al.*, 2013). En général, les marqueurs SNP présents sur la puce LD sont uniformément distribués (distance physique) mais la densification en marqueurs au niveau des points chauds de recombinaison et télomères améliore la précision (Wellman *et al.*, 2013, Bolormaa *et al.*, 2015). Le nombre d'individus dans la population d'étalonnage en regard de la taille efficace de la population est un facteur prédominant pour la précision dans le cadre d'une imputation populationnelle (He *et al.*, 2015), car la mesure du déséquilibre de liaison et l'estimation de l'effet des marqueurs sont sensibles au nombre d'individus génotypés. De même, cette mesure dépend de la représentativité de la population d'étalonnage au regard de la population cible. L'enrichissement de la population d'étalonnage par des animaux de race différente n'entraîne pas de gain de précision (Moghaddar *et al.*, 2014). Par contre, dans le cadre d'une imputation familiale, la précision est très sensible au degré d'apparentement entre population d'étalonnage et cible et est peu

sensible au nombre d'individus appartenant à la population d'étalonnage et à la taille efficace de la population (Huang *et al.*, 2012).

Les conséquences sur la précision des valeurs génomiques estimées sont limitées, pertes inférieures à quelques pourcents, pour des précisions d'imputation élevées et importantes (Bolormaa *et al.*, 2015), pertes supérieures à 20%, pour des taux de concordance et corrélation inférieurs à 0.80 (Moghaddar *et al.*, 2014). Entre ces deux niveaux, les conséquences semblent affectées par la méthodologie utilisée pour l'évaluation génomique et l'apparentement entre population de référence et cible.

La diffusion du progrès génétique

Le transfert du progrès génétique de la population en sélection à la population commerciale peut se faire par différents moyen tels que l'insémination ou la vente de reproducteurs mâles ou femelles. En France, L'insémination est principalement utilisée au sein des programmes de sélection et dans les élevages commerciaux de quelques races laitières tandis que la vente d'embryon est quasi-inexistante (Loywyck et Lagriffoul, 2016 ; Fatet *et al.*, 2008). Le mode de diffusion d'une race donnée est plutôt orienté sur la vente de mâles ou bien de femelles en fonction du mode d'utilisation de la race dans les élevages de production : diffusion de mâles pour une utilisation en race pure ou pour le croisement et diffusion de femelles utilisées comme support maternel pour le croisement.

Déterminants du progrès génétique annuel

Le progrès génétique annuel au sein du programme de sélection, ΔG_a, est dépendant de quatre facteurs : l'intensité de la sélection i, la précision des évaluations génétiques ρ, la variabilité génétique σ_g et l'intervalle de génération IG, $\Delta G_a = \frac{i*\rho*\sigma_g}{IG}$.

L'intensité de la sélection caractérise la sévérité du choix des reproducteurs. Si le tri est strictement réalisé sur les valeurs génétiques estimées, l'intensité de sélection est fonction du taux de retenus (pression de sélection) correspondant au ratio du nombre de reproducteurs retenus sur le nombre de candidats disponibles. Une intensité de sélection élevée est favorable au progrès génétique par génération.

La précision de l'estimation des valeurs génétiques correspond à la corrélation entre le critère de sélection et la valeur génétique vraie.

En l'absence d'information génomique, cette corrélation dépend essentiellement de la quantité et qualité de l'information phénotypique connue pour l'individu évalué (*i.e.* nombre de phénotypes, degré d'apparentement des individus qui expriment ces phénotypes) et de l'héritabilité du ou des caractères. Elle peut être théoriquement obtenue à partir de l'expression des variances-covariances telles que formalisées par la théorie des indices ou plus difficilement pour le BLUP à partir de l'inverse de la matrice des coefficients des équations du modèle mixte et du paramètre α. Compte-tenu d'un nombre élevé de performances et individus dans les évaluations en routine, de quelques dizaines de milliers à quelques millions pour les populations ovines et caprines, les précisions des estimations sont généralement approximées dans le cas de l'utilisation du BLUP.

En plus de la quantité d'information phénotypique disponible, les principaux facteurs modulant la précision des valeurs génétiques en présence d'information génomique sont la taille efficace de la population (nombre d'individus d'une population idéale, au sens de Falconer (1960), pour lequel on aurait un degré de dérive génétique équivalent à celui de la population réelle) dont dépend le nombre de segments chromosomique indépendants, la taille de la population de référence, l'héritabilité du caractère et le nombre de marqueurs (Hayes et Goddard, 2008). Si l'existence de ces effets est démontrée, leur niveau est modulé par le degré d'apparentement entre les individus génotypés et disposant d'information phénotypique (sur performances propres et/ou d'apparentés) et les individus génotypés candidats à la sélection (Clark *et al.*, 2012). La présence d'apparentés dans la population de référence accroît la précision des candidats à la sélection.

La variabilité génétique est une donnée propre à la population et au caractère donné. Plusieurs facteurs affectent l'évolution de la variabilité génétique au cours du temps, migration, mutation, dérive génétique et sélection (Falconer, 1960). Deux facteurs induisent une diminution de la variabilité génétique : la sélection et la limitation des effectifs qui entraînent une augmentation de la consanguinité moyenne et une baisse de l'hétérozygotie moyenne.

L'effet de la sélection sur la variabilité génétique de la population au démarrage de la sélection, induisant un déséquilibre de liaison entre loci, a été formalisé par Bulmer (1976). La perte de variabilité génétique est dépendante de l'intensité de la sélection et se produit lors des premiers cycles de sélection. La variance génétique se stabilise ensuite à une valeur limite. Si les paramètres génétiques sont estimés sur une population en cours de sélection, la variance génétique estimée est alors plus proche de cette valeur limite. (Ollivier, 2002).

La taille limitée de la population, le déséquilibre entre nombre de mâles et femelles et le déséquilibre de taille des descendances, notamment des mâles, sont des facteurs qui favorisent la dérive génétique, c'est-à-dire la perte aléatoire d'allèles par échantillonnage au cours des cycles de reproduction successifs, et contribuent ainsi à une réduction de la variabilité génétique.

Des modes de gestion des populations basés principalement sur l'exploitation de l'information généalogique et plus récemment de l'information génomique, permettent de réduire la perte de variabilité liée dans les populations sélectionnées. Ces méthodes visent, dans le cadre d'une organisation de sélection bien établie, à choisir les reproducteurs et déterminer les accouplements en tenant compte de l'évolution de la consanguinité. Plusieurs méthodes d'optimisation existent telles que l'utilisation de fonction de pénalité (Henryon, 2015) ou d'optimisation sous contraintes, *i.e.* application d'une contrainte sur le niveau maximal d'accroissement de consanguinité autorisé (Meuwissen, 1997).

L'intervalle de génération est défini comme l'âge moyen des reproducteurs à la naissance de leurs descendants, eux-mêmes candidats à la sélection. La longueur de l'intervalle de génération dépend de la contribution de chaque classe d'âge de reproducteurs mâles et femelles au renouvellement. Cette contribution est fonction de l'organisation du programme de sélection incluant le(s) mode(s) de reproduction des reproducteurs et leurs limites zootechniques.

L'intensité de sélection, la précision et l'intervalle de génération ne sont pas indépendants. Favoriser l'un des facteurs se fait généralement au détriment d'au moins un des deux autres. Par exemple le gain en précision lié à la mise en place d'une évaluation sur descendance des

mâles s'accompagne d'un accroissement de l'intervalle de génération moyen. De même pour un nombre total de filles de testage contraint, un accroissement de l'intensité de sélection obtenu par une augmentation du nombre de mâles mis en testage s'accompagnent par une réduction de la précision de leur valeurs génétiques estimées due à un moindre nombre de filles par mâle. L'optimisation du programme de sélection revient à déterminer la combinaison optimale de ces trois facteurs qui maximise le progrès génétique tout en maîtrisant la perte de variabilité génétique.

En théorie, si le mode de diffusion du progrès génétique est constant au cours du temps et que la stratégie de sélection est stable et appliquée depuis plusieurs générations, les gains génétiques aux niveaux du programme de sélection et de la population commerciale sont identiques. Cependant il y a un décalage entre le niveau moyen des reproducteurs du noyau et de la population commerciale, en général au détriment de la population commerciale, qui dépend du progrès génétique annuel et de son mode de diffusion (Bichard, 1971 ; Elsen, 1993).

Economie de la sélection

En général le traitement des aspects économiques met en œuvre le calcul des éléments de recettes et de coûts sur un horizon de temps donné (Poutous *et al.*, 1962 ; Hill, 1971). La prise en compte de l'ensemble des coûts de la sélection, liés par exemple à l'enregistrement des performances ou encore au calcul des valeurs génétiques, est fastidieuse. Les coûts sont divers et le périmètre difficile à définir notamment entre activités d'élevage et sélection, la seconde activité étant indissociable de la première. On peut donc se limiter à la seule prise en compte

des éléments de coûts variables entre les stratégies que l'on souhaite comparer. Les recettes sont généralement liées à l'expression du caractère sous sélection (et caractères corrélées). Ces recettes sont décalées dans le temps et s'échelonnent au cours des générations suivantes selon des modalités qui dépendent du caractère (*e.g.* caractères répétés tout au long de la carrière de l'animal), des propriétés du progrès génétique (cumul du gain au cours des générations) et des échanges de reproducteurs au sein du noyau (organisation de la sélection) et entre le noyau et la population commerciale (organisation de la diffusion). Afin de tenir compte de leur échelonnement dans le temps, les coûts et recettes sont actualisés (Weller, 1994). Les méthodes d'actualisation permettent de considérer l'évolution de la valeur de l'argent au cours du temps. Cette évolution dépend de deux facteurs : i) le coût du temps, un euro aujourd'hui vaut plus qu'un euro demain (autrement dit si j'investis un euro aujourd'hui, j'aurai demain un euro augmenté de l'intérêt de mon investissement) et ii) le coût du risque, un euro certain vaut plus qu'un euro espéré mais incertain. Les flux financiers F espérés au cours du temps t sur un horizon H sont donc rapportés à leur valeur actuelle Fa en tenant compte d'un taux d'actualisation d. $Fa = \sum_{t=1}^{t=H} F(t) * (1+d)^{-t}$. La plupart des études basées sur des recettes et coûts actualisés dans les programmes de sélection animaux utilisent des taux d'actualisation compris entre 5 et 15% (Weller, 1994).

Les principes du flux de gènes actualisé peuvent également être utilisés pour déterminer le nombre d'expressions génétiques actualisées d'un reproducteur selon son mode d'utilisation (*e.g.* utilisation terminale ou descendants retenus pour la reproduction) et du mode d'expression du caractère considéré (Elsen, 1977 ; Amer, 1999). Ce nombre d'expressions est ensuite utilisé pour quantifier la valeur économique d'un animal en fonction de sa valeur génétique.

Modélisation des plans de sélection et estimation du progrès génétique.

Objectifs de la modélisation des programmes de sélection

La modélisation des programmes de sélection vise à traduire leur fonctionnement réel en relations algébriques pour prédire leur efficacité technique et économique (Elsen, 1992). Les critères d'efficacité technique utilisés sont généralement le progrès génétique annuel et la vitesse d'accroissement de la consanguinité. Les critères d'efficacité économique sont moins universels. Le bénéfice (recettes-coûts) et le taux de rendement du capital investi ((recette-coûts) / coûts) ou encore le délai de récupération (durée pour laquelle les recettes équivalent aux coûts) sont des critères d'efficacité économique communément utilisés (König *et al.*, 2016, Shumbusho *et al.*, 2016).

D'autres approches consistent par exemple i) à comparer des stratégies sur leurs critères techniques mais en contraignant leurs coûts de telle sorte qu'ils soient identiques, ii) à comparer le coût du gain d'un écart-type génétique de différentes stratégies (Schaeffer, 2006) ou encore iii) à estimer les gains génétiques de nouvelles stratégies en fonction de leurs coûts additionnels par rapport à une stratégie de référence (Tribout *et al.*, 2013b).

Modèles déterministes et stochastiques

Deux types de modèles sont utilisés pour la modélisation des programmes de sélection. Dans les modèles déterministes, le programme de sélection est complètement décrit par des

relations entre variables (Elsen, 1992). Pour un jeu de données en entrée du modèle correspond un résultat unique en sortie. Ces modèles, associés généralement à des temps de calcul très courts, permettent d'évaluer un très grand nombre de situations (*i.e.* jeu de données en entrée du modèle). Les modèles stochastiques se basent sur la simulation des individus et de leur génome. Les valeurs génétiques vraies des individus, leur(s) phénotype(s) ou encore certains évènement tels que les évènements démographiques dépendent de tirages aléatoires dans des distributions connues *a priori*. Ces modèles permettent de considérer des situations difficiles à décrire par des relations déterministes telles que la précision de l'évaluation génomique. Ces modèles sont néanmoins associés à des temps de calcul importants ce qui limite le nombre de situations qu'il est possible d'évaluer.

Analyse démographique de la population modélisée

L'analyse démographique est nécessaire à l'établissement des modèles qu'ils soient déterministes ou stochastiques. La population est définie comme un ensemble d'individus. L'unité de temps pour décrire la population dépend de l'espèce considérée et de l'objet de l'étude. En général la durée d'un cycle de reproduction ou d'un intervalle de génération est choisie (Khang, 1983). Au sein de cette population, une stratification plus ou moins fine en classes est précisée en fonction de certaines caractéristiques telles que l'appartenance des individus à une race, à un programme de sélection (population en sélection ou population commerciale), à un élevage donné ou encore à un mode de reproduction (Elsen, 1992). Le degré de stratification dépend des facteurs engendrant une variabilité du niveau génétique ou de la précision de son estimation entre individus. Les classes, utilisées pour la modélisation de la

population dans le cadre des modèles déterministes, sont ainsi constituées d'individus dont les attributs sont homogènes et soumis aux mêmes règles de fonctionnement.

L'analyse démographique consiste à décrire l'état de la population (sa taille, ses effectifs par classe) et son évolution au cours du temps, évolution déterminée par les règles de fonctionnement (reproduction, sélection, migration) relatives à chacune des classes. Cette analyse repose sur deux visions complémentaires, la vision transversale, observation des phénomènes démographiques (entrées et sorties d'individus) à un temps donné et la vision longitudinale, observation des phénomènes démographiques au cours du temps pour un groupe d'individus appartenant à une même cohorte et soumis aux mêmes règles de fonctionnement (Khang, 1983).

Paramètres et variables utilisés pour la modélisation

Les paramètres et variables inclus dans un modèle sont de plusieurs ordres : variables d'état, paramètres de la population, variables de décision et variables internes (Elsen, 1992).

L'utilité des variables d'état est liée à la subdivision de la population en classes homogènes telle que décrite précédemment. Chacune des classes est caractérisée par une ou plusieurs variables d'état qui lui est propre, par exemple la valeur génétique moyenne des individus la composant et qui est susceptible d'évoluer au cours du temps dans les modèles dynamiques (versus asymptotiques). Les variables d'état sont définies afin de calculer les différents critères de mesure de l'efficacité technique et économique de la population.

Les paramètres sont les variables « imposées » et sont spécifiques d'une situation donnée (population*caractère étudié, prix d'une technologie,…). On peut distinguer les

paramètres génétiques (*e.g.* héritabilité du caractère), démographiques (*e.g.* taux de mortalité d'une classe d'âge), zootechniques (*e.g.* âge à la puberté) et économiques (*e.g.* coût unitaire d'une insémination). Bien que ces paramètres soient imposés, ils sont susceptibles de varier en fonction de la population ou bien au cours du temps. Il convient donc d'évaluer la sensibilité des résultats issus de la modélisation à ces paramètres.

Les variables de décision permettent de décrire les composantes du programme de sélection. Paramètres et variables de décision sont reliés aux variables d'état par l'intermédiaire de variables internes. Les relations entre variables permettent de traduire les contraintes démographique, biologique ou organisationnelle ainsi que d'établir leurs relations fonctionnelles. L'optimisation d'un programme de sélection revient donc à déterminer la combinaison des variables de décision qui maximise son efficacité technique et/ou économique.

Estimation du progrès génétique dans le cadre d'un modèle déterministe

Différentielle de sélection

La différentielle de sélection résultant d'une sélection par troncature sur indice de sélection correspond à la valeur moyenne des indices des individus sélectionnés exprimés en écart à la moyenne des candidats à la sélection. Le calcul des différentielles de sélection est réalisé pour chacune des sub-populations ou classes du modèle qui sont caractérisées par i) une distribution des indices des candidats à la sélection et ii) une proportion d'individus sélectionné.

Estimation du progrès génétique annuel asymptotique

Dans ce type de modèle, seuls les effets à long terme de la sélection sont pris en compte. La définition des classes se limite généralement à celles nécessaires pour dissocier les quatre voies du progrès génétiques : mâles à mâles MM, mâles à femelles MF, femelles à mâles FM et femelles à femelles FF.

Trois relations sont établies pour dériver la formule du progrès génétique asymptotique (Rendel et Robertson, 1950) :

1) La valeur génétique moyenne des candidats mâles M et femelles F au temps t dépend de celle de leurs parents au temps t-1.

$$M(t) = 0.5 * \left[MM_{(t-1)} + FM_{(t-1)}\right] \qquad F(t) = 0.5 * \left[MF_{(t-1)} + FF_{(t-1)}\right]$$

2) Pour une catégorie donnée, l'écart entre t-1 et t dépend du progrès génétique annuel Δ_g et de l'intervalle de génération L de cette catégorie

$$MM_{(t-1)} = MM_{(t)} - L_{MM}\Delta_g \qquad FM_{(t-1)} = FM_{(t)} - L_{FM}\Delta_g$$

$$MF_{(t-1)} = MF_{(t)} - L_{MF}\Delta_g \qquad FF_{(t-1)} = FF_{(t)} - L_{FF}\Delta_g$$

3) La valeur génétique des reproducteurs sélectionnés au temps t dépend de la valeur génétique des candidats et de la différentielle de sélection Δ appliquée pour la catégorie

$$MM_{(t)} = M(t) + \Delta_{MM} \qquad FM_{(t)} = F(t) + \Delta_{FM}$$

$$MF_{(t)} = M(t) + \Delta_{MF} \qquad FF_{(t)} = F(t) + \Delta_{FF}$$

En combinant l'ensemble des relations, on peut exprimer le progrès génétique annuel comme la somme des différentielles de sélection divisée par la somme des intervalles de génération :

$$\Delta_g = \frac{\Delta_{MM}+\Delta_{MF}+\Delta_{FM}+\Delta_{FF}}{L_{MM}+L_{MF}+L_{FM}+L_{FF}}$$

Cette expression peut être adaptée pour des cas où l'ensemble des individus d'un sexe donné peuvent être issus de différentes catégories de pères et de mères. Dans ce cas, la formulation devient plus complexe et peut présenter des difficultés d'interprétation.

Estimation du progrès génétique annuel par la méthode de flux de gènes

Le modèle de flux de gènes (Hill, 1974 ; Elsen et Mocquot, 1974) est un modèle dynamique qui permet de décrire l'évolution des valeurs génétiques moyennes par classe d'individu au cours du temps. Le vecteur des valeurs génétiques moyennes par classe au temps $t+1$, $\mathbf{g_{t+1}}$ dépend du vecteur $\mathbf{g_t}$, d'une matrice de passage $\mathbf{P_t}$ décrivant les flux de gènes entre t et $t+1$ et du vecteur $\mathbf{d_t}$ des différentielles de sélection appliquées sur chaque classe : $\mathbf{g_{t+1}} = \mathbf{P_t} * \mathbf{g_t} + \mathbf{d_t}$

En général les paramètres démographiques et décisions de sélection sont supposées constantes au cours du temps et $\mathbf{P_t} = \mathbf{P}$ et $\mathbf{d_t} = \mathbf{d}$. Dans ce cas l'estimation du progrès génétique quand t tend vers ∞ est égale à celle du modèle asymptotique précédent (Hill, 1974 ; Elsen et Mocquot, 1974). L'avantage du modèle de flux de gènes est sa capacité à décrire l'évolution parfois erratique du niveau génétique de la population lors de la mise en place d'une stratégie, description nécessaire pour la mesure de sa rentabilité économique.

La méthode de flux de gènes ne considère pas l'hérédité de l'avantage sélectif entre générations successives. La qualité de prédiction des proportions asymptotiques de gènes provenant des classes âge*sexe de reproducteurs par cette méthode a ainsi été remise en cause (Woolliams *et al.*, 1999 ;Bijma et Woolliams, 1999). Cependant, bien que le lien entre progrès

génétique asymptotique et proportions asymptotiques de gènes provenant des classes âge*sexe (dépendant des cycles de sélection successifs) soit remis en cause, la méthode de flux de gènes permet de prédire très correctement le gain génétique asymptotique (Bijma et Woolliams, 2000).

Estimation de la précision des valeurs génétiques, du progrès génétique et de l'évolution de la consanguinité dans le cadre d'un modèle stochastique

Dans un modèle stochastique, certaines variables sont aléatoires et donc associées à une distribution de probabilités. Les génomes sont simulés et les effets des QTLs sont aléatoirement obtenus par tirage dans une distribution connue *a priori*. La valeur génétique vraie de chaque individu dépend de leurs génotypes aux QTL. Les phénotypes peuvent être simulés en se basant sur le modèle génétique qui suppose que les phénotypes sont la somme d'un effet génétique et d'un effet de l'environnement. Les phénotypes d'un individu sont simulés en fonction de leur valeur génétique vraie et des paramètres génétique du caractère considéré (héritabilités, répétabilités). Ces phénotypes sont alors utilisés pour le calcul des valeurs génétiques estimées, par exemple par un BLUP dans le cadre classique ou par GBLUP ou single step GBLUP en présence d'information génomique. Le gain génétique annuel peut être estimé par la régression sur le temps de la moyenne des valeurs génétiques d'une catégorie de la population, de préférence avant sélection (cas des femelles en première parité). De même, la vitesse d'accroissement annuel de la consanguinité peut être estimée par la régression sur le temps de la moyenne des coefficients de consanguinité. Les précisions peuvent être estimées pour différentes catégories d'animaux par le coefficient de corrélation

de Pearson entre les valeurs génétiques vraies et estimées. Pour un ensemble de valeurs des variables de décision, on obtient une variabilité de réponses liée à la prise en compte des phénomènes aléatoires par le modèle. Cette variabilité de réponses renseigne sur les stratégies évaluées, leur stabilité (instabilité) relative. Il est ainsi possible d'inclure la notion de risque : deux stratégies donnant (quasiment) la même espérance de progrès génétique peuvent être associées à des variabilités (de réponses) très différentes. La variabilité de réponse permet par ailleurs de tester si les différences d'espérances, par exemple de progrès génétique, entre stratégies sont significatives.

Optimisation d'un programme de sélection

Pour une situation donnée, c'est-à-dire une combinaison de valeurs prises par les variables de décision, les modèles permettent d'évaluer une fonction objectif. Cette fonction peut être strictement technique et combiner les critères de gain génétique annuel et taux d'accroissement de la consanguinité ou bien économique en intégrant le calcul des recettes et coûts liés à la situation analysée. Pour l'optimisation des programmes de sélection de populations animales, les méthodes sont le plus souvent appliquées sur des modèles déterministes qui, par essence, possèdent une solution unique pour une combinaison donnée des variables de décision. L'intervalle de variation des variables de décision ou zone de faisabilité est défini par un ensemble de contraintes (Gill *et al.*, 1981). Il s'agit ensuite de déterminer la combinaison des valeurs qui maximise ou minimise la fonction objectif par des algorithmes d'optimisation sous contraintes tels que les méthodes d'optimisation quadratique, algorithme du recuit simulé ou algorithmes génétiques (Dréo *et al.*, 2005). Costard et Elsen (2011) ont comparé l'optimisation

d'une sélection assistée par gène, à l'aide d'un algorithme génétique, dans le cadre d'approches stochastiques et déterministes. Ils ont montré la quasi-équivalence entre les solutions optimales produites par les deux types de modèle pour des populations de taille infinie. Dans le cas de populations de taille finie, le modèle déterministe surestime le gain génétique et produit un jeu de variables de décisions sub-optimales.

Les programmes de sélection des populations ovines et caprines en France

Les modèles présentés dans cette thèse ayant été développés pour répondre aux enjeux techniques et économiques de l'utilisation d'un panel SNP très basse densité dans les populations ovines et caprines françaises, cette partie de la bibliographie aborde plus particulièrement leurs caractéristiques.

Organisation générale des programmes de sélection

Dans les programmes de sélection, les objectifs de sélection, c'est-à-dire l'ensemble des caractères pris en compte lors de la sélection des animaux, sont définis par race. Pour les races laitières ovines et caprine, les objectifs intègrent en partie ou totalité les quantités de lait, qualités de lait - matières grasses et protéiques, numération cellulaire - et aptitudes fonctionnelles - morphologie de la mamelle (Palhière *et al.*, 2015 ; Barillet *et al.*, 2016). En races ovines allaitantes, les objectifs de sélection incluent des aptitudes maternelles - capacité d'allaitement des agneaux, viabilité et prolificité - et notamment pour les races dites « lourdes » des aptitudes bouchères – croissance post-sevrage, état d'engraissement et

développement musculaire (Cheype *et al.*, 2013). D'autres caractères, ne bénéficiant pas d'évaluation génétique officielle (standard de race, capacité de reproduction, …) ou dont les modèles d'évaluation génétique sont en cours de développement (résistance au parasitisme) sont également pris en compte lors de la sélection des reproducteurs.

Phases	Caractères inclus dans l'ISOL[1]	Races		
		Lacaune	ROLP [2]	Corse
Conception des outils		1955-1965		
Démarrage (lait)	QL	1965-1975	1975-1987	1995-2005
Croisière (lait)	QL	1975-1985	1987-2000	2005-
Démarrage (taux)	QMG, QMP	1985-1992	2000-2008	
Croisière (taux)	QMG, QMP, TB, TP	1992-2005	2008-2016	
Démarrage (fonctionnel)	QMG, QMP, TB, TP, CCS, morpho [3]	2005-2010	2016-	
Croisière (fonctionnel)	QMG, QMP, TB, TP, CCS, morpho [3]	2010-		

[1] plus sélection pour la résistance à la tremblante (gène PrP, à partir de 1998 ou 2001 selon la race).
[2] ROLP = Races Ovines Laitières des Pyrénées.
[3] morphologie mammaire externe.

Tableau 1 : Niveau d'organisation et évolution dans le temps du schéma de sélection de chaque race de brebis laitières françaises (sources :idele, INRA et CNBL 2015) – paru dans Barillet *et al.*, 2016.

Caractère	Unité du caractère	Système laitier			Système fromager		
		Pondération pour une unité du caractère (€)	Pondération pour un écart-type génétique du caractère (€)	Pondération relative	Pondération pour une unité du caractère (€)	Pondération pour un écart-type génétique du caractère (€)	Pondération relative
Vecteur lait	+ 1 Kg par lactation	+0,11	+8,98	19,0 %	-0,03	-2,45	2,0 %
MG	+ 1 Kg par lactation	+0,40	+1,28	2,7 %	+5,30	+17,01	14,0 %
MP	+ 1Kg par lactation	+8,20	+18,94	40,2 %	+31,10	+71,84	59,0 %
CCS	- 1 point de SCS par lactation	+7,20	+3,88	8,2 %	+4,30	+2,32	1,9 %
Morphologie	+ 1 point d'IMC	+1,20	+14,1	29,9 %	+2,40	+28,20	23,1 %

Tableau 2 : Valeur des pondérations économiques des cinq caractères modélisés, en race Saanen (Palhière *et al.*, 2015).

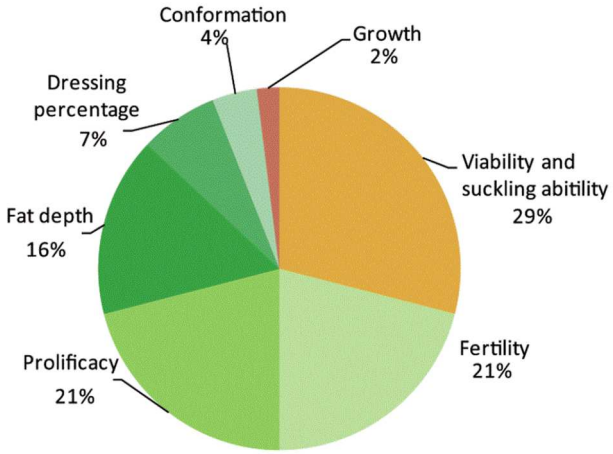


Figure 1 : Objectif de sélection de la race Blanche du Massif Central déterminé à partir des poids économiques relatifs de chaque caractère (Cheype *et al.*, 2013).

L'information généalogique, les phénotypes, caractéristiques physiologiques et la description du milieu sont recueillis soit en élevage notamment pour les caractères exprimés par les femelles (contrôle de performances), soit au niveau des outils collectifs de gestion des mâles, centres d'élevage des jeunes mâles et centres d'insémination (Ménissier et Bouix, 1992 ; Cheype, 2016 ; Douguet *et al.*, 2017). Quelles que soient les populations, les filiations maternelles sont déduites au moment de la mise-bas qui se déroule généralement sous surveillance des éleveurs. En conséquence les filiations maternelles sont quasi-complètes.

Le mode d'obtention de la filiation paternelle est intimement lié au mode de reproduction des mères des jeunes candidats à la sélection. Les pères sont déduits de la déclaration lors de la mise en lutte de la mère de l'individu. La qualité de la filiation est dépendante de l'application de règles strictes pendant la période de lutte que ce soit par

insémination artificielle ou monte naturelle (utilisation de la monte en main ou lot de lutte limité à un seul mâle). Par exemple en ovin, les lots de luttes naturelles doivent être réalisés avec un seul mâle et la lutte doit être interrompue pendant au moins 12 jours avant l'introduction d'un nouveau mâle dans le lot. Concernant les inséminations, elles doivent être faites "en paternité" (semence provenant d'un seul mâle) et sans repasse avant 12 jours après la mise en place. Dans le cas de lots de lutte comprenant plusieurs mâles ou de repasse immédiate après insémination, la filiation ne peut être déduite de la lutte. Dans des élevages, dont la taille s'accroît et dont la main-d'œuvre est de plus en plus restreinte, ces règles deviennent très contraignantes pour les luttes naturelles. Elles imposent la conduite de nombreux petits lots, activité chronophage, et difficilement réalisable pour certaines conduites d'élevage (pâturage, transhumance).

Très majoritairement, les pères des candidats mâles à la sélection sont identifiés. Pour les femelles, les taux de filiations paternelles connues sont plus variables. Dans les élevages caprins en contrôle laitier officiel, le taux avoisine 50%. Dans l'espèce ovine, les taux varient de 24% en moyenne chez les races allaitantes rustiques à 82% en moyenne dans les races allaitantes bouchères et races laitières (Douguet *et al.*, 2017 ; Tiphine *et al.*, 2017).

Le mode de sélection des reproducteurs diverge selon le sexe. Les femelles sont triées en élevages en fonction des valeurs génétiques de leurs parents, de leur standard et de la qualité de leurs aptitudes fonctionnelles et mises ensuite à la reproduction. Pour les principales races ovines laitières, l'utilisation de l'insémination est prédominante avec un taux de brebis inséminées supérieur à 60% en race Manech Tête Rousse et proche de 100% en Lacaune. Excepté pour la race Lacaune viande pour laquelle le taux est très élevé, les autres races ovines laitières, les races caprines Alpine et Saanen et les races ovines allaitantes utilisant

l'insémination ont des taux de femelles inséminées compris entre 20 et 50% (Fatet *et al.*, 2008 ; Loywyck et Lagriffoul, 2016). Une partie des races ovines allaitantes utilise exclusivement la monte naturelle.

Race	brebis agnelées	paternité connue	mises-bas sur IA	
			nombre	brebis agnelées
AURE ET CAMPAN	1220			
AVRANCHIN	283	50,3 %		
BAREGEOISE	2560	0,4 %		
BERRICHON DE L'INDRE	1167	52,0 %		
BERRICHON DU CHER	2798	92,3 %	1 117	39,6 %
BIZET	3158	41,5 %		
BLANC DU MASSIF CENTRAL	22115	33,1 %	4 331	16,8 %
BLEU DU MAINE	1288	91,2 %		
BOULONNAISE	1787	69,1 %		
CASTILLONAISE	1580	6,6 %		
CAUSSENARDE DES GARRIGUES				
CAUSSENARDE DU LOT	28900	6,5 %	2 076	6,5 %
CHARMOISE	2048	69,5 %	4	0,2 %
CLUN FOREST	267	84,4 %		
COTENTIN	105	95,2 %		
DORSET DOWN	113	98,2 %		
EST A LAINE MERINOS	2962	28,8 %	307	9,4 %
FINNOISE	159	32,1 %		
GRIVETTE	5815	32,4 %		
HAMPSHIRE	808	80,7 %		
ILE DE FRANCE	10823	83,1 %	1 707	15,6 %
LACAUNE VIANDE	15107	44,8 %	7 237	39,1 %
LACAUNE VIANDE GID	1409	67,9 %	969	58,1 %
LIMOUSINE	9096	31,0 %	446	4,6 %
LOURDAISE	482	2,5 %		
MARTINIK	584	71,4 %		
MERINOS D'ARLES	13200	11,8 %	250	1,8 %
MERINOS DE RAMBOUILLET	105	91,4 %		
MONTAGNE NOIRE	869	23,9 %		
MOUREROUS	4800			
MOUTON CHAROLLAIS	7112	96,9 %	767	10,8 %
MOUTON VENDEEN	7707	86,6 %	1 263	15,9 %
NOIRE DU VELAY	6395	36,8 %	136	1,7 %
PREALPES DU SUD	5000	18,9 %	217	4,2 %
RAVA	8279	42,9 %		
ROMANE	23166	25,7 %	2 402	9,7 %
ROMANOV	561	23,7 %		
ROUGE DE L'OUEST	3834	82,5 %	522	13,3 %
ROUGE DU ROUSSILLON				
ROUSSIN	1134	87,4 %		
SOLOGNOTE	2297	73,9 %		
SOUTHDOWN	317	95,6 %		
SUFFOLK	3143	90,5 %	462	14,7 %
TARASCONNAISE	7990	22,2 %	228	2,8 %
TEXEL	4224	93,0 %	395	9,3 %
THONES ET MARTHOD				

Tableau 3 : Nombre de brebis agnelées, % de mises-bas issues d'une lutte en paternité, nombre et % de mises-bas sur IA – races allaitantes, campagne 2016 (paru dans Tiphine *et al.*, 2017).

L'organisation des accouplements ne concerne que les seules femelles mises à l'IA lorsque ce mode de reproduction est dominant. Les femelles luttées par monte naturelle sont incluses dans le cas contraire. Afin de produire de jeunes candidats mâles, les meilleures femelles, choisies sur valeurs génétiques, sont accouplées aux meilleurs mâles, par exemple aux mâles d'insémination sélectionnés sur descendance pour les programmes mettant en œuvre ce type d'évaluation. Dans certaines populations, un tri des plus mauvaises femelles sur leur valeur génétique estimée est réalisé après leur première performance mais, plus généralement, la réforme opérée après leur entrée en production résulte de causes indépendantes de cette valeur. Le pourcentage de femelles contrôlées est très variable entre les différentes populations, plutôt faible en ovin allaitant avec près de 8% de la population, intermédiaire en caprin avec près de 35% et élevé en ovin lait avec 20% en contrôle laitier officiel, base de l'amélioration génétique et supérieur à 60% en incluant le contrôle simplifié (Douguet *et al.*, 2017 ; Tiphine *et al.*, 2017).

Races	Nbre de brebis *	Brebis contrôlées *	Béliers			IA *
			Génotypés	*Centre d'élevage* *	*Testés sur descendance* **	
Lacaune	830 000	680 551	2 524	2 524	272	414 538
Races pyrénéennes						
- Basco-béarnaise	80 000	30 664	-	79	44	15 201
- Manech tête noire	95 000	21 619	-	44	26	6 042
- Manech tête rousse	290 000	105 891	-	255	146	61 307
Corse	85 000	31 067	-	296	24	7 234
Total	1 380 000	869 792	2 524	3 198	512	504 322

Tableau 4 : Effectifs ovins laitiers. *Données 2016, **Données 2015/2016. Sources : Institut de l'élevage, Association Nationale de l'Insémination Ovine, Comité Nationale Brebis Laitières (paru dans « Dispositif génétique chiffres clefs ruminants 2016 » – France Génétique Elevage).

Races	Nombre de chèvres *	Chèvres contrôlées	Boucs testés sur descendance *	IA *
Alpine	476 000	168 611	44	42 503
Saanen	349 000	120 215	33	27 180
Autres	25 000	9 147	-	-
Total	850 000	297 973	77	69 683

Tableau 5 : Effectifs caprins laitiers. * Période 01/10/2014 au 30/09/2015. Sources : Institut de l'élevage, France Contrôle Elevage, Capgènes (paru dans « Dispositif génétique chiffres clefs ruminants 2016 » – France Génétique Elevage).

Races	Nombre de brebis *	Brebis contrôlées *	Béliers centre d'élevage	Béliers testés sur descendance	IA
Races bouchères	962 800	82 670	1 214	78 (boucher)	179 576
Races rustiques	2 449 700	210 385	2 580	59 (boucher) 57 (maternel)	132 952
Croisées/Autres	385 500	33 115	-	-	-
Total	3 798 000	326 170	3 794	194	312 528

Tableau 6 : Effectifs ovins allaitants. *Données campagne 2016. Races bouchères : Berrichon du Cher, Charmoise, Ile de France, Rouge de l'Ouest, Suffolk, Texel, Vendéenne. Races rustiques : Blanc du Massif Central, Causses du Lot, Lacaune, Limousine, Mérinos d'Arles, Noire du Velay, PréAlpes du sud, Rava, Romane, Tarasconnaise. Sources : Institut de l'élevage, Association Nationale de l'Insémination Ovine (paru dans « Dispositif génétique chiffres clefs ruminants 2016 » – France Génétique Elevage).

La sélection des mâles est généralement réalisée collectivement au niveau du noyau de sélection, que ce soit pour une utilisation en monte naturelle ou bien par insémination artificielle. Les males sont premièrement sélectionnés sur les mêmes critères que les femelles, standard et aptitudes fonctionnelles, auxquels peuvent s'ajouter des critères de gestion de la variabilité génétique et de l'information génomique pour un locus particulier tel que le gène

PrP (associé à la résistance à la tremblante classique) en ovin. Hormis pour la croissance post-sevrage en ovin allaitant (< 10% des élevages en contrôle), les mâles ne sont pas évalués en élevages.

Les individus issus des meilleurs parents intègrent les centres d'élevage au niveau desquels est réalisée une sélection sur performances propres (aptitudes bouchères, aptitudes à la reproduction,...) ou sur index génomique tel qu'en race Lacaune lait (Astruc *et al.*, 2016). Environ 3200 et 3800 jeunes béliers par an entrent dans ces structures en ovin lait et viande et environ 120 boucs par an dans les centres caprins.

A l'issue du contrôle individuel, les mâles sont d'abord sélectionnés pour l'insémination. Les mâles pour la lutte en monte naturelle dans le noyau sont ensuite choisis parmi les candidats non retenus pour l'IA. Les mâles non sélectionnés sont diffusés dans les élevages commerciaux ou éliminés. Dans les schémas non génomiques, les males utilisés par insémination sont généralement mis à l'épreuve de la descendance avant une utilisation plus large. Pour les caractères maternels, ils sont accouplés à une partie des femelles des élevages du noyau. Les filles issues de ces accouplements et conservées par les éleveurs rentrent en production et fournissent ainsi les phénotypes nécessaires à l'évaluation de leurs pères. Seules les meilleurs pères sont conservés et utilisés comme améliorateurs. Le délai entre choix des mâles pour le testage et évaluation sur descendance dépend de la précocité sexuelle des mâles et des femelles. En ovin allaitant, certains mâles sont également testés pour leurs aptitudes bouchères. Ils sont accouplés avec des femelles d'élevage ne faisant pas partie du programme de sélection. Les agneaux issus de ces femelles sont contrôlés fournissant ainsi les phénotypes nécessaires à l'évaluation de leurs pères. Les effectifs de males testés annuellement sont très variables entre races, d'une vingtaine en race Corse à plus de 250 en race Lacaune lait pour un total de plus de 500 mâles en ovins laitiers, une trentaine en Saanen et quarantaine en Alpine

pour les caprins et compris entre une dizaine et une trentaine par race sur leurs aptitudes maternelles et/ou bouchères pour un total de plus de 150 males en ovins allaitant (Capgènes, 2017 ; Loywyck et Lagriffoul, 2017).

Utilisation de l'assignation de parenté

Les marqueurs génétiques peuvent contribuer à une meilleure connaissance des parentés. Si les parentés sont établies a priori (système déclaratif), les marqueurs peuvent les confirmer ou les infirmer, en détectant les incompatibilités mendéliennes : il s'agit de contrôles de filiation. Si les parentés sont inconnues, on peut les inférer en calculant la probabilité de pères putatifs d'être le père d'un individu compte-tenu de leurs marqueurs. On parle dans ce cas d'assignation de parenté sur marqueurs moléculaires. L'assignation de parenté suscite un vif intérêt, exprimé à la fois par les éleveurs-sélectionneurs et les gestionnaires des programmes de sélection ovin et caprin (Hamelin, 2014 ; Nicolas, 2014). La technique est porteuse d'espoir pour l'obtention des parentés en monte naturelle tout en limitant les contraintes énoncées précédemment.

En ovin, les premiers tests d'assignation de parenté ont été effectués en France avec le panel de dix microsatellites utilisé pour la réalisation des contrôles de filiation (Raoul *et al.*, 2012). Les résultats obtenus avec ce panel dans diverses situations (parents avec ou sans analyse, parenté proches entre reproducteurs putatifs) ont été décevants que ce soit au niveau du taux d'agneaux assignés ou bien du taux d'erreur d'assignation. Ainsi un nouveau panel a été développé, spécifique pour l'assignation et basé sur des marqueurs SNPs (Tortereau *et al.*,

2017). Ce choix a été renforcé par i) une automatisation plus aisée des analyses de marqueurs SNPs et ii) une possibilité d'analyses simultanées de marqueurs d'assignation et de mutations en ségrégation dans les populations ovines. A partir du génotypage sur puces SNPs moyenne densité (≈50 000 marqueurs) d'une trentaine de races ovines, 1432 SNPs ont été sélectionnés comme marqueurs candidats pour l'assignation pour l'ensemble des populations. Un panel d'assignation comprenant environ 249 marqueurs a été élaboré sur technologie Sequenom et testé en condition réelle en race Blanche du Massif Central, incluant le processus de collecte des échantillons biologiques. Le taux d'assignation était de 95% avec un taux d'erreur de 1%.

En caprin, une démarche internationale a été initiée à partir des données disponibles de l'initiative ADAPTmap. Un panel de 195 SNPs a été sélectionné à partir de 3887 individus provenant de 106 populations (Talenti et Palhière *et al*., 2017). Ce panel, testé sur un dispositif en race Alpine et Saanen, permet un taux d'assignation de 100%. Les résultats obtenus en ovin et caprin sont encourageants et permettent d'envisager une assignation à grande échelle.

Le frein majeur qui limite aujourd'hui le développement de l'assignation est son coût de revient pour l'éleveur, situé autour de 20 euros par individu assigné. Une enquête réalisée auprès d'éleveurs ovins et caprins adhérents au contrôle de performances indiquait que 75% des éleveurs étaient intéressés par l'utilisation de la technique dans leur élevage avec un consentement à payer compris entre 5 et 10 euros maximum (Hamelin, 2014 ; Nicolas, 2014). Les conditions de la mise à disposition de cette technique aux acteurs de la sélection ovine et caprine ne semblent donc pas encore réunies. Les enjeux pour les populations et leurs élevages sont néanmoins différents en fonction du taux de paternités connues des femelles et de leur(s) mode(s) de reproduction dominant : insémination, monte naturelle en paternité ou en mélange.

Lorsque les filiations paternelles sont déduites de l'insémination artificielle, l'utilisation de l'assignation n'est pas un enjeu majeur. En effet, l'assignation de paternité ne constitue pas une alternative aux autres fonctions remplies par l'insémination telles que l'accroissement de la descendance des meilleurs reproducteurs mâles ou la connexion génétique entre troupeaux.

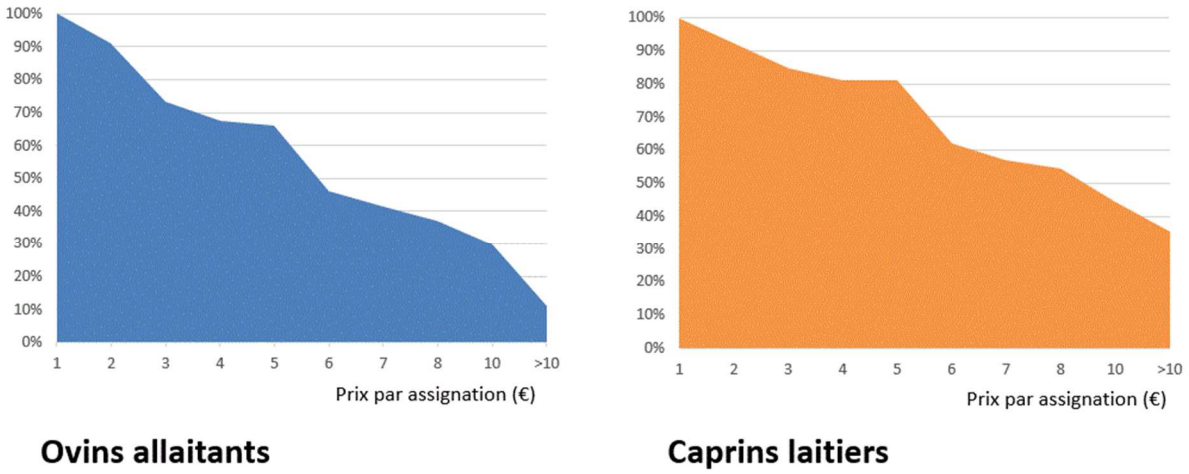


Figure 2 : % cumulé d'éleveurs en contrôle de performances désirant utiliser l'assignation de parenté en fonction de son coût (d'après Hamelin, 2014 et Nicolas, 2014).

Lorsque les filiations paternelles sont déduites d'une conduite spécifique de la reproduction en monte naturelle (monte en main, lutte en lot de paternité), l'assignation de paternité constitue une alternative et peut permettre de contourner ou réduire les contraintes afférentes aux méthodes classiques (temps de travail, main d'œuvre). L'enjeu est principalement économique : substitution du coût actuel des parentés issues des lots de monte naturelle en lutte contrôlée par le coût de l'assignation. Le choix d'adopter l'assignation sera déterminé, par l'éleveur et pour sa situation, en fonction du bilan coût-bénéfice de chacune des méthodes.

Lorsque le taux de paternité connue des femelles est faible, l'enjeu est également technique car une méthode moins contraignante pourrait se traduire par une augmentation du taux de femelles de paternité connue. Le coût lié à l'assignation est potentiellement associé à une meilleure efficacité technique du programme de sélection qu'il convient d'intégrer pour mesurer l'intérêt de la méthode.

La gestion des gènes majeurs : le cas particulier des gènes d'ovulation dans les populations ovines.

La notion de « gène majeur » résulte de la part importante de la variance génétique expliquée par un seul polymorphisme (SNPs, combinaison de SNPs ou autres types de polymorphisme) et constitue un écart au modèle génétique infinitésimal. Ce type de caractère est dit d'hérédité mixte : polygénique et gène majeur. La possibilité de fixer l'allèle favorable au gène majeur, les méthodes de gestion adaptées à ce type de situation et les conséquences sur la vitesse de fixation de l'allèle favorable et les gains génétiques à court et long termes sont bien établies (Larzul *et al.*, 1997 ; Manfredi *et al.*,1998 ; Dekkers et Van Arendonk, 1998 ; Villanueva *et al.*, 1999 ; Gomez-Raya *et al.*, 1999 ; Sánchez *et al.*, 2005 ; Costard *et al.*, 2014 ; Fontanesi *et al.*, 2015 ;.Carillier-Jacquin *et al.*, 2016).

Plusieurs gènes majeurs ont été identifiés dans les populations ovines et caprines tel que le gène de la myostatine impliqué dans la variabilité du développement musculaire en ovin (Clop *et al.*, 2006), le gène de la caséïne impliqué dans la variabilité des aptitudes fromagères

du lait en caprin (Boulanger *et al.*, 1984) ou encore le gène PrP (Hunter *et al.*, 1996) impliqué dans la variabilité de la résistance génétique à la tremblante en ovin et caprin. Une partie de ces gènes majeurs, tel que PrP en ovins (Palhière *et al*., 2004), est pris en compte dans la sélection, entrainant un accroissement de la fréquence de leur allèle favorable. En ovin, de nombreux gènes majeurs affectent le taux d'ovulation et indirectement la taille de portée (Bodin *et al.*, 2011). Pour ces gènes majeurs, la situation est souvent particulière car si l'allèle hyper-prolifique est intéressant chez les femelles hétérozygotes, son effet est souvent trop important ou délétère (problèmes de fertilité) chez les femelles homozygotes porteuses.

En France, les recherches sur les gènes majeurs d'ovulation ont débuté en 1981-1982 (très peu de temps après la découverte du gène Booroola (gène BMPR1B) en Australie) suite à un don à l'INRA par le CSIRO de cinq mâles Booroola. Depuis, de nombreux gènes majeurs présents sur le chromosome X ou sur un autosome ont été découverts, et régulièrement l'implication de nouveaux gènes ou polymorphismes de gènes connus sont mis en évidence dans de nouvelles populations ovines. En mérinos d'Arles, l'introgression du gène Booroola dans une partie de la population a été décidée en concertation avec les sélectionneurs de la race (Bodin *et al.*, 2011). A l'issue du programme d'introgression piloté par l'INRA, la diffusion a été dans un premier temps limitée à trois éleveurs privés. En 2010, un programme de production d'agnelles porteuses hétérozygotes a été initié pour élargir la diffusion. Dans la population Lacaune viande Ovi-test, une première mutation, portée sur le gène BMP15, a été mise en évidence en 2007 (Bodin *et al.*, 2007). De faible fréquence et induisant une stérilité des porteuses homozygotes, la mutation a été éradiquée. Une seconde mutation dans le gène FecL (aussi appelé gène Lacaune) a été mis en évidence au début des années 2000. Depuis 2010, toutes les femelles de renouvellement sont génotypées. Estimés en 2014, l'allèle favorable a un effet de +0.5 agneau par mise-bas sur la prolificité tandis que le poids à 30 jours est plus faible

et la mortalité des nés double légèrement plus élevée (Martin *et al.*, 2014). La gestion de cette mutation repose sur l'interdiction des accouplements susceptibles de générer des femelles porteuses homozygotes. Dans les populations Noire du Velay, Grivette, Ile de France, Mouton Vendéen, Causses du Lot et Belle-Ile, des mutations sont suspectées ou en cours de caractérisation (Bodin *et al.*, 2011 ; Bodin, 2017, communication personnelle).

Les règles de gestion classiques des reproducteurs reposent sur l'hypothèse d'une hérédité polygénique. Pour des allèles à effet fort, leur présence dans une population où la prolificité a une certaine importance économique nécessite d'ajuster ces règles de gestion. L'estimation des valeurs génétiques dont la transmission n'obéit plus aux lois de la génétique polygénique classique est biaisée. Ignorer la présence de ces allèles hyper-prolifiques conduit à une augmentation non contrôlée de leur fréquence se traduisant par un accroissement du nombre des femelles porteuses homozygotes, indésirables, et à une réduction de la variabilité polygénique de la prolificité. Cela impose donc de rechercher activement ces mutations lorsqu'elles sont suspectées et de les caractériser précisément en estimant leur fréquence dans la population et leurs effets sur la prolificité et autres performances zootechniques. En fonction de leurs fréquences et effets, ces mutations doivent être conservées en gérant de façon optimisée les fréquences des différents génotypes lorsque c'est jugé nécessaire au développement de la population ou éradiquées.

Opportunités et frein au développement de la sélection génomique dans les populations ovines et caprines

L'intérêt de l'information génomique nécessite que la précision des prédictions, corrélation entre valeurs génétiques estimées et valeurs génétiques vraies, atteigne un niveau minimum. Ce niveau de précision requis dépend des variations sur les autres composantes du progrès génétiques, intensité de sélection et intervalle de génération, induites par la réorganisation du programme de sélection permise par la valorisation d'informations génomiques.

Dans un schéma classique de petits ruminants, la mise à la reproduction des femelles et mâles de monte naturelle est précoce, les taux de renouvellement moyen (20%) à élevés (33%) et la durée du testage plutôt faible : âge des mâles compris entre 4 et 5 ans lors de l'obtention des résultats. Lorsque la monte naturelle a une part importante, voire prépondérante, la réduction de l'intervalle de génération est alors très faible. En termes de progrès génétique, les gains attendus liés à la réduction de l'intervalle de génération pour les petits ruminants (Shumbusho *et al.*, 2013) sont moindres que ceux attendus pour les bovins (Schaeffer, 2006). En outre, selon la part relative de l'insémination et de la monte naturelle, les différentielles de sélection appliquées pour la sélection sur caractères femelles des mâles est très différente : les mâles d'insémination sont présélectionnés sur ascendance puis sélectionnés sur descendance avec une précision élevée tandis que les males de monte naturelle sont sélectionnés sur ascendance avec une précision faible. Le mode de reproduction des mâles affecte également la quantité et qualité de leurs phénotypes, c'est-à-dire de l'information disponible (*e.g.* le nombre de filles) pour l'évaluation de leurs valeurs génétiques.

La sélection génomique est historiquement basée en Holstein sur une population de référence comprenant des mâles évalués sur descendance et génotypées avec une puce moyenne densité (environ 50000 SNPs) puis constamment élargie par le génotypage de nouveaux mâles (Boichard *et al.*, 2012). Les facteurs principaux affectant la précision de l'évaluation génomique sont le déséquilibre de liaison entre marqueurs et QTL, la distribution des effets des QTL, la densité en marqueurs, la précision des phénotypes, la taille de la population de référence et son degré d'apparentement avec les candidats à la sélection (Hayes *et al.*, 2009).

Ces dernières années, plusieurs techniques ont vu le jour : le développement de puces « basse densité », entre 10 000 et 15000 SNPs, principalement utilisées pour le génotypage des femelles et des candidats à la sélection et de puces « haute densité », entre 600 000 et 700 000 SNPs, pour le génotypage des pères à mâles. Associés aux méthodes d'imputation, ces deux outils ont permis des gains économiques dus au coût moindre des génotypages des candidats, et techniques liés au gain de précision des index basés sur des génotypes haute densité pour certains caractères (Solberg *et al.*, 2011). Le projet 1000 bull genomes a également accéléré le séquençage du génome de mâles stratégique mais les gains pour la sélection restent jusqu'ici limités (VanRaden *et al.*, 2017).

Les situations ovines et caprines sont généralement défavorables par rapport à la situation Holstein. Le nombre moyen de filles avec performances par mâle est plus faible à beaucoup plus faibles en raisons de contraintes technico-économiques telles que la part des populations soumise au contrôle des performances, l'insémination en semence fraîche en ovin ou la part de la monte naturelle. De plus, les tailles efficaces de population sont généralement plus élevées (Hall, 2016).

Malgré une sensibilité importante aux paramètres de la population de référence, taille et nombre effectif de segments chromosomiques (Brard et Ricard, 2014), les formules dérivant les précisions des prédictions génomiques montrent que des quantités et qualités inférieures des phénotypes des individus de la population de référence (nombre de phénotypes, degré d'apparentement des individus qui expriment ces phénotypes) se traduisent par une moindre précision.

En plus de l'effet des modes de reproduction sur la précision des prédictions génomiques, Shumbusho *et al*. (2013) montrent, à précision équivalente, que des schémas basés sur des taux d'inséminations variables présentent des gains potentiels différents liés à la mise en place de la sélection génomique.

Les facteurs limitant la précision des prédictions génomiques peuvent être compensés par l'accroissement du nombre d'individus dans la population de référence. Etant donné la taille de sa population en sélection et une utilisation quasi-exclusive de l'insémination, l'intérêt de la sélection génomique a été démontré en race ovine Lacaune et mise en place en 2015. La race ovine laitière Manech Tête Rousse et les races caprines Alpine et Saanen se préparent également à adopter la sélection génomique.

Pour les populations de plus petites tailles, le nombre d'individus dans la population de référence est limité lorsqu'elle est composée seulement de mâles. Accroître le nombre d'individus par des approches multiraciales n'a pour l'instant pas donné de gains de précision significatifs (Moghaddar *et al.*, 2014). La prise en compte des phénotypes et génotypes des mâles présents dans les élevages commerciaux est une stratégie prometteuse (Santos *et al.*, 2017). Cependant, les performances et la généalogie des femelles des élevages commerciaux sont rarement enregistrées.

Les formules déterministes dérivant la précision des index génomiques ne prennent pas en compte l'apparentement entre populations de référence et candidats (Clark *et al.*, 2014). Lorsque le degré d'apparentement entre ces deux populations est élevé, la précision des index génomiques pourrait être supérieure à celle prédite. L'obtention d'un degré d'apparentement élevé serait possible dans le cas des populations françaises : la connaissance des pères des candidats mâles à la sélection est généralement exhaustive compte-tenu de leur taille et de leur fonctionnement basé sur la gestion collective de la voie mâle.

Au-delà de l'intérêt technique, la rentabilité économique de la sélection génomique doit être étudiée avant que sa mise en place soit effective. Les gains génétiques supplémentaires induits par la sélection génomique se répercutent au niveau des élevages du programme de sélection mais aussi de la population commerciale en fonction de l'importance de la diffusion. En contrepartie, les coûts nécessaires à l'obtention de l'information génomique, prélèvements biologiques, génotypages et calculs, se concentrent essentiellement au niveau du programme de sélection. Shumbusho *et al.* (2016) montrent que si les schémas génomiques induisent généralement un coût supérieur par rapport au schéma classique, la prise en compte du surplus de recettes générées au niveau de la population commerciale se traduit par une rentabilité supérieure de certaines stratégies de sélection génomique. L'étude souligne cependant que l'optimisation des schémas classiques serait plus rentable que l'introduction d'information génomique.

Chapitre 2 – Accroître le taux de paternité des femelles en sélection par l'assignation de parenté : intérêt génétique et rentabilité économique

Résumé de l'article 1

Dans les élevages commerciaux, les femelles sont luttées en lot avec plusieurs mâles. Les pères des nouveau-nés issus de ces luttes ne peuvent être identifiés. Dans les élevages en sélection, les brebis peuvent être inséminées ou luttées par monte naturelle en paternité par exemple via des lots de luttes de taille limitée et comprenant un seul mâle ou encore par la monte en main. Cependant l'utilisation de ces techniques est parfois limitée dans les élevages en sélection, ce qui aboutit, pour certaines populations, à des proportions variables et parfois assez faibles de femelles dont le père est identifié. Dans ce contexte l'assignement des parentés sur marqueurs moléculaires, rendu possible par le développement du panel SNPs très basse densité, constitue une technique attractive pour les éleveurs.

Le progrès génétique annuel créé au sein de la population en sélection est modulé par plusieurs facteurs dont la précision des valeurs génétiques estimées (index). Pour un individu donné, la précision des index dépend des paramètres génétiques du caractère (héritabilité, répétabilité) et des performances connues pour l'individu et ses apparentés. Les manques d'information dans les généalogies limitent la quantité d'informations disponibles et donc diminuent la précision des index. Cependant, les conséquences de la connaissance partielle des filiations paternelles des femelles sur le progrès génétique annuel n'ont jamais été évaluées dans des situations où la monte naturelle est le mode de reproduction prépondérant.

Nous avons dans cet article évalué trois types de programmes de sélection à l'aide de simulations déterministes : un programme dans lequel les pères à mâles sont exclusivement utilisés par insémination après avoir été présélectionnés sur ascendance puis évalués et sélectionnés sur performances de leur descendance, un programme dans lequel les pères à mâles sont exclusivement utilisés par insémination après avoir été sélectionnés uniquement sur ascendance et un programme dans lequel les pères à mâles sont exclusivement utilisés par monte naturelle après avoir été sélectionnés sur ascendance. Pour ces programmes nous avons calculé le gain génétique annuel asymptotique d'un caractère maternel pour différentes valeurs du taux de filiation paternelle des femelles. Les reproducteurs mâles et femelles étaient regroupés en catégories définies de telle sorte que la précision des index intra-catégorie soit homogène. Le statut (mâle de monte naturelle, en testage, améliorateur) et le nombre de cycles de reproduction étaient considérés pour les catégories mâles et la parité, le statut et le nombre de cycles de reproduction du père (incluant le statut « inconnu ») pour les catégories femelles. Pour chacune des catégories, en se basant sur un modèle démographique, les informations moyennes disponibles pour évaluer les reproducteurs ont été calculées : nombre de performances propres, nombre moyen d'apparentés identifiés, nombre moyen de performances des apparentés identifiés. La précision moyenne des index de chacune des catégories a ensuite été dérivée en se basant sur la théorie des indices et les accouplements et la sélection des candidats au cours du temps ont été décrits par un modèle de flux de gènes. Afin d'évaluer la rentabilité économique de l'assignation, ce modèle de flux de gènes a été étendu à la population commerciale dans laquelle sont transférés des reproducteurs mâles ou femelles nés dans le noyau. Ce modèle a permis de suivre l'évolution des niveaux génétiques au sein de la population à partir de la mise en place de l'assignation de parenté en fonction i) du niveau initial du taux de filiations paternelles, ii) du mode de diffusion du progrès

génétique (vente de mâles ou de femelles) et ii) de la taille relative du noyau et de la population commerciale. Le profit relatif et le temps de retour sur investissement ont été calculés pour différents ratios gain/coût (gain monétaire dû à un écart-type génétique supplémentaire divisé par le coût unitaire de l'assignation).

Les gains génétiques annuels supplémentaires qui ont été obtenus dépendent de l'organisation du programme de sélection, du taux de filiations paternelles initial et des paramètres génétiques du caractère sélectionné. Pour un caractère peu héritable (0.10) sélectionné dans un schéma basé exclusivement sur la monte naturelle, le passage d'un taux de filiations paternelles de 5% à un taux de 100% a permis un gain additionnel de +16.9%. Ce gain additionnel chutait à +6% pour ce même type d'organisation quand le taux de filiations paternelles initial était élevé (70%). Pour les programmes de sélection basés sur l'insémination mais sans testage, le gain additionnel variait entre +13.5% et +5.9% pour des taux de filiations paternelles initiaux de 25 à 70%. Pour les programmes de sélection basés sur l'insémination avec testage, le gain additionnel était inférieur à +5% même pour un taux de filiations paternelles initial relativement bas (40%). Pour des caractères plus héritables, les gains additionnels des schémas monte-naturelle et IA sans testage diminuaient soit autour de +10% pour h^2=0.2 et autour de +5% pour h^2=0.4.

En termes d'intérêt économique, nos résultats indiquent que l'augmentation du taux de filiations paternelles par l'assignation de parenté n'était pas toujours rentable et que le résultat était variable en fonction du type de programmes de sélection, de la taille de la population commerciale et du mode de diffusion des reproducteurs. Au niveau du seul noyau, un ratio gain/coût de 5 (gain monétaire dû à un écart-type génétique supplémentaire divisé par le coût unitaire de l'assignation) était nécessaire pour assurer une rentabilité quel que soit le mode de diffusion et le type de programme de sélection. Compte tenu du caractère, exprimé

par les femelles, la diffusion de femelles permettait un retour sur investissement plus rapide. Cependant, ce mode de diffusion limite la capacité de diffusion du noyau vers la population commerciale. Lorsque le noyau représentait un tiers de la population totale, un ratio gain/coût de 2 était suffisant pour que l'assignation soit rentable quel que soit le mode de diffusion et le type de programme. Cependant le profit obtenu était dans ce cas très faible.

Article 1 : Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs

J. Raoul, I. Palhière, J.M. Astruc and J.M. Elsen. Journal of Animal Sciences. 2016. 94: 3663-3683.

Published October 6, 2016

Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs[1]

J. Raoul,*†[2] I. Palhière,† J. M. Astruc,* and J. M. Elsen†[2]

*Institut de l'Elevage, F-31321 Castanet-Tolosan, France;
and †GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT, Castanet Tolosan, France

ABSTRACT: In sheep and goat breeding programs, the proportion of females for which the sire is known (known paternity rate [KPR]) can be very low. In this context, paternity assignment using SNP is an attractive tool. The annual genetic gain (AGG) is impacted by the accuracy of the EBV. In populations with a low KPR, the number of known relatives for a given individual is low, and the EBV that are based on this information are imprecise. However, the impact of partially known paternal filiations, in terms of potential genetic and economic losses, has never been quantitatively evaluated in situations where natural mating is the main reproductive mode. A deterministic model was developed to assess, for a panel of real breeding programs, the influence of the female KPR on the AGG and economic benefit. First, males were divided into categories according to their status (natural mating or AI sire) and breeding cycle and females according to parity, sire status (including unknown sire), and breeding cycle of the sire. Second, a demographic model described, for each category, the accumulation of known records for individuals and their close relatives. The output from this model was used to compute the average accuracy of the EBV per category. Then, a genetic model based on the gene flow between categories over time was described. Using the average accuracy of EBV per category, it provided the asymptotic AGG of the nucleus given its KPR. In the economic studies, changes to the mean genetic values in the nucleus and the commercial population after an increase in KPR and various gain:cost ratios (monetary gain due to an extra genetic SD of the selected trait divided by the cost of 1 assignment) were considered. Relative profit and payback periods were computed. We showed that SNP-based parentage assignment aimed at increasing the female KPR was not always profitable and that the type of breeding program and the size of the commercial population should be taken into consideration. Notably, achieving a profit was largely dependent on obtaining a favorable gain:cost ratio. The maximum supplementary AGG (16.9%) was obtained for breeding programs using only natural mating. In such programs without AI, a gain:cost ratio of 5 was needed to make assignment profitable at the nucleus level whereas a gain:cost ratio of 2 was sufficient if the nucleus represented a third of the total population.

Key words: breeding programs, deterministic model, economic, genetic gain, parentage, small ruminants

 J. Anim. Sci. 2016.94:3663–3683
doi:10.2527/jas2015-0165

INTRODUCTION

The availability of SNP chips has opened new opportunities such as the implementation of genomic selection (Meuwissen et al., 2013). The benefits of genomic selection have been assessed in goats and sheep (Pickering et al., 2013; Shumbusho et al., 2013, 2016; Van der Werf, 2014; Rupp et al., 2016) and the genomic predictions have been studied and proposed for use by goat and sheep breeders in a few countries (Daetwyler et al., 2013; Moghaddar et al., 2013; Auvray et al., 2014; Baloche et al., 2014; Carillier et al., 2014; Swan et al., 2014). Moreover, the availability of SNP chips has allowed panels for accurate parentage testing to be designed (Bell et al., 2013; Clarke et al., 2014; Heaton et al., 2014; Tortereau et al., 2015).

[1]This study was funded by Institut de l'Elevage, INRA, and Apis-gène. We gratefully acknowledge the funding.

[2]Corresponding authors: jerome.raoul@toulouse.inra.fr; Jean-Michel.elsen@toulouse.inra.fr

Received December 4, 2015.
Accepted July 11, 2016.

In goat and sheep nuclei, known pedigrees are incomplete because only some of females are inseminated by or mated to a single male. In populations with a low female known paternity rate (**KPR**), the number of known relatives of a given individual is low, leading to less accurate EBV. Up to now, few studies have investigated the influence of KPR. Stochastic simulations showed that genealogical information reduces bias and increases the accuracy of EBV (Schenkel and Schaeffer, 2000; Schenkel et al., 2002), in particular in goats (Analla et al., 1995), sheep (Besbes, 1990), and cattle (Harder et al., 2005).

However, the impact of KPR on both the genetic gain and profit has never been quantitatively evaluated in breeding programs combining AI and natural mating (**NM**). Such assessment would provide useful information as to whether using SNP for paternity identification is profitable. Using a detailed model, our objective was first to investigate the influence of female KPR on the annual genetic gain (**AGG**) in 3 different breeding programs. Second, we assessed the potential economic benefit of using SNP to increase the female KPR and then applied our model to various case studies.

MATERIAL AND METHODS

In most small ruminant breeding programs, sires of breeding males are generally known (most potential sire dams are inseminated by or mated with a single male). Paternity of breeding females is generally known when they were born from AI sires but only sometimes when they were born from NM sires. We modeled situations where sires of females born from NM were assigned using DNA-based methods. Various breeding managements of selected males were considered: NM and AI after or without progeny testing. Sire breeding management might have some consequences on expected additional gain due to an increase in female KPR and on profitability of using DNA-based assignment. In situations studied in this paper, increasing female KPR does not allow better estimation of breeding values for traits recorded on only males (e.g., meat trait in French sheep breeding plans). Situations where traits are recorded on lambs or kids (males and females) were not studied because they require too many DNA tests. Therefore, only a maternal trait has been considered.

General Overview

Our deterministic model, developed in the FORTRAN language, comprised several steps. First, a demographic model was developed: males were divided into categories according to their status (i.e., NM sires or AI sires) and breeding cycle (i.e., first, second, or third reproduction cycles during which they were used) and females according to their parity (subsequently designated "rank"), sire status (including unknown sire), and the breeding cycle of their sire. The number of individuals belonging to each category was driven by a set of input parameters, including the female KPR. For each of these categories, the demographic model described the accumulation of records for the individuals themselves and their close relatives over time. We assumed that the NM sires of females, when they were known, were obtained using DNA-based paternity assignment. We did not consider any sire misidentification. In a second phase, the output from this demographic model was used to compute the average accuracy of EBV per category. It is essential to note that due to the female categories we defined and the hypotheses we made, the female KPR had an effect on the number of females belonging to each category. Except for female categories for which the sire status was NM sire, the female KPR did not have any effect on the average accuracy of the EBV for a given category. This will be shown later. In a third step, a genetic model based on gene flow between categories over time was developed using standard matrix methodology (Elsen and Mocquot, 1974; Hill, 1974). The contribution of each sire category to the selected newborn males and females (newborns chosen for replacement of the population) as well as the selection differentials along the sire path, depend on the breeding program assessed. The contribution of each dam category to the selected newborn males was computed assuming a selection above a single truncation threshold: whatever the category, candidate females displaying an EBV above this threshold were retained as sire dams for the next period of time. On the whole, the female KPR influenced both the average accuracy of the EBV for the entire female population (in the nucleus) and the genetic contribution to selected newborn males of each female category.

We used this model to compare the asymptotic annual genetic gain (**aAGG**) given different levels of female KPR. We also estimated the economic value of increasing the initial value of the female KPR (**init.KPR**) to 1 by measuring, for females of the nucleus, the short-term evolution of the mean genetic values in the nucleus and the commercial population after an increase of this rate. Then, we computed the cumulated discounted revenues provided by the additional genetic gain achieved and the cumulated discounted costs due to assignment of newborns. Several economic values for the trait, assignment cost, and commercial population sizes were considered.

Stratification of the Population and Breeding Programs Assessed

For the 3 breeding programs modeled, the selection criterion was a maternal trait, with repeated measures

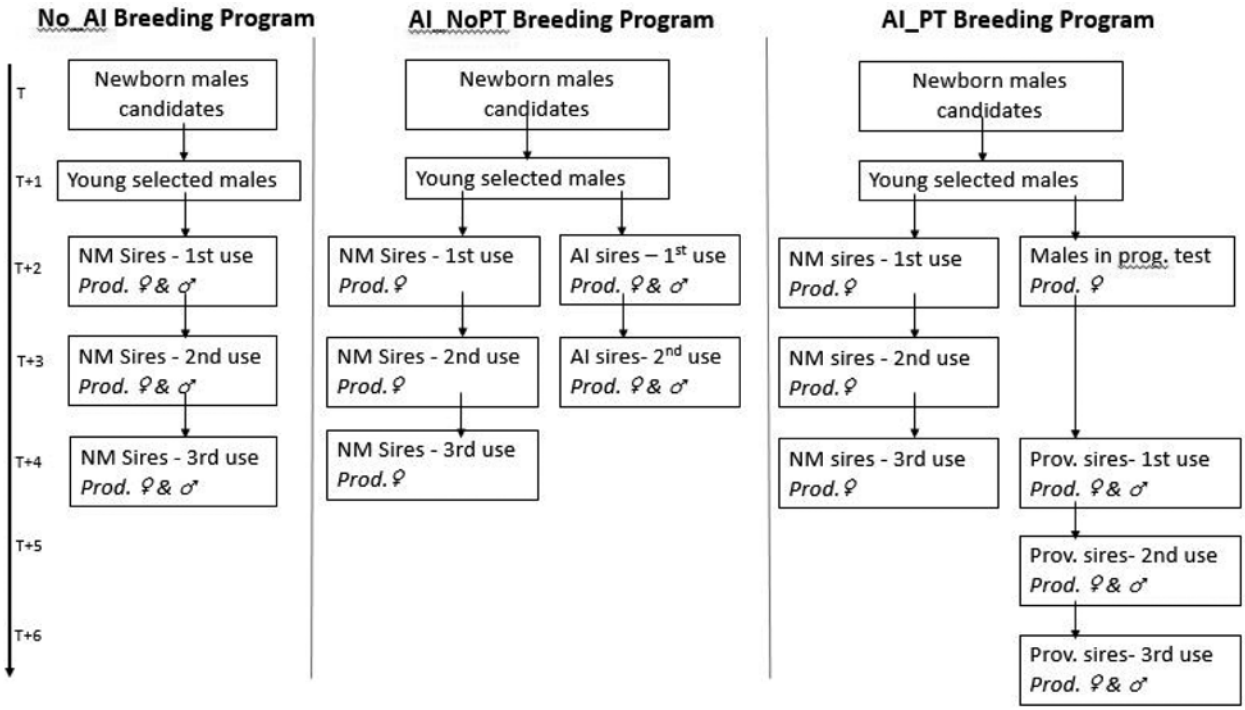


Figure 1. Description of male categories. AI_PT = male candidates were born from progeny-tested AI sires; AI_NoPT = male candidates were born from non-progeny-tested AI sires. No_AI = male candidates were born from natural mating (NM) sires; Males in prog. test = males being progeny tested (by AI); Prov. sires = proven AI sires; *T* = Time expressed in reproduction cycles; Prod.♀ = sire categories contributing to newborn female candidates; Prod. ♀ & ♂ = sire categories contributing to newborn female and male candidates.

on only females. Reproduction was based on both NM and AI. These models reflected 3 types of representative programs used in French small ruminant populations. In the "No AI" (**No_AI**) breeding program, only NM sires were used. In the "AI No Progeny Testing" (**AI_NoPT**) breeding program, all candidate males were born from AI sires that were not progeny tested. As the use of AI is limited, not all selected females were born from AI sires. All available newborn females born from an AI sire were kept for replacement in the nucleus (25% of selected newborn females). The other selected females (75%) were born from NM sires. In the "AI Progeny Testing" (**AI_PT**) breeding program, all male candidates were born from proven AI sires but only some female candidates was born from AI sires. All available females born from both AI sires being progeny tested (approximately 9% of selected newborn females) and proven sires (approximately 31% of selected newborn females) were kept for the nucleus. The other selected females (60% of selected newborn females) were born from NM sires. We assumed that the sires of selected young males were known, whatever the program. Given the parameter used in the model, 5% of fertile dams mated with a single male were sufficient to produce young males needed for replacement. As there was no selection along the dam–female path, this assumption implied a minimum female KPR of 0.05 (sisters of male candidates). We also assumed that the sires of females born from AI were known. Based on these assumptions, minimum values of the female KPR were obtained for the different breeding programs (0.05 for No_AI, 0.25 for AI_NoPT, and 0.40 for AI_PT).

Figure 1 describes the 11 categories of males, their selection process, and their use in each breeding program. The first category assembles the newborn male candidates selected on their dams' EBV. In all programs, only males whose sire is known and born from the best 20% dams are considered as male candidates. There was no additional selection to sort males for NM or AI when they were not progeny tested or when results of the progeny test were still unknown. When progeny testing was performed (AI_PT Breeding Program), proven males were a selected subset of progeny-tested males. The number of males in each category depended on input parameters such as the number of AI males or the number of progeny-tested males by reproduction cycle.

Table 1 describes the 70 possible female categories for the various breeding programs. Female categories were determined by rank (i.e., parity, from 1 to a maximum of 7), sire status (unknown sire, sire being progeny tested, proven sire, AI sire [not progeny tested], and NM sire), and breeding cycle of the sire (first, second, third, or fourth use of the sire). The amount of information available for a given female and her close relatives depended on these 3 factors. A NM sire could be used for 3 reproduction cycles whatever the scheme

Table 1. Description of female categories according to their sire status

Sire status	Maximum number of breeding cycles per sire[1]	Maximum parity	Number of categories	Programs in which the category is implemented[2]
Unknown		7	7	No_AI, AI_NoPT, and AI_PT
Natural mating	3	7	21	No_AI, AI_NoPT, and AI_PT
AI sires (not progeny tested)	2	7	14	AI_NoPT
AI sires being progeny tested	1	7	7	AI_PT
Proven sires	3	7	21	AI_PT

[1]The maximum number of breeding cycles reflects the maximum reproductive life of males, for a given status, expressed in reproduction cycles.

[2]AI_PT = male candidates were born from progeny-tested AI sires; AI_NoPT = male candidates were born from non-progeny-tested AI sires; No_AI = male candidates were born from natural mating sires.

considered. Because AI sires were not progeny tested in the AI_NoPT breeding program, they were only used twice to limit their individual contribution. In the AI_PT breeding program, sires being progeny tested were bred once and proven sires could be used for 3 reproduction cycles. The selection of dams of future females was neglected, assuming that all fertile females contributed to selected newborn females. The selection of dams of future candidate males is fully described in the genetic model section: their EBV must be higher than a given year-specific threshold. In a given year, females' EBV distributions vary between categories. As detailed below, the number of females in each category depended on input parameters such as the female KPR, the number of AI males, and the percentage of AI. To model the effect of selection steps achieved on males (on their dams' EBV and after progeny testing), we computed the average accuracy of EBV for each female category and for progeny-tested candidate males. Therefore, we first developed a demographic model to compute the amount of information (number of records) available for the individuals of each female category and their close relatives and for males in the progeny testing category.

Parameters of the Model

All the parameters and variables included in the model are listed in Table 2. Following the methodology described by Elsen (1992), we distinguished state variables that characterize states of the population (in practical terms, states of categories), parameters (genetic, demographic, zootechnical, and economic parameters), decisional variables that describe the breeding program components, and internal variables that link parameters and decisional variables to state variables.

Demographic Model

We used a stationary demographic model in which the number of individuals belonging to a rank class was constant over cycles of reproduction (Khang, 1983). The female rank was parity for both meat sheep and dairy sheep and goats. It was assumed that unfertile females were culled at each reproduction cycle. As a consequence, each rank class comprised only 1 age class. Because of the assumption of stationarity, the proportion of females belonging to the lth rank class, α_l, was constant over time and equal to

$$\alpha_l = s_{\mathrm{f}}^{l} \times f^{l} / \sum_{i=1}^{a} s_{\mathrm{f}}^{i} \times f^{i},$$

in which a is the maximum number of rank classes, s_{f}, is the survival rate of adult females, and f is the global fertility.

The proportion of categories within a rank class depended on various parameters listed in Table 2: KPR of females, population size, and proportion of females born from an AI sire ($PropF_{AI}$).

The equations below indicate the formulae used to compute the number of females (ei,t) belonging to each category (i) within a given rank (l) depending on the status and breeding cycle y of the sire.

The proportion of sires of breeding cycle y given the maximum number of breeding cycles (b), and the survival rate of males (s_{m}) between 2 reproduction cycles was

$$\lambda(y) = s_{\mathrm{m}}^{y} / \sum_{y=1}^{b} s_{\mathrm{m}}^{y}.$$

The number of females within a parity l was then equal to $\mathrm{F}*(1 - KPR)*\alpha_l$ for females born from unknown sires and $\lambda(\mathrm{y})*[F*(KPR - PropF_{AI})]*\alpha_l$ for females born from NM sires (breeding cycle y) with F the total number of females, α_{l} the proportion of female in parity l and $PropF_{AI}$ the proportion of selected young females born from AI sires. In the AI_NoPT breeding program, the number of females within a parity l born from non-progeny-tested AI sires (breeding cycle y) was equal to $\lambda(\mathrm{y})*[F*PropF_{AI}]*\alpha_l$. In the AI_PT breeding program, the number females within a parity l was equal to $F*PropF_{AI}*PropF_{AIPT}*\alpha_l$ for those born from males being progeny tested, and $\lambda(\mathrm{y})*[\mathrm{F}*PropF_{AI}*(1-PropF_{AIPT})]*\alpha_l$ for those born from proven sires (breeding cycle y) with $PropF_{AIPT}$ the relative proportion of females born from AI males being progeny-tested.

Table 2. List of parameters and variables included in the model[1]

A. Variables				
Name	A1. State variables			
$\mu_{(t)}$	average genetic value of the female population at time t			
$g_{i(t)}$	average genetic value of category i at time t			
$d_{i(t)}$	selection differential of category i at time t			
aAGG	asymptotic annual genetic gain			
iaAGG	initial asymptotic annual genetic gain (= *aAGG*, given an initial value of the female known paternity rate)			
SGG(*t*, *n*)	supplementary genetic gain in the nucleus between t-1 and t			
SGG(*t*, *c*)	supplementary genetic gain in the commercial population between t-1 and t			
$e_{i,t}$	number of females belonging to the category i at time t			
relative profit	relative economic benefit			
Name	A.2 Decisional variables	N_AI	AI_NoPT	AI_PT
KPR	known paternity rate of females		several assessed	
init.KPR	Initial known paternity rate of females (before shifting to KPR = 1)	0.05	0.25	0.40
M_{AI}	number of AI sires selected per reproduction cycle (not progeny tested)	0	20	0
M_{PT}	number of progeny-tested sires per year	0	0	25
ND_{PT}	number of daughters per progeny-tested sire	0	0	20
$p_{(1,j)}2$	Genetic contribution of male category *j* at (*t* − *1*) to male candidates at time *t*[3]			
	Natural Mating-first (*j* = 3)	0.20		
	Natural Mating- second use (*j* = 4)	0.15		
	Natural Mating-third use (*j* = 5)	0.15		
	Progeny-tested sire (AI; *j* = 6)			0.00
	Proven sire (AI) – first use (*j* = 8)			0.20
	Proven sire (AI) – second use (*j* = 9)			0.15
	Proven sire (AI) – third use (*j* = 10)			0.15
	AI sire – first use (*j* = 11)		0.30	
	AI sire – second use (*j* = 12)		0.20	
r_{DM}	proportion of females selected as dams of sires		0.20	
tst	selection intensity after progeny-testing			0.50
Name	A3. Internal variables			
α_l	proportion of females of parity l			
β_{AI}	proportion of females mated to an AI sire			
β_{MN}	proportion of females mated to a natural mating sire			
λ_i	proportion of sires in their ith breeding cycle			
$\sigma_{i(t)}$	accuracy of EBV of the ith category at time *t*			
$\omega_{i,k(t)}$	probability that a female from category *i* be a dam of a male given the threshold at time *t*			
$\chi_{j(t)}$	proportion of dams of males belonging to category *j* among the dams of males at time *t*			
C	additional costs due to the increase in KPR			
Das	number of fertile dams mated per sire in the assigned sub-flock			
K(*t*)	single truncation threshold to select dams of males at time *t*			
$K_{pt}(t)$	truncation threshold to select proven AI males among males being progeny tested at time *t*			
$\bar{m}_x(i)$	Average number of records (of individual or close relatives x) per animal of the ith category (maternal grand-dams [x=mgd], paternal grand-dams [x=pgd], dams, [x=d], individual [x=p], paternal half-sisters [x=phs], maternal half-sisters [x=mhs], daughters [x=da])			
$\bar{n}_x(i)$	Average number of known relatives x per animal of the ith category (paternal half-sisters [x=phs], maternal half-sisters [x=mhs], daughters [x=da])			
ND_{AI}	number of daughters selected per AI sire at each reproduction cycle			
ND_{NM}	number of daughters selected per natural mating sire and reproduction cycle			
ND_{PV}	number of daughters selected per proven sire at each reproduction cycle			
$p_{i,j}^2$	genetic contribution of category *j* at (*t* − *1*) to the category *i* at time *t*			
$PropF_{AIPT}$	relative proportion of AI females born from males being progeny-tested			
$q_{i,K(t)}$	proportion of dams of males belonging to category among the female population at time			
R	additional revenues due to the increase in KPR			
tdf	proportion of commercial newborn females born in the nucleus			
tdm	proportion of commercial newborn males born from natural mating or AI sires in the nucleus			
tdm_{AI}	proportion of commercial newborn males born from AI sires in the nucleus			
tdm_{NM}	proportion of commercial newborn males born from natural mating sires in the nucleus			

Continued

Table 2. (cont.)

B. Parameters				
Name	Genetic parameters	N_AI	AI_NoPT	AI_PT
h^2	Heritability		several assessed	
σ^2_g	genetic variance		1	
r	Repeatability		several assessed	
σ^2_p	phenotypic variance		several assessed (function of h^2)	
Name	Demographic and zootechnical parameters	N_AI	AI_NoPT	AI_PT
a	maximum parity		7	
b	maximal length of reproductive life of males		3	
Das_min	minimum number of fertile dams mated per sire in the assigned sub-flock		15	
Das_max	maximum number of fertile dams mated per sire in the assigned sub-flock		36	
dum_c	length of reproductive life of males (expressed in number of reproduction cycles) in the commercial population		3	
F	size of the fertile population		24000	
f	global fertility (natural or induced estrus+ returns)		0.90	
f_{AI}	fertility on induced estrus/AI		0.55	
$PropF_{AI}$	proportion of selected young females born from AI sires (nucleus level)	0.00	0.25	0.40
$prolif_{NM}$	prolificacy on natural estrus		1.50	
$prolif_{AI}$	prolificacy on induced estrus/AI		1.65	1.65
ps	relative size of the commercial population compared with the nucleus population		several assessed	
r_{Farn}	rate of newborn females available for replacement at the nucleus level (1- proportion culled due to functional faults and/or breed standard).		0.67	
r_{Farc}	rate of newborn females born in the nucleus and available for transfer to commercial flocks. (1- proportion culled due to functional faults)		0.72	
r_{Marc}	ratio of newborn males available for transfer to commercial flocks divided by newborn males.		1:9	
$r_{F/M}$	ratio between the number of females and the number of males in the commercial population		30:1	
s_f	survival rate of adult females		0.90	
s_m	survival rate of adult males		0.90	
s_{nb}	survival rate of newborn animals		0.85	
sex ratio	Proportion of females at birth		0.50	
Name	Economic parameters	N_AI	AI_NoPT	AI_PT
C_p	cost of sire assignment per animal		several assessed	
d	discounting rate		0.05	
MG	monetary gain per genetic standard deviation of the selected trait expressed		several assessed	
ratio			several assessed	
T	investment period (expressed in reproduction cycles)		60	

[1]AI_PT: male candidates were born from progeny-tested AI sires. AI_NoPT: male candidates were born from non-progeny-tested AI sires. No_AI: male candidates were born from natural mating sires. Note that Parameters and decisional variables (including paternal origin of males) are input data whereas internal variables and state variables depend on input data.

[2]Note that p_{ij} is defined as an internal variable except for $i = 1$ and $j = 1,12$ (paternal contributions are considered as decisional variables)

[3]Only categories that might have a genetic contribution to males and females different from 0 are represented (9 categories among the 12 male categories)

The demographic model provided the average number of available records (records for a given individual and its close relatives) for each female and male category. A complete description of the formulae used to compute all variables is available in Supplementary Material S1 (see the online version of the article at http://journalofanimalscience.org). We took into account the information collected for relatives available during the reproductive life of target individuals and for which the kinship coefficient was greater than or equal to 0.125 (assuming an inbreeding coefficient of 0 for all individuals). Such information included the number of the individual's own records ($\bar{m}_p(i)$), the average number of dam records ($\bar{m}_d(i)$), the average number of maternal granddam ($\bar{m}_{mgd}(i)$) and paternal granddam ($\bar{m}_{pgd}(i)$) records, the average number of maternal half sisters ($\bar{n}_{mhs}(i)$), their average number of records ($\bar{m}_{mhs}(i)$), and the average number of paternal half sisters ($\bar{n}_{phs}(i)$) and their average number of records ($\bar{m}_{phs}(i)$). We neglected full-sib information as their numbers are limited in breeding programs that do not use specific reproductive methods (e.g., Multiple Ovulation and Embryo Transfer). This assumption was made after comparison of average accuracies obtained by the model and real data (see, hereafter, the model fit in RESULTS AND DISCUSSION). For male categories, only the information about the males belonging to the category "males being progeny tested" (AI_PT

program) was computed. In addition to the variables obtained for females, the average number of daughters ($\bar{n}_{da}(i)$) and their average number of daughter records ($\bar{m}_{da}(i)$) were computed. In the No_AI and AI_NoPT breeding programs, sires were not selected on their progeny records. Therefore, computing their EBV was not necessary.

To compute this information, transversal and longitudinal approaches were combined into the demographic model. A transversal approach studies the population present at a given time. A longitudinal approach studies the history of a given category over time. For example, the proportion α_3 of females of rank 3 present at time t comprises females from the rank 2 class at time $t - 1$ that were present in proportion α_2 and from the rank 1 class at time $t - 2$ present in proportion α_1. By definition, the number of records known for a rank 3 female present at time t is 3. Combining longitudinal and transversal approaches, the average number of records for females present at time t was $\sum_{l=1}^{a} \alpha_l \times (l)$ (note that $\sum_{l=1}^{a} \alpha_l = 1$).

The following assumptions were used to determine the values associated with each category for all of the variables listed above:

As there was no selection along the dam–female path, the probability that a dam, whatever her category, produced a selected newborn female was equal to the proportion of rank 1 females (α_1).

The probability that a dam belonging to category i at time t produced a selected newborn male was equal to the contribution of ith category to male candidates: $\chi_i(t)$. Details as to how the contributions $\chi_i(t)$ were computed are provided hereafter in the description of the genetic model.

The number of known daughters by sire and by reproduction cycle (also used to compute the number of paternal half sisters of females) depends on sire status.

For NM sires, we assumed that daughters were assigned using markers and we did not consider any sire misidentification. Using a limited number of markers caused ambiguous parentage assignment (Dodds et al., 2005; Van Eenennaam et al., 2007) but larger SNP panels allow for accurate parentage testing (Bell et al., 2013; Clarke et al., 2014; Heaton et al., 2014; Tortereau et al., 2015). As only some newborn females could be assigned, we divided the female population mated with NM sires into 2 subflocks: the assigned subflock, containing females whose progeny is assigned, and the unassigned subflock, containing females whose progeny is not assigned. The number of daughters selected per NM sire and reproduction cycle (ND_{NM}) depends on the number of fertile dams mated per sire in the assigned subflock (*Das*) and the number of selected and fertile newborn females born from such dams. This number depends on various demographic and zootechnical parameters (prolificacy on natural estrus [*prolifNM*], the survival rate of newborn animals [*s*nb], the global fertility [*f*], sex ratio, the rate of newborn female available for replacement at the nucleus level [*rFarn*], and the survival rate of adult female [*s*f]):

$$ND_{NM} = Das \times prolif_{NM} \times s_{nb} \times f \times sex\ ratio \times r_{Farn} \times s_f.$$

For AI sires that were not progeny tested (the AI_NoPT program), the number of daughters selected per AI sire at each reproduction cycle (ND_{AI}) depends on the proportion of selected newborn females born from an AI sire ($F \times PropF_{AI} \times \alpha_1$) and the total number of AI sires ($M_{AI}(1 + s_m)$), in which M_{AI} is the number of AI sires selected per reproduction cycle and sm the survival rate of adult males:

$$ND_{AI} = (F \times PropF_{AI} \times \alpha_1)/[M_{AI}(1 + s_m)].$$

For the AI sires being progeny tested (the AI_PT program), the number of daughters per progeny-tested sire (ND_{PT}) is an input parameter.

For proven AI sires (the AI_PT program), the number of daughters selected per proven sire at each reproduction cycle (ND_{PV}) depends on 5 input parameters: $PropF_{AI}$, the number of progeny-tested sires per year (M_{PT}), NDPT, the proportion of males selected after progeny testing (*tst*), and (s_m):

$$ND_{PV} = F \times PropF_{AI} \times (1 - PropF_{AIPT}) \times \alpha_1/[M_{PT} \times tst \times (s_m + s_m^2 + s_m^3)].$$

Computation of the Average Reliability of EBV and Model Fit

The output from the demographic model was used to compute the average reliability of EBV ($r^2_{\hat{g}\hat{g}}$) per category. The general formula used was

$$r^2_{\hat{g}\hat{g}} = \left[\text{cov}(g, \mathbf{y}) \times \text{var}(\mathbf{y})^{-1} \times \text{cov}(\mathbf{y}, g)\right] / \text{var}(g),$$

in which g is the candidate genetic value and $\mathbf{y}$ is the vector of information known for an individual and its close relatives. Details are given in Supplementary Material S2 (see the online version of the article at http://journalofanimalscience.org).

Considering females born from a NM sire, variation in the female KPR affected the information provided by their paternal half sisters because the number of known daughters by NM sire varies. Considering females born from an unknown sire or an AI sire (being progeny tested,

proven, or not progeny tested), variation in the female KPR did not affect the information provided by their paternal half sisters because the numbers of known daughters per sire was completely independent from the female KPR. Likewise, variation in the female KPR had no effect, within a given category, on the information provided by paternal and maternal granddams, dams, and maternal half sisters. Variation in the female KPR had an effect on the information provided by aunts, but given the adjustment of our model (as shown later), we did not include any information from aunts in our model. A change in KPR is relevant only for the average accuracy of EBV for categories of females born from natural mating sires. However, whatever the female category, the relative weight of each female category (and therefore the weighted average accuracy of the total female population) was affected by the female KPR.

To decide which relatives to take into account in our model (minimum coefficient of kinship with the target individual) and check the quality of the EBV accuracy estimates, we compared the output from our demographic model with real data from official French genetic evaluations. The traits considered were prolificacy for meat sheep and protein yield for dairy goats. Only females with at least 1 record during the year prior to the genetic evaluation computing were retained. For each female, we extracted the EBV accuracy, rank (parity or lactation rank), sire status (as defined in our model), and breeding cycle of the sire. Given this information, we assigned individuals to categories. Depending on the breed, we removed approximately 10% of active females of rank > 7 or born from very old sires (frozen semen, for example). Then, the real average accuracy of EBV [**R_ear**] was computed for each category and compared with the accuracies computed by the model [**M_ear**], using appropriate input parameters for each breed. In a final step, adjusted heritabilities that minimized the mean square error between R_ear and M_ear were computed and compared with heritabilities used as the input parameter in our study.

Genetic Model

For a given reproduction cycle, each category is characterized by an average genetic merit value. The average genetic value of a category i at reproduction cycle t depends on the average genetic values and relative contributions of categories at cycle $t - 1$ and the selection differential applied to category i as described previously (Elsen and Mocquot, 1974; Hill, 1974; Elsen, 1993):

$$g_{i(t)} = \sum_j p_{i,j(t)} \times g_{j(t-1)} + d_{i(t)},$$

in which $g_{i(t)}$ is the average genetic value of category i at cycle t, $p_{i,j(t)}$ is the genetic contribution of category j at $t - 1$ to the category i at t, and $d_{i(t)}$ is the selection differential applied to category i at t.

This can be written using a standard matrix notation:

$$\mathbf{g_t} = \mathbf{P} \times \mathbf{g_{t-1}} + \mathbf{d_t},$$

in which **P** is the transition matrix describing the gene flow between categories after 1 reproduction cycle.

To describe the gene flow from the nucleus to the commercial population, we extended the $\mathbf{g_t}$ vector, the **P** matrix, and the $\mathbf{d_t}$ vector to include commercial population categories. The way in which the genetic progress is transferred (sales to the commercial flocks of either male reproducers or female reproducers born in the nucleus) and the relative size of the commercial population compared with the nucleus population (*ps*; from 0.5 to 50) both influence the improvement lag between the nucleus and the commercial population. We assumed that all the reproducers born within the nucleus and available for distribution were provided to commercial flocks. This situation reflected the potential economic benefit of increasing the female KPR under the most favorable circumstances. The proportion of commercial newborn females born in the nucleus (*tdf*) and the proportion of commercial newborn males born from natural mating or AI sires in the nucleus (*tdm*) and intended for commercial distribution was the ratio between the maximum distribution potential and the number of selected newborns required. When females were used to transfer the genetic progress, we set a maximum relative size for the commercial population (*ps* = 5). Beyond this size, the proportion of selected newborns (<0.1) produced by the nucleus was considered too low and unrealistic.

Details of the **P** matrix, including the computation of tdm and tdf, are provided in Supplementary Material S3 (see the online version of the article at http://journalofanimalscience.org).

Dam–Male Path. The contributions of each dam category to selected newborn males were computed assuming selection on their EBV ($\hat{g}_{i(t)}$) at time t above a single truncation threshold to select dams of males at time t ($K(t)$). to select a proportion of females selected as dams of sires r_{DM}.

Assuming a normal distribution of EBV in each category, the probability for a female belonging to category i to be the dam of a male was

$$\text{prob}\left(\hat{g}_{i(t)} > K(t)\right) = \omega_{i,K(t)} =$$

$$\int_{K(t)}^{+\infty} \left\{1 / \left[\left(2\pi\right)^{1/2} \times \sigma_{i(t)}\right]\right\} \times e^{-0.5 \times \left(\left\{u - \left[G_{i(t)} - \mu_{(t)}\right]\right\} / \sigma_{i(t)}\right)^2} du,$$

in which $g_{i(t)}$ and $\sigma_{i(t)}$ are the mean EBV and the accuracy of the EBV for females of category i at t, respectively, and $\mu_{(t)}$ is the mean EBV of the female population at t.

The threshold $K(t)$ was calculated using the iterative method developed by Ducrocq and Quaas (1988) with a precision of 10^{-6} as $\left|r_{DM} - \sum_i q_{i,K(t)}\right| \leq 10^{-6}$, where $q_{i,K(t)}$ denotes the proportion of dams of males belonging to category i among the female population at time t. $\sum_i q_{i,K(t)} = (1/F)\sum_i e_{i,t} \times \omega_{i.K(t)}$, in which F is the total population size and $e_{i,t}$ is the number of females belonging to category i at time. $\omega_{i,K(t)}$ was numerically computed with the library routine G01EAF (The NAG Fortran Library, The Numerical Algorithms Group (NAG), Oxford, United Kingdom. www.nag.com) giving the 1-tail probability for the standard normal distribution, and $\chi_{i(t)}$, the proportion of dams of males belonging to category i among the dams of males at time t, is given by $\chi_{i(t)} = q_{i,K(t)}/r_{DM}$.

Knowing the single truncation threshold $K(t)$ and assuming a normal distribution of EBV, the newborn mean genetic superiority of selected newborn males due to their dam's selection is given by

$$d_{1(t)} = 0.5 \times \sum_i \chi_{i(t)} \times \frac{\frac{1}{(2\pi)^{1/2}} \times e^{-0.5\left\{\frac{K(t)-\left[g_{i(t)}-\mu_{(t)}\right]}{\sigma_{i(t)}}\right\}^2}}{\omega_{i,K(t)}} \times \sigma_{i(t)},$$

where 1 denotes the category of selected newborn males and i the female categories.

Selection Differential after Progeny Testing. Using the same methodology and accuracy as for the dam–male path, the truncation threshold $K_{pt}(t)$ that satisfies the tst was given by

$$\text{tst} = \int_{K_{pt}(t)}^{\infty} \frac{1}{(2\pi)^{1/2} \times \sigma_{pt(t)}} e^{-0.5\left[u/\sigma_{pt}\right]^2} du,$$

in which $\sigma_{pt(t)}$ is the accuracy of the males' EBV after progeny testing.

Knowing the truncation threshold $K_{pt}(t)$ and assuming a normal distribution of EBV, the selection differential $d_8(t)$ of proven males was

$$d_{8(t)} = \frac{\frac{1}{(2\pi)^{1/2}} \times e^{-0.5\left[K_{pt(t)}/\sigma_{j(t)}\right]^2}}{tst} \times \sigma_{8(t)},$$

where 8 denotes the proven male category used for their first breeding cycle. Note that for $j \neq 1$ and $j \neq 8$, $d_{j(t)} = 0$.

Genetic Evolution: Asymptotic Annual Genetic Gain Estimation and Supplementary Genetic Gain over Time

The genetic evolution of the population between $t-1$ and t was given by $(1/F)\sum_i \left[e_{i,t-1} \times g_{i(t-1)} - e_{i,t} \times g_{i(t)}\right]$. First, the genetic gain achieved in the nucleus given various levels of female KPR was computed. For a fair comparison, we assumed that the population had been selected over a long period (from $t = -\infty$ to $t = 0$). Reproduction cycles were implemented until a steady state was reached at $t = 0$. The steady state vector of mean genetic values, $\mathbf{g}_{t=0}$, was such that the elements of $\mathbf{g}_{t=0} - \mathbf{g}_{t=-1}$ were all equal to the AGG, that is, the aAGG. Second, the supplementary genetic gain (**SGG**) over time due to a shift from an init.KPR to complete genealogy (KPR = 1) was computed. To get the SGG, the following scenario was considered: starting from a steady state at init.KPR and time $t = 0$, a progressive shift from the initial to the final value (KPR = 1) was implemented and only newborn selected females born from unknown sires were assigned. Because only the newborns were assigned and the maximum rank was set at 7 in the model, the female KPR at the nucleus level was equal to 1 at time $t = 0 + 7$. Given the evolution of the population over time, the SGG(t,n) in the nucleus and SGG(t,c) in the commercial population was computed as follows:

for $i \in$ nucleus female categories in parity 1 at t,

$$\text{SGG}(t,n) = \sum_t \sum_i (1/F \times \alpha_1) \times \left[e_{i,t-1} \times g_{i(t-1)} - e_{i,t} \times g_{i(t)}\right] - t \times \text{iaAGG},$$

and for $i \in$ commercial female categories in parity 1 at t,

$$\text{SGG}(t,c) = \sum_t \sum_i (1/ps \times F \times \alpha_1) \times \left[e_{i,t-1} \times G_{i(t-1)} - e_{i,t} \times G_{i(t)}\right] - t \times \text{iaAGG},$$

in which the initial aAGG (iaAGG) is the aAGG given KPR = init.KPR.

The value of init.KPR was the minimal value of KPR for each breeding program: 0.05 (No_AI), 0.25 (AI_NoPT), and 0.40 (AI_PT). As previously mentioned, SGG (t, c) was influenced by both the method of transfer from the nucleus to the commercial population (males or females) and the relative size of the commercial population (ps).

Economic Benefit Estimation

The potential economic benefit of increasing the female KPR depends on 1) the additional genetic gain created in the nucleus (SGG(t, n)) and in the commercial population (SGG(t, c)) for the considered trait, 2) ps (from 0.5 to 50), and 3) the gain:cost ratio defined as the monetary gain due to an extra genetic SD of the considered trait divided by the cost of 1 assignment (cost of sire assignment per animal).

Gain:Cost Ratios. The monetary gain by genetic SD of the studied trait depends on the trait and species considered. The cost of 1 assignment also varies

Table 3. Adjustment of the model to various real data sets from official French genetic databases

Breed[1]	Trait[2]	Program[3]	KPR[4]	Official h^2 [5]	Adjusted h^2 [5]	M_ear − R_ear[5]
BMC	Prolif.	AI_PT	0.5	0.10	0.111	−0.03
Romane	Prolif.	AI_NoPT	0.35	0.10	0.118	−0.04
Merinos A	Prolif.	No_AI	0.05	0.10	0.113	−0.04
Rava	Prolif.	No_AI	0.7	0.10	0.125	−0.07
Saanen	PY	AI_PT	0.65	0.37	0.300	0.08

[1]Four meat sheep breeds (Blanche du Massif Central [BMC], Romane, Merinos d'Arles [Merinos A], and Rava) and 1 dairy goat breed (Saanen) were considered.

[2]Two maternal traits were considered: prolificacy (Prolif.) and protein yield of milk (PY).

[3]AI_PT = male candidates were born from progeny-tested AI sires; AI_NoPT = male candidates were born from non-progeny-tested AI sires; No_AI = male candidates were born from natural mating sires.

[4]KPR = known paternity rate (for females; the percentage of females in the nucleus for whom the sire was known).

[5]h^2 is the heritability used in the official genetic evaluation (official h^2) performed in 2015. Adjusted h^2 is the heritability that minimizes the difference between the real average accuracy of EBV (R_ear) and accuracies computed by the model (M_ear).

from one country to another, for example, and new technical devices might reduce current prices. Ranges from €5 to 50 for the monetary gain and from €5 to 25 for the assignment cost were assumed giving a range of gain:cost ratios between 0.2 and 10.

Cumulated Discounted Costs and Revenues and Economic Indicators Computed. The SGG was achieved by increasing the init.KPR to a complete genealogy. The value of init.KPR was the minimal value of the female KPR for each breeding program: 0.05 (No_AI), 0.25 (AI_NoPT), and 0.40 (AI_PT).

The revenues due to this SGG were discounted assuming a discounting rate (d) of 0.05 and an investment period (T) of 60 reproductive cycles. This long period was chosen to minimize border effects (lack of genetic gain generated by costs occurring at the end of the investment period). So the revenues (R) and costs (C) can be expressed as follows:

$$R = MG \times \left\{ \begin{array}{l} \sum_{t=1}^{T} \mathrm{SGG}(t,n) \times \left[1/(1+d)\right]^t \times \\ \sum_{i=1}^{a} F \times \alpha_i \times \left[1/(1+d)\right]^t \\ + \sum_{t=1}^{T} \mathrm{SGG}(t,c) \times \left[1/(1+d)\right]^t \times \\ \sum_{i=1}^{a} F \times ps \times \alpha_i \times \left[1/(1+d)\right]^t \end{array} \right\},$$

in which MG is the monetary gain by genetic SD expressed in euros. The delay between selection and expressions (including their decrease in number over time) were taken into account by multiplying the number of genetic expressions $\sum_{i=1}^{a} F \times \alpha_i$ by the term $\sum_{i=1}^{a} \left[1/(1+d)\right]^i$, in which d is the discounting rate and a is the maximum parity. And

$$C = C_p \times \sum_{t=1}^{T} (\mathrm{KPR} - \mathrm{init.KPR}) \times \alpha_1 \times F \times \left[1/(1+d)\right]^t,$$

in which C_p is the cost of sire assignment per animal expressed in euros (considered only at the nucleus level).

To assess the economic benefit, a final step was implemented to compute 2 indicators: the relative profit (equal to $(R - C)/C$) and the payback period (the minimum investment period needed for a positive cash flow). We assumed that fixed costs were identical in each situation.

RESULTS AND DISCUSSION

Model Fit

Table 3 shows the adjustment of the model to various sets of real data extracted from official French genetic databases. The adjusted h^2, allowing the best fit from our model to real data, remained close to the h^2 used in the official genetic evaluations ("official h^2"). As not all of the available information was considered in our model, it could be expected that the reliabilities of EBV would be underestimated. To obtain similar reliabilities of EBV with less information, the h^2 has to be higher. The results obtained for meat sheep breeds are consistent in this respect. In contrast, in the Saanen goat breed, the model tended to overestimate the reliabilities. This overestimation could be due, in part, to the quality control performed on the data, which leads to removal of some records that are considered in our model prior to official genetic evaluation. The quantity of data removed is high for dairy traits (protein yield), with 10.4% of goat lactation data being removed before computing the official French genetic evaluations in 2014 (Douguet et al., 2015). Given the adjustment obtained, we concluded that our demographic model and the methodology used to compute the average accuracies of EBV per category were relevant. The official h^2 value was used to study the technical and economic value of increasing the female KPR.

Concerning the genetic model, our algebra was checked using different estimations of the expected genetic gain: the aAGG was computed both by convergence of the Markov chain (Elsen and Mocquot, 1974; Hill, 1974;

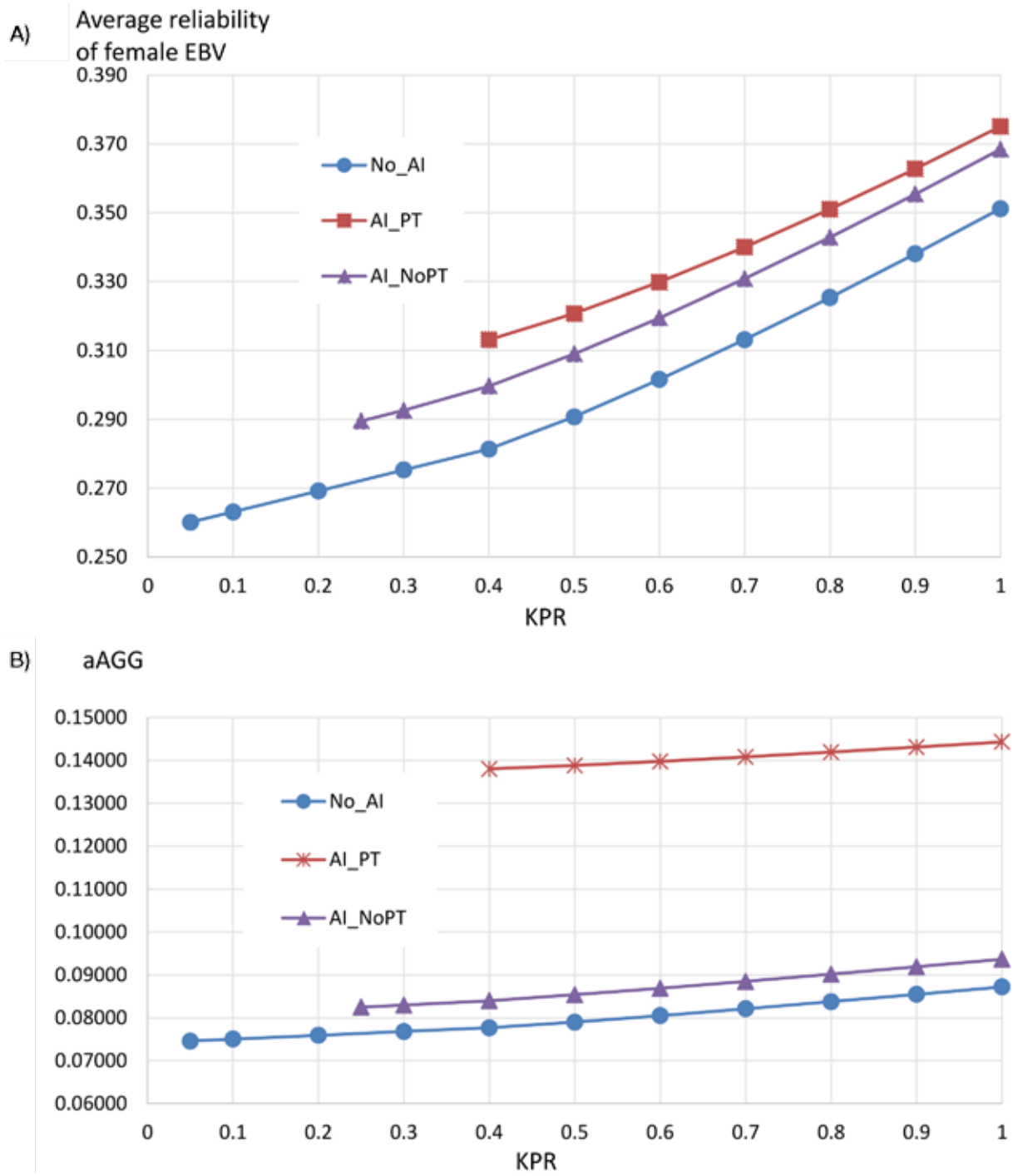


Figure 2. (a) Average reliability of female EBV as a function of the female known paternity rate (KPR) in 3 breeding programs for a low-heritability trait ($h^2 = 0.1$, $r = 0.2$). (b) asymptotic annual genetic gain (aAGG) expressed in genetic SD units as a function of the female KPR in 3 breeding programs for a low-heritability trait ($h^2 = 0.1$, $r = 0.2$). Each curve represents a breeding program. AI_PT = male candidates were born from progeny-tested AI sires; AI_NoPT = male candidates were born from non-progeny-tested AI sires; No_AI = male candidates were born from natural mating sires.

Elsen, 1993) and by adapting the Rendel and Robertson (1950) model. These methods gave the same estimates of aAGG, and aAGG was also found to be equal to the genetic evolution of the population between $t - 1$ and t as t tends to infinity. Our model did not take into account the consequences over time of the selective advantages of reproducers, as proposed in the model described by Woolliams et al. (1999). However, Bijma and Woolliams (1999, 2000) showed that the genetic gain may be accurately predicted using the conventional gene flow theory.

Supplementary Annual Genetic Gain Produced by the Nucleus

In a first step, we assessed the effect of KPR on the average reliability of females (Fig. 2a) and the aAGG

Table 4. Comparison of asymptotic annual genetic gain (aAGG)[1] and the reliability of female EBV between an initial value of the female known paternity rate (init.KPR) and complete pedigree (known paternity rate [KPR] = 1) in 3 breeding programs

	No_AI,[2] init.KPR = 0.05		AI_NoPT,[2] init.KPR = 0.25		AI_PT,[2] init.KPR = 0.40	
Value of KPR	aAGG	Av. r^2[2,3]	aAGG	Av. r^2	aAGG	Av. r^2
init.KPR	0.075	0.26	0.083	0.29	0.138	0.31
KPR = 1	0.087	0.35	0.094	0.37	0.144	0.38
Increase[4]	0.013	0.09	0.011	0.08	0.006	0.06
	16.9%	35.0%	13.5%	27.2%	4.6%	19.8%

[1]aAGG is expressed in genetic SD units.

[2]AI_PT = male candidates were born from progeny-tested AI sires; AI_NoPT = male candidates were born from non-progeny-tested AI sires. No_AI = male candidates were born from natural mating sires.

[3]Av. r^2 is the weighted reliability of females.

[4]Differences in aAGG et Av.r^2 between init.KPR and KPR=1.

(Fig. 2b) in the 3 breeding programs for a low-heritability trait (h^2 = 0.1, r = 0.2). As shown in Fig. 2a, as the female KPR increased, the average reliability of the EBV for females was improved whatever the breeding scheme considered. Similarly, the aAGG was improved in an almost linear manner (Fig. 2b). The supplementary aAGG obtained for the 3 breeding programs given their current situation (init.KPR = 0.05, 0.25, and 0.40 for No_AI, AI_NoPT, and AI_PT, respectively) is shown in Table 4. The increases in both accuracy and gain depend on the breeding scheme. The greatest increase was obtained for the No_AI breeding program, with female reliability increasing from 0.26 to 0.35 (+35.0%). This improvement resulted in a substantial supplementary AGG of 0.013 σ_g (+16.9%). The AI_PT breeding program was the least influenced by KPR, with the average reliability of female EBV increasing only from 0.31 to 0.38 (19.8%). This increase in female reliability resulted in a SGG for the AI_NoPT breeding program (+0.006 σ_g [+4.6%]) that was half that of the No_AI program. To evaluate the breeding programs on an equivalent basis, we compared the total gains obtained with the same init.KPR (Fig. 3). This corresponds to the difference of aAGG (on the y axis of Fig. 2a) between KPR = 1 and a unique init.KPR. Figure 3 shows that the gain is similar for the No_AI and AI_NoPT programs (approximately 0.010 σ_g for init.KPR = 40%) and significantly higher for the AI_PT program. However, the difference in extra gain between these programs and the AI_PT program decreased when the init.KPR was higher (approximately 0.004 for init.KPR = 40% and approximately 0.002 for init.KPR = 70%).

The greater positive effect of an increase of the female KPR in the No_AI and AI_NoPT programs compared with the AI_PT program is consistent with our hypotheses. In AI_PT programs, a large part of the AGG is attributable to sire selection after progeny testing, the EBV of proven sires being higher and more accurate than those of males used in other schemes. The dams of males selected with a single truncation threshold mainly belong to categories of females born from proven sires. The contribution of females born from NM sires categories was low. As a consequence, increasing the accuracy of the EBV of the latter categories of females had little effect on the dam–male path and, thereby, on the AGG. In contrast, in the No_AI and AI_NoPT programs, no or few dams of males were born from AI sires. Therefore, the contribution of the categories of females born from NM sires to the dam–male path was higher. The increase of the average reliability of female EBV led to an increase of the average reliability of the EBV of females selected as dams of newborn males and had, therefore, a greater effect on the AGG.

Table 5 presents the supplementary aAGG obtained in each of the breeding programs modeled with different heritability values. The higher the heritability, the lower the supplementary aAGG obtained as KPR increases to reach KPR = 1. Whatever the breeding program considered, the supplementary aAGG is halved when the heritability of the maternal trait is multiplied by 4 (from h^2 = 0.1 to h^2 = 0.4). This is not surprising because, for identical genealogical information, lower heritabilities lead to lower EBV reliabilities and the lower the initial EBV reliabilities, the higher the effect of additional information.

Comparison of the results obtained for various breeding programs and heritabilities showed that paternity assignment would mostly benefit 1) programs with an important proportion of NM and 2) programs selecting on low-heritability traits, such as prolificacy. These results are consistent with others reported in the literature. Several studies, conducted mainly on dairy cattle programs, examined the effect of sire misidentification or missing sire information on the estimation of genetic parameters (Schenkel and Schaeffer, 2000; Roughsedge et al., 2001; Senneke et al., 2004; Parlato and Van Vleck, 2012; Winkelman, 2013; Garritsen et al., 2015). The pedigree error rate for various sheep and dairy cattle populations reached about 10 to 15% (Dodds et al., 2005). In

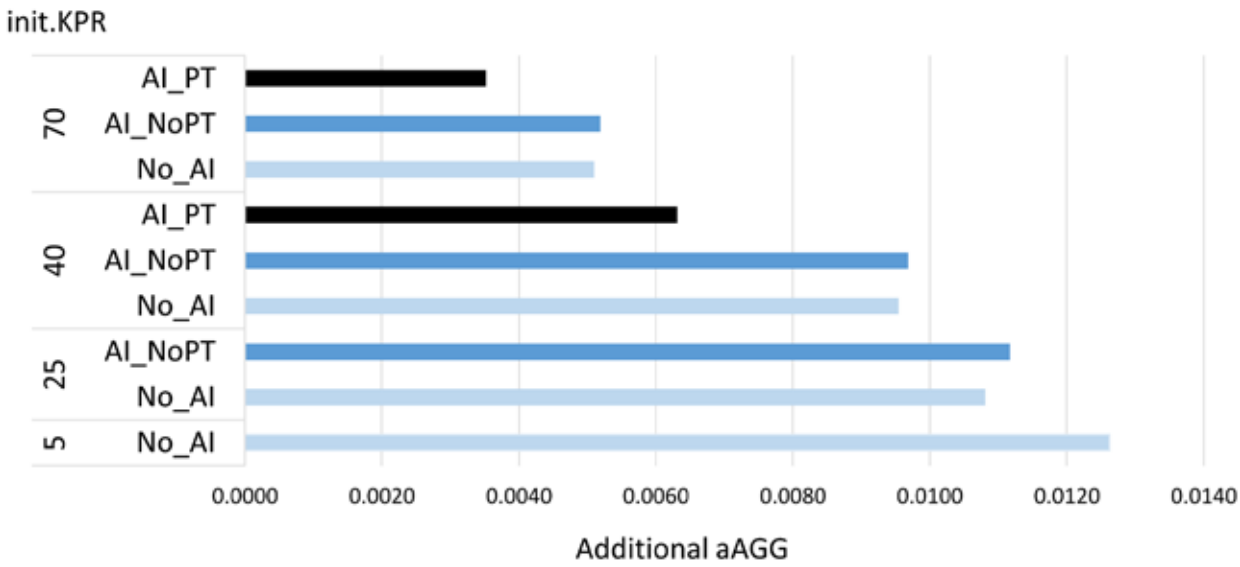


Figure 3. Comparison of the supplementary asymptotic annual genetic gain (aAGG) for the 3 breeding programs: AI_PT (in which male candidates were born from progeny-tested AI sires; black), AI_NoPT (in which male candidates were born from non-progeny-tested AI sires; dark blue), and No_AI (in which male candidates were born from natural mating sires; light blue). The additional aAGG is expressed in genetic SD. Various levels of the initial value of the female known paternity rate (init.KPR) were assessed depending on the minimal level required by each program (e.g., a minimal init.KPR of 40% was required to implement progeny testing in the AI_PT program).

France, the sire misidentification rate for meat sheep breeds reached 7.7% in 2011 (Raoul, 2011). Although most of the studies did not quantify these effects on the genetic gain, Israel and Weller (2000) reported a loss of genetic gain of 4.3% with 10% of incorrect paternity for a trait with a heritability of 0.25. In a similar manner, Harder et al. (2005) predicted a decrease of the response to selection for proven sires (8.6%) and males being progeny tested (12.6%) for a trait with a heritability of 0.25 when 40% of the sire information was missing. Both studies were performed on dairy cattle breeding programs. In another simulation conducted in dairy cattle, Sanders et al. (2006) reported that the simultaneous effect of missing information (10%) and wrong sire information (7%) decreased the accuracy of sire EBV by 4 to 10% depending on the size of the daughter group (from 100 to 10 daughters per sire) for a trait with a heritability of 0.10. In accordance with our results, the effect of missing sire information decreased when the heritability of the considered trait increased. The same study reported that the effect of wrong sire information was greater than that of missing sire information on the reliability of sire EBV. Likewise, Sullivan (1995) estimated a loss of genetic gain of almost 15% when 50% of sires were unknown for a trait with a heritability of 0.05. The effects described are, therefore, of the same order of magnitude as those obtained in our study. However, the explanations are probably different. In the abovementioned studies, wrong and missing sire information mainly affected the accuracy of the EBV for males being progeny testing and proven males. The consequences were mainly a lower selection differential along the sire–male path, which resulted in a loss of genetic progress. In our study, the selection differential that was the most affected was that of the dam–male path because the female KPR had no effect on the sire–male path and only a moderate effect on the sire–female path. In the AI_PT and AI_NoPT programs, male sires are AI sires. Therefore, improving the knowledge of the paternal ancestry of females born from NM sires did not affect the reliability of the EBV of the sires of males.

Potential Economic Benefit of Increasing Known Paternity Rate

Figure 4 shows the relative profit (margin per cost unit) in the entire population for a gain:cost ratio of 5 (monetary gain due to an extra genetic SD divided by

Table 5. Supplementary asymptotic annual genetic gain (aAGG) for various values of trait heritability (and repeatability) in the 3 breeding programs (initial value of the female known paternity rate = 0.05, 0.25, and 0.40, respectively, for the No_AI, AI_NoPT, and AI_PT programs[1])

Breeding Program	h^2 (r)			
	0.1 (0.2)	0.2 (0.35)	0.3 (0.45)	0.4 (0.6)
No_AI	0.013	0.010	0.008	0.006
	16.9%	11.1%	7.9%	5.9%
AI_NoPT	0.011	0.009	0.007	0.006
	13.5%	9.0%	6.5%	4.8%
AI_PT	0.006	0.005	0.004	0.003
	4.6%	3.3%	2.5%	1.9%

[1]AI_PT = male candidates were born from progeny-tested AI sires; AI_NoPT = male candidates were born from non-progeny-tested AI sires; No_AI = male candidates were born from natural mating sires.

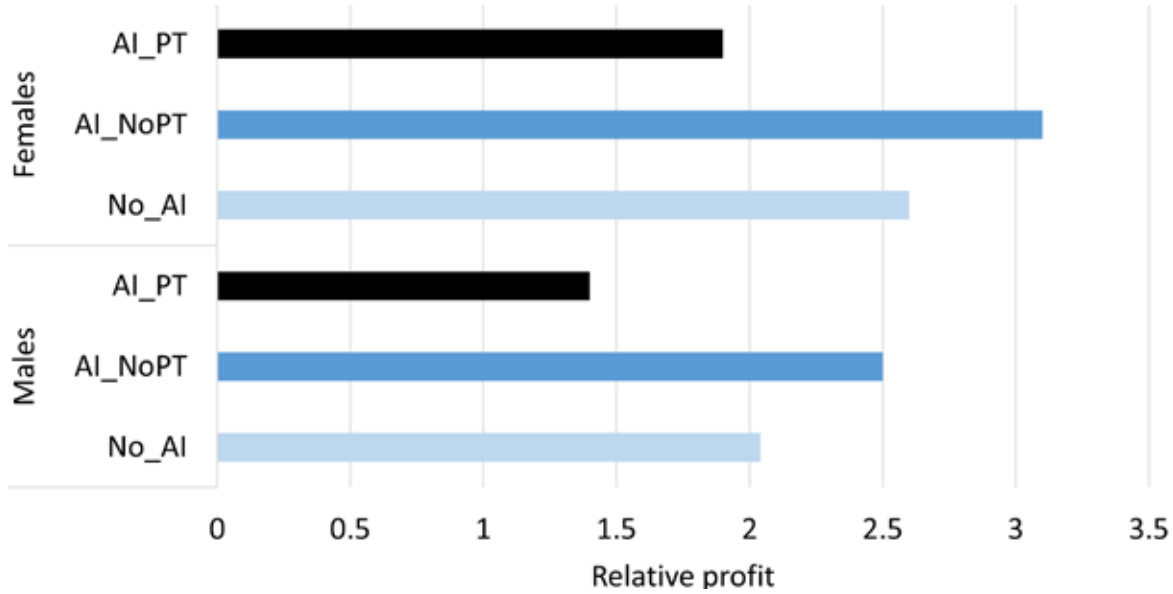


Figure 4. Relative profits (economic margin per cost unit) in the total population when genetic progress was transferred by females (upper part) or males (lower part) in the 3 breeding programs: AI_PT (male candidates were born from progeny-tested AI sires (initial known paternity rate [KPR] = 0.4); black), AI_NoPT (in which male candidates were born from non-progeny-tested AI sires (initial KPR = 0.25); dark blue), and No_AI (in which male candidates were born from natural mating sires (initial KPR = 0.05); light blue). The relative size of the commercial population compared with the nucleus population was 0.5 and the gain:cost ratio was 5. The gain:cost ratio was defined as the monetary gain due to an extra genetic SD divided by the cost of 1 assignment. Relative profits reflect a slow increase in KPR (only selected replacement was assigned) with the shift from initial KPR to complete genealogy (KPR = 1) lasting over 7 reproduction cycles.

the cost of 1 assignment) and a relative size of the commercial population of 0.5. Whatever the way genetic gain was transferred (females or males), the highest profit was obtained for the AI_NoPT breeding program and the lowest for the AI_PT program. The relative profits seem moderate (1.4 to 3.1) given the relatively high gain:cost ratio (in terms of current costs) and the investment period of 60 cycles considered. The ranking (in terms of relative profitability) between breeding programs was consistent but differences were enhanced as the gain:cost ratio and/or the population size increased and lessened as the gain:cost ratio or population size decreased. For a relative size of the commercial population of 1, the profit was higher when the genetic gain was transferred by females rather than by males. This result was as expected for a maternal trait. Indeed, newborn females transferred directly from the nucleus to the industry express the SGG as soon as they start breeding, whereas for newborn males, only half of the SGG is expressed when their daughters start breeding. Even if the asymptotic genetic gain is identical for both methods of transfer, the lag is greater when transferred via males. However, the relative size of the commercial population was limited to 5 at the most when females were used. Male dissemination is, therefore, the only way to transfer the gain from the nucleus to the industry for larger commercial populations and, in these situations, resulted in a higher relative profit. When the 3 breeding programs were compared, it could be observed that the differences in terms of economic results were lower than the differences of additional genetic gain achieved. For example, if the net supplementary AGG for the AI_NoPT program was nearly twice that of the AI_PT program (0.011 and 0.006, respectively), the profit was only 63% higher (3.1 and 1.9, respectively) when transferred via females and 79% higher when transferred via males (2.5 and 1.4, respectively).

As the results obtained for various gain:cost ratios and various relative sizes of the commercial population were consistent among programs, only the results for the No_AI program are shown in subsequent figures. The results for the AI_NoPT and AI_PT programs are available in the Supplementary Results (see the online version of the article at http://journalofanimalscience.org). Figures 5a and 5b show the profit over an investment period of 60 reproduction cycles as a function of the gain:cost ratio and for various relative sizes of the commercial population when males are used to transfer the genetic gain. We observed a linear relation between the profit and the gain:cost ratio whatever the relative size of the commercial population. This result confirmed the strong relationship between the economic value and the ratio, especially the cost of parentage assignment. Some minimal conditions (in terms of minimal gain:cost ratio and/or minimal size of the commercial population) were needed to make assignment profitable. At the nucleus level (i.e., ps = 0), a gain:cost ratio of 2 was insufficient. These results were consistent whatever

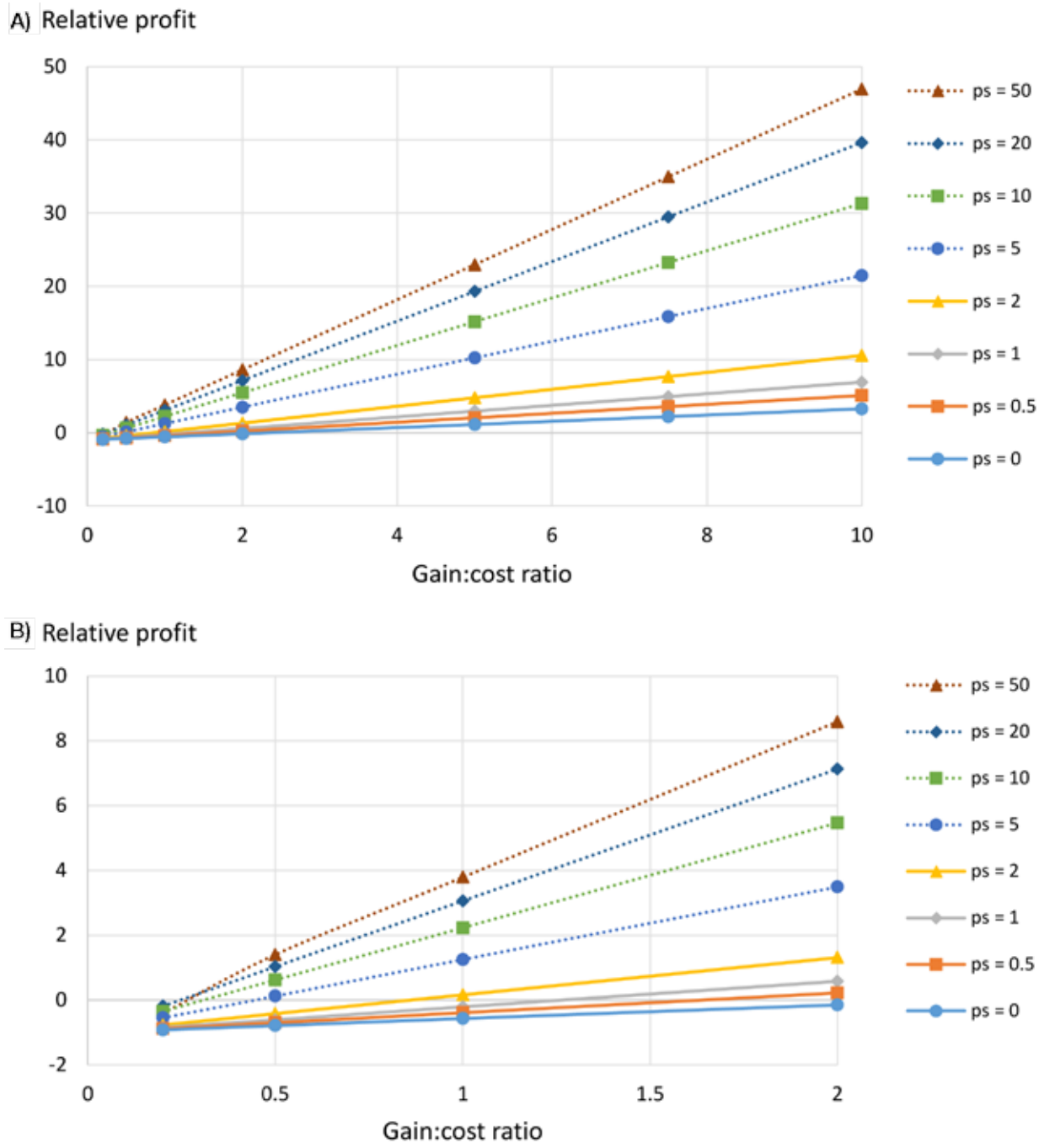


Figure 5. Relative profit resulting from an increase in the female known paternity rate for various gain:cost ratios and relative sizes of the commercial population. Those results were obtained for No_AI program (in which male candidates were born from natural mating sires) and a transfer of genetic progress from the nucleus to the commercial population by males. (a) Full range of gain:cost ratios considered. (b) Limited range of gain:cost ratios [0.2, 2]. We considered an investment period of 60 reproduction cycles. Each curve corresponds to a given population size. *ps* = the relative size of the commercial population compared with the nucleus population (nucleus size = 1). The gain:cost ratio is defined as the monetary gain due to an extra genetic SD divided by the cost of 1 assignment.

the method of transfer used and the breeding program considered. A low gain:cost ratio might be offset by the size of the commercial population. However, for a ratio of 0.2, increasing KPR was never profitable and for a ratio of 0.5, the relative size had to be large ($ps \geq 5$) while the corresponding profit remained very low: relative profit = 0.12 when ps = 5 and relative profit = 1.4 when $ps = 50$. Figure 6a shows the profit as a function of the relative size of the population when males are used to transfer the genetic gain. When the relative size increases, *tdm* decreases: $tdm = 1$ ($ps \leq 5$), $tdm = 0.57$ ($ps = 10$), $tdm = 0.28$ ($ps = 20$), and $tdm = 0.11$ ($ps = 50$). Figure 6b shows the same graph but for commercial population sizes for which all selected young males were born in the nucleus ($tdm = 1$). For a given gain:cost ratio, Fig. 6a shows that the profit increased as the commercial population size grew. However, this increase seemed to converge and reach a plateau. The

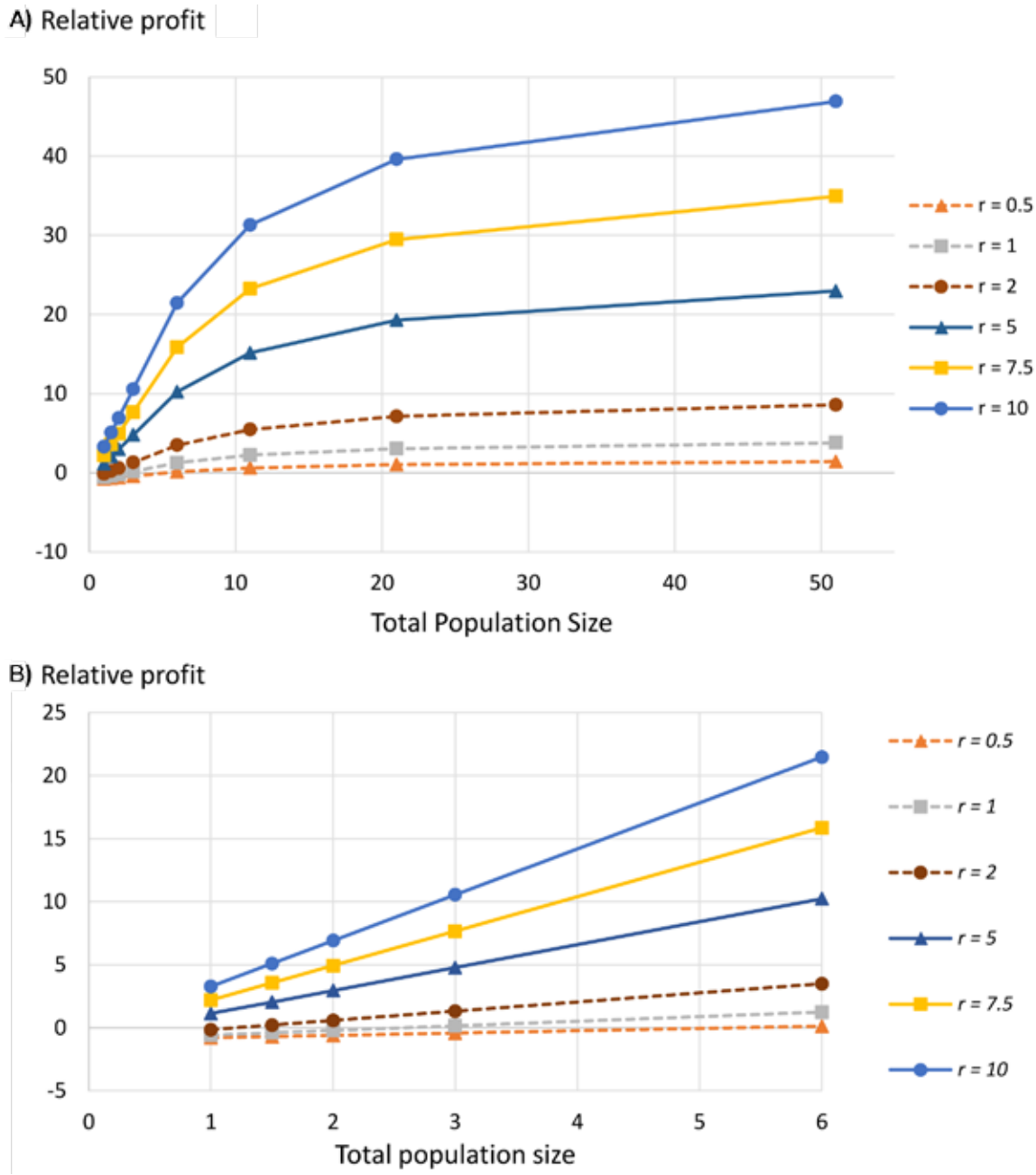


Figure 6. Relative profit as a function of the total population size resulting from an increase of the female known paternity rate for various gain:cost ratios. Those results were obtained for No_AI program (in which male candidates were born from natural mating sires) and a transfer of genetic progress from the nucleus to the commercial population by males. (a) Full range total population sizes considered. (b) Limited range of population sizes for which all replacement males were born in the nucleus. We considered an investment period of 60 reproduction cycles. Each curve corresponds to a given gain:cost ratio ("ratio") defined as the monetary gain due to an extra genetic SD divided by the cost of 1 assignment. The size of the population was defined in relation to the nucleus size (nucleus size = 1).

higher the ratio, the larger the total population size must be to reach this plateau. As shown in Fig. 6b, for smaller population sizes ($ps \leq 5$) in which all sires could be replaced by males from the nucleus, the increase in profit was found to be linear. In such cases, the increase in profit is related to the evolution of the average genetic level of the commercial population. For larger commercial populations, the proportion of selected newborn males born in the nucleus was lower and the SGG diminished. Therefore, for low gain:cost ratios (ratio = 0.5), the increase to profit due to the use of larger commercial populations was very limited because the product of both factors (low ratio × low SGG) tended to 0.

The second economic indicator assessed was the payback period, that is, the minimum period needed to get a positive cash flow. Figures 7a and 7b show the payback period for the No_AI program as a function of the gain:cost ratio when either males or females are used to transfer the genetic gain (Fig. 7a and 7b, respectively). For low gain:cost ratios, the payback period was high and the relative size of the commercial population had a large effect. When the gain:cost ratio increased, both the pay-

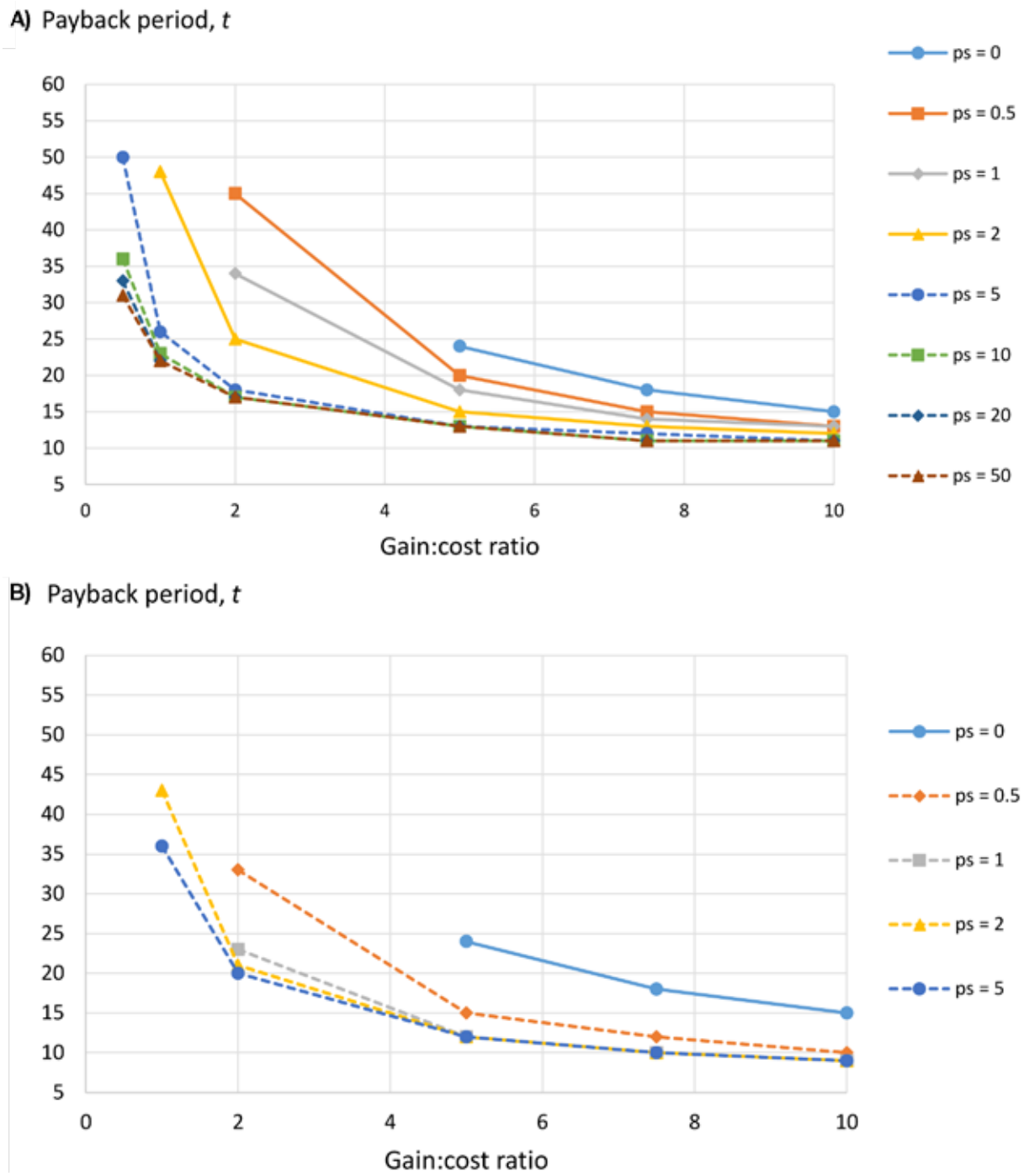


Figure 7. Payback period for various relative sizes of the commercial population as a function of the gain:cost ratio when males or females are used to transfer the genetic gain. Those results were obtained for No_AI program (in which male candidates were born from natural mating sires). (a) Transfer via males. (b) Transfer via females. Each curve corresponds to a given population size. *ps* = the relative size of the commercial population compared with the nucleus population (nucleus size = 1). The gain:cost ratio is defined as the monetary gain due to an extra genetic SD divided by the cost of 1 assignment.

back period and differences observed for different relative sizes were reduced. When transferred by males, and when the gain:cost ratio was not less than 7.5, the payback period converged to approximately 11 reproduction cycles for a population size of at least 5. When females were used to transfer the improvement, the payback period converged to 9 for high gain:cost ratios (≥7.5) and moderately sized commercial populations (≥2). The minimum period of investment remained quite long even for high ratios and large populations due to 1) the delay between assignment (costs) of selected newborns and their first production, 2) the very gradual shift from init.KPR to KPR = 1 (hence the low SGG observed in the nucleus immediately after the shift), and 3) the lag between improvement in the nucleus and the commercial population. The comparison of payback periods for males and females (Fig. 7a and 7b, respectively) showed that for a given gain:cost ratio and a given population size, the

Table 6. Case studies. Potential benefit of increasing the known paternity rate (KPR) in 3 French breeding programs

	Breed[1]		
	Noire du Velay	Romane	Corse
	Breeding program[2]		
Parameter and variable names	No_AI	AI_NoPT	AI_PT
Initial KPR	0.55	0.35	0.5
Size of the nucleus fertile population (F)	6,000	15,000	15,000
Total population ($F(1 + ps^3)$)	18,000	52,500	26,250
Relative size of the commercial population (ps)	2	2.5	0.75
Method of transfer of genetic progress	Females	Females	Males
Proportion of females born from AI sires ($PropF_{AI}$)	0	0.25	0.5
Prolificacy (natural estrus)	1.7	2	1.1
Monetary gain of the selected trait (euros per genetic SD)	7.5	7.5	35
Cost (euros per animal assigned)	18	18	18
Gain:cost ratio	0.4	0.4	1.9
aAGG[4] (initial KPR; expressed in σ_g)	0.0801	0.0834	0.1189
aAGG (KPR = 1; expressed in σ_g)	0.0879	0.0948	0.1210
Supplementary aAGG	9.6%	13.7%	1.8%
Relative profit ($T^5 = 60$), total population	−0.29	−0.12	−0.41
Maximum cost (euros) for assignment to be profitable at $T = 20$ (total population)	5.9	7.1	3.7

[1]Noire du Velay and Romane are 2 meat sheep breeds. Corse is a dairy sheep breed.

[2]AI_PT = male candidates were born from progeny-tested AI sires; AI_NoPT = male candidates were born from non-progeny-tested AI sires; No_AI = male candidates were born from natural mating sires.

[3]ps = relative size of the commercial population compared with the nucleus population.

[4]aAGG = asymptotic annual genetic gain.

[5]T = investment period.

period required to recover the cost of investment is either shorter with females or equivalent. This result was expected for a maternal trait because the gene flow from the nucleus to the commercial population is faster through females. If there is no extra return from the commercial population (i.e., considering the payback period at only the nucleus level), the payback period of 15 reproduction cycles is still relatively high even with a gain:cost ratio of 10. A similar economic study, conducted by Ron et al. (1996) in the Israeli dairy cattle population, reported that profit became positive by year 10.

Practical Case Study

In France, the female KPR in nucleus flocks is highly variable depending on breeds. This heterogeneity is due to variable use of AI and female KPR requirements prescribed by breeders' associations, which determine whether sire information is required or not for the females. In some situations, pedigrees are well known (mostly in terminal sire breeds with a high KPR, close to 1), and in others, they are scarcely known (such as hardy breeds with a low KPR, close to 0.05). We used the model we developed to estimate the potential benefit of increasing the KPR in 3 French breeding programs: the Noire du Velay breed (meat sheep) as an example of a No_AI breeding program (init.KPR = 0.55), the Romane breed (meat sheep) as an example of an AI_NoPT program (init.KPR = 0.35 and $PropF_{AI} = 0.25$), and the Corse breed (dairy sheep) as an example of an AI_PT program (init.KPR = 0.50 and $PropF_{AI} = 0.25$). Table 6 provides the values for the main parameters and the potential benefits in terms of additional genetic progress and economic gain.

The SGG was low for the Corse breed (+1.8%) and quite high for the Noire du Velay (+9.6%) and Romane (+13.7%) breeds. The economic value depended on the SGG transferred from the nucleus to the commercial population and the gain:cost ratio. Due to a medium to high prolificacy, the main method of transferring progress consists in selling newborn females for both the Noire du Velay ($prolifNM = 1.7$ and $ps = 2$) and the Romane ($prolifNM = 2.0$ and $ps = 2.5$) breeds. Transfer through females is, however, quite limited in the Corse breed due to its low prolificacy ($prolifNM = 1.1$) but genetic progress is transferred via NM males ($ps = 0.75$). The economic gain per genetic SD is rather low in meat sheep breeds (€7.5) and, in contrast, relatively high (€35) in the Corse dairy sheep (Cheype et al., 2013). The cost of 1 assignment (€18) used in this case study is the current price in France in 2015. These values resulted in a gain:cost ratio (as defined in our study) that was very low for meat sheep breeds (ratio = 0.4) and moderate for the dairy sheep breed (ratio = 1.8). However, the use of parentage assignment was not profitable for any of the 3 breeding programs. The maximum cost per assignment

for it to become profitable after 20 reproduction cycles was very low for the Corse breed (€3.7 per assignment) and €6 to 7 for the meat sheep breeds (€5.9 for Noire du Velay and €7.1 for Romane). Consequently, the use of SNP-based assignment to increase the knowledge of paternal filiation (female KPR) does not seem profitable in the short term for the breeds in this example given the current price of assignment. However, the relative profit was probably slightly underestimated because it did not take into account the partial reduction of misidentification that can be realized when dams and selected newborn females are genotyped. Moreover, improved profitability could be achieved by combining parentage assignment and genotyping for the genes currently managed in breeding programs such as the *PrP*, myostatin, or ovulation genes (Palhière et al., 2003; Grasset et al., 2009; Martin et al., 2014). The SNP tools currently under development in France for use as parentage tools (Tortereau et al., 2015) or the Illumina sheep low density array (Illumina Inc., San Diego, CA) combine SNP used for parentage assignment and detection of causal mutations. The use of such tools would make parentage assignment profitable for a larger number of breeding programs. The cost of genetic assignment should also be compared with classical procedures used to control paternity that consist of breeding many small female flocks to a single male and also have a cost. Indeed, they require specific flock management, which is time-consuming compared with commercial flock conditions where females are mated with several males (multiple NM). Depending on each breeder's herd management constraints, the use of a SNP-based procedure can represent an attractive alternative. In the future, the parentage tool might offer a larger number of SNP genotyped at a price close to the current one and the value of building a reference population for genomic selection, based on females, should be assessed.

On the whole, this study demonstrates that using SNP-based parentage assignment to increase the KPR is not always profitable and depends on assumptions such as the gain:cost ratio or commercial population size. However, our study did not consider the partial decrease of parentage misidentification and the better genetic connection among flocks (which would result in improved genetic evaluations with better estimations of fixed effects and animal breeding values), which would be associated with parentage assignment and an increase of the KPR. Both of these factors, for which the beneficial effects on genetic progress have been previously discussed (Roughsedge et al., 2001; Sanders et al., 2006; Parlato and Van Vleck, 2012), would increase the value of parentage assignment.

Conclusion

The objective of this study was to predict the technical efficiency, in terms of extra genetic gain, and the potential economic benefit of increasing the female KPR by parentage assignment in small ruminant breeding programs. Depending on how the breeding program was organized, the additional genetic progress achieved in the nucleus ranged from 4 to almost 17%. The 2 economic indicators computed, the long-term profit and payback period, showed that the potential economic benefit was widely influenced by how the genetic progress was transferred to the commercial population (male or female), the relative size of the commercial population, and the gain:cost ratio of assignment. Because small ruminant breeding programs are diverse, there is no single and definitive answer to the question of implementing parentage assignment. Achieving a profit is largely dependent on obtaining a favorable gain:cost ratio, and the model described in this paper represents a very useful tool for assessing the value of SNP-based parentage assignment in different situations, notably because it allows specific parameters to be taken into consideration. For example, the model can take into account various origins of sires of selected males (a combination of NM sires and progeny tested sires).

Taking into account the SGG and additional economic revenue that could be achieved by reducing parentage misidentification should slightly improve the value of parentage assignment. Furthermore, the use of SNP tools combining parentage assignment and major gene genotyping would make parentage assignment beneficial for a larger number of breeding programs.

LITERATURE CITED

Analla, M., A. Sanchez-Palma, A. Munõz-Serrano, and J. M. Serradilla. 1995. Simulation analysis with BLUP methodology of different data structures in goat selection schemes in Spain. Small Rumin. Res. 17:51–55. doi:10.1016/0921-4488(95)00656-6

Auvray, B., J. C. McEwan, S.-A. Newman, M. Lee, and K. G. Dodds. 2014. Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. J. Anim. Sci. 92:4375–4389. doi:10.2527/jas.2014-7801

Baloche, G., A. Legarra, G. Sallé, H. Larroque, J.-M. Astruc, C. Robert-Granié, and F. Barillet. 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. J. Dairy Sci. 97:1107–1116. doi:10.3168/jds.2013-7135

Bell, A. M., J. M. Henshall, S. Gill, K. Gore, and J. W. Kijas. 2013. Success rates of commercial SNP based parentage assignment in sheep. In: Proc. Assoc. Adv. Anim. Breed. Genet., vol. 20. p. 278–281.

Besbes, B. 1990. Evaluation du contrôle de la paternité chez les ovins: Comparaison des réponses à la sélection obtenues en présence et en absence du contrôle de la paternité. (In French.) PhD Diss., CIHEAM, Zaragoza, Spain.

Bijma, P., and J. A. Woolliams. 1999. Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151:1197–1210.

Bijma, P., and J. A. Woolliams. 2000. On the relation between gene flow theory and genetic gain. Genet. Sel. Evol. 32:99–104. doi:10.1186/1297-9686-32-1-99

Carillier, C., H. Larroque, and C. Robert-Granié. 2014. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genet. Sel. Evol. 46:67. doi:10.1186/s12711-014-0067-3

Cheype, A., J. Guerrier, F. Tortereau, D. François, J. P. Poivey, K. Chile, and J. Raoul. 2013. Economical weighting of breeding objectives and definition of total merit indexes in BMC sheep breed. In: Book of abstracts of the 64th annual meeting of the European Federation of Animal Science, Nantes, France. p. 319.

Clarke, S. M., H. M. Henry, K. G. Dodds, T. W. D. Jowett, T. R. Manley, R. M. Anderson, and J. C. McEwan. 2014. A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. PLoS One 9:e93392. doi:10.1371/journal.pone.0093392

Daetwyler, H. D., S. Bolormaa, D. J. Brown, J. H. J. Van der Werf, and B. J. Hayes. 2013. A genomic prediction cross-validation approach combining ewe repeated phenotypes and ram daughter trait deviations. In: Proc. Assoc. Adv. Anim. Breed. Genet., vol. 20. p. 360–363.

Dodds, K. G., M. L. Tate, and J. A. Sise. 2005. Genetic evaluation using parentage information from genetic markers. J. Anim. Sci. 83:2271–2279.

Douguet, M., J.-M. Astruc, and G. Thomas. 2015. Résultats de contrôle laitier – France 2014. (In French.) http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/resultats-de-controle-laitier-france-2013.html (Accessed 15 November 2015.)

Ducrocq, V., and R. L. Quaas. 1988. Prediction of genetic response to truncation selection across generations. J. Dairy Sci. 71:2543–2553. doi:10.3168/jds.S0022-0302(88)79843-4

Elsen, J. M. 1992. La gestion des populations: De l'optimisation au progrès génétique réalisé dans les schémas de sélection. (In French.) Prod. Anim. (Paris, Fr.) 5:237–242.

Elsen, J. M. 1993. Prediction of annual genetic gain and improvement lag between populations. Genet. Sel. Evol. 25:75–82. doi:10.1186/1297-9686-25-1-75

Elsen, J., and J. Mocquot. 1974. Méthode de prévision de l'évolution du niveau génétique d'une population soumise à une opération de sélection et dont les générations se chevauchent. (In French.) INRA Bull. Tech. Dépt. Génét. Anim., vol. 17. p. 30–54.

Garritsen, C., E. Van Marle-Köster, M. A. Snyman, and C. Visser. 2015. The impact of DNA parentage verification on breeding value estimation and sire ranking in South African Angora goats. Small Rumin. Res. 124:30–37. doi:10.1016/j.smallrumres.2015.01.008

Grasset, D., J. Bouix, B. Bibé, H. Leveziel, M. Georges, and E. Laville. 2009. Le gène d'hypertrophie musculaire du «Texel Belge»: Identification, impact, introgression. In: Rencontres autour des recherches sur les ruminants. (In French.) Paris, France. p. 415–418.

Harder, B., J. Bennewitz, N. Reinsch, M. Mayer, and E. Kalm. 2005. Effect of missing sire information on genetic evaluation. Arch. Tierz. 48:219–232.

Heaton, M. P., K. A. Leymaster, T. S. Kalbfleisch, J. W. Kijas, S. M. Clarke, J. McEwan, J. F. Maddox, V. Basnayake, D. T. Petrik, B. Simpson, T. P. L. Smith, and C. G. Chitko-McKown, and the International Sheep Genomics Consortium. 2014. SNPs for parentage testing and traceability in globally diverse breeds of sheep. PLoS One 9:e94851. doi:10.1371/journal.pone.0094851

Hill, W. G. 1974. Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. 18:117–139. doi:10.1017/S0003356100017372

Israel, C., and J. I. Weller. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. J. Dairy Sci. 83:181–187. doi:10.3168/jds.S0022-0302(00)74869-7

Khang, J. V. T. 1983. Méthodes d'analyse des données démographiques et généalogiques dans les populations d'animaux domestiques. (In French.) Genet. Sel. Evol. 15:263–298. doi:10.1186/1297-9686-15-2-263

Martin, P., J. Raoul, and L. Bodin. 2014. Effects of the FecL major gene in the Lacaune meat sheep population. Genet. Sel. Evol. 46:48. doi:10.1186/1297-9686-46-48

Meuwissen, T., B. Hayes, and M. Goddard. 2013. Accelerating improvement of livestock with genomic selection. Annu. Rev. Anim. Biosci. 1:221–237. doi:10.1146/annurev-animal-031412-103705

Moghaddar, N., A. A. Swan, and J. H. J. Van der Werf. 2013. Accuracy of genomic prediction from multi-breed sheep reference population. In: Proc. Assoc. Adv. Anim. Breed. Genet., vol. 20. p. 220–223.

Palhière, I., M. Brochard, J. M. Astruc, and F. Barillet. B. Bed'Hom, B. Bibé, J. Bouix, O. Catrou, J. M. Elsen, D. François, L. Griffon, E. Jullien, M. Orlianges, G. Perret, P. Tribon. 2003. Breeding for scrapie resistance in France. In: Proceedings of the international workshop on major genes and QTL in sheep and goats, CD-ROM communication no. 3-03.

Parlato, E., and L. D. Van Vleck. 2012. Effect of parentage misidentification on estimates of genetic parameters for milk yield in the Mediterranean Italian buffalo population. J. Dairy Sci. 95:4059–4064. doi:10.3168/jds.2011-4855

Pickering, N. K., K. G. Dodds, B. Auvray, and J. C. McEwan. 2013. The impact of genomic selection on genetic gain in the New Zealand sheep dual purpose selection index. In: Proc. Assoc. Adv. Anim. Breed. Genet., vol. 20. p. 175–178.

Raoul, J. 2011. Contrôles de filiations ovins allaitants 2011. Bulletin Geovial 11. (In French.) http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/geovial-n-11-2012.html (Accessed 15 November 2015)

Rendel, J. M., and A. Robertson. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J. Genet. 50:1–8. doi:10.1007/BF02986789

Ron, M., Y. Blanc, M. Band, E. Ezra, and J. I. Weller. 1996. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. J. Dairy Sci. 79:676–681. doi:10.3168/jds.S0022-0302(96)76413-5

Roughsedge, T., S. Brotherstone, and P. M. Visscher. 2001. Bias and power in the estimation of a maternal family variance component in the presence of incomplete and incorrect pedigree information. J. Dairy Sci. 84:944–950. doi:10.3168/jds.S0022-0302(01)74552-3

Rupp, R., S. Mucha, H. Larroque, J. McEwan, and J. Conington. 2016. Genomic application in sheep and goat breeding. Anim. Front. 6(1). doi:10.2527/af.2016-0006

Sanders, K., J. Bennewitz, and E. Kalm. 2006. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. J. Dairy Sci. 89:315–321. doi:10.3168/jds.S0022-0302(06)72096-3

Schenkel, F. S., and L. R. Schaeffer. 2000. Effects of nonrandom parental selection on estimation of variance components. J. Anim. Breed. Genet. 117:225–239. doi:10.1046/j.1439-0388.2000.00262.x

Schenkel, F. S., L. R. Schaeffer, and P. J. Boettcher. 2002. Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. Genet. Sel. Evol. 34:41–59. doi:10.1186/1297-9686-34-1-41

Senneke, S. L., M. D. MacNeil, and L. D. Van Vleck. 2004. Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. J. Anim. Sci. 82:2307–2312.

Shumbusho, F., J. Raoul, J. M. Astruc, I. Palhière, and J. M. Elsen. 2013. Designing, technical evaluation and profitability estimation of breeding strategies based on molecular information for small ruminant species. J. Anim. Sci. 91:3644–3657. doi:10.2527/jas.2012-6205

Shumbusho, F., J. Raoul, J. M. Astruc, I. Palhiere, S. Lemarié, A. Fugeray-Scarbel and J. M. Elsen. 2016. Economic evaluation of genomic selection in small ruminants: A sheep meat breeding program. Animal 10:1033–1041. doi:10.1017/S1751731115002049

Sullivan, P. G. 1995. Alternatives for genetic evaluation with uncertain parentage. Can. J. Anim. Sci. 75:31–36. doi:10.4141/cjas95-004

Swan, A. A., D. J. Brown, H. D. Daetwyler, B. J. Hayes, M. Kelly, N. Moghaddar, and J. H. J. Van der Werf. 2014. Genomic evaluations in the Australian sheep industry. In: Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada. paper 334.

Tortereau, F., C. Moreno, G. Tosser-Klopp, L. Barbotte, L. Genestout, and J. Raoul. 2015. Development of a SNP parentage assignment panel for French sheep breeds. In: Book of abstracts of the 64th annual meeting of the European Federation of Animal Science, Warsaw, Poland. p. 510.

Van der Werf, J. H. J. 2014. Genomic selection in sheep breeding programs. In: Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada. paper 351.

Van Eenennaam, A. L., R. L. Weaber, D. J. Drake, M. C. T. Penedo, R. L. Quaas, D. J. Garrick, and E. J. Pollak. 2007. DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting. J. Anim. Sci. 85:3159–3169. doi:10.2527/jas.2007-0284

Winkelman, A. M. 2013. Effect of daughter misidentification on dairy sire evaluation. In: Proc 20th Conf. Assoc. Adv. Anim. Breed. Genet., Translating Science into Action, Napier, New Zealand, 20th–23rd October 2013. p. 25–28.

Woolliams, J. A., P. Bijma, and B. Villanueva. 1999. Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153:1009–1020.

SUPPLEMENTARY MATERIALS-1: ***INFORMATION CONSIDERED FOR A FEMALE TARGET INDIVIDUAL (TI).***

We considered that the maximum rank (parity) for females was 7 ($a = 7$).
l = lambing (lactation) rank

The number of own records, $\bar{m}_p(i)$, is equal to the rank of the considered category.

The average number of records per dam, $\bar{m}_d(i)$***,*** depended on the number of measures known for a dam at the time of birth of the target individual and all new measures for the dams cumulated until the target individual reaches the rank considered.
The example below illustrates how $\bar{m}_d(i)$ was computed for females (TI) of rank 1.
Considering the absence of selection along the dam-female path, all cohorts of females contributed to the selected newborn females in proportion to their numbers. Individuals were born from α_1 rank_1-dams, α_2 rank_2-dams, α_3 rank_3-dams and so on. Thus the average number of records known per dam at an individual's birth was $\sum_{l=1}^{7} \alpha_l * (l)$.
As the population was in a steady state, we knew that the females present in proportion α_l at time t would be present in proportion α_{l+1} at time t+1 ($l = [1, 7-1]$).
We assumed that the first rank of the target individual occurs at 2 reproduction cycles after its birth (birth(TI)). At birth(TI)+1, rank_1-dams were in proportion α_2, rank_2-dams in proportion α_3 and so on. Rank_7-dams were culled. Thus the additional records for dams at birth(TI)+1 were $\sum_{l=2}^{7} \alpha_l$.
Following the same process, the additional records for dams at birth(TI)+2 were $\sum_{l=3}^{7} \alpha_l$.

To sum up, $\bar{m}_m(i)$ of females (TI) of rank 1 was equal to:
$\bar{m}_d(i) = \sum_{l=1}^{7} \alpha_l * (l) + \sum_{l=2}^{7} \alpha_l + \sum_{l=3}^{7} \alpha_l$

Over time, until all TI dams are culled, new records of TI dams accumulate as the rank of TI increases. The following table summarizes the increase in dam records given the rank of the target individual.

TI Rank	New records of TI dams
2	$\sum_{l=4}^{7} \alpha_l$
3	$\sum_{l=5}^{7} \alpha_l$
4	$\sum_{l=6}^{7} \alpha_l$
5	$\sum_{l=7}^{7} \alpha_l$
6	All dams have been culled

The average number of records for maternal grand-dams, $\bar{m}_{mgd}(i)$, and paternal grand-dams, $\bar{m}_{pgd}(i)$, were computed in the same way as for dam records. The average number of records for paternal grand-dams in 'unknown sire' categories was 0.

We computed $\bar{m}_{mgd}(i)$ for target individuals of rank 1 (time = birth(TI)+2)

For ease of understanding, we rename $d\alpha_l$, the proportion of rank-l dams of target individuals, and $gd\alpha_l$ the proportion of rank-l grand-dams of target individuals.

We know that the proportion of rank_1-dams at the birth of the target individual is $d\alpha_1$

Rank_1 dams were born at birth(TI)-2 from l rank grand-dams present in proportion $gd\alpha_l$.

At birth(TI)-2, $gd\alpha_1$ had one record, $gd\alpha_2$ had two records and so on. The total number of records were $\sum_{l=1}^{7} gd\alpha_l * (l)$.

At birth(TI)-1, rank_1-grand-dams were present in proportion $gd\alpha_2$, rank_2 grand-dams in proportion $gd\alpha_3$ and so on. The total number of new records was $\sum_{l=2}^{7} \mathrm{gd}\alpha_l$ at birth(TI)-1, $\sum_{l=3}^{7} gd\alpha_l$ at birth(TI), $\sum_{l=4}^{7} \mathrm{gm}\alpha_l$ at birth(TI)+1 and $\sum_{l=5}^{7} \mathrm{gm}\alpha_l$ at birth(TI)+2.

The total number of records for maternal grand-dams for a TI born from a rank_1 dam was:

$$d\alpha_1 * [(\sum_{l=1}^{7} gd\alpha_l * (l) + \sum_{l=2}^{7} \mathrm{gd}\alpha_l + \sum_{l=3}^{7} gd\alpha_l + \sum_{l=4}^{7} \mathrm{gd}\alpha_l + \sum_{l=5}^{7} \mathrm{gd}\alpha_l)]$$

Applying this reasoning to other rank dams we obtained:

$$\begin{aligned}\bar{\mathrm{m}}_{\mathrm{mgd}} = {} & d\alpha_1 * [(\sum_{l=1}^{7} gd\alpha_l * (l) + \sum_{l=2}^{7} gd\alpha_l + \sum_{l=3}^{7} gd\alpha_l + \sum_{l=4}^{7} gd\alpha_l + \sum_{l=5}^{7} gd\alpha_l)] + d\alpha_2 * \\ & [(\sum_{l=1}^{7} gd\alpha_l * (l) + \sum_{l=2}^{7} gd\alpha_l + \sum_{l=3}^{7} gd\alpha_l + \sum_{l=4}^{7} gd\alpha_l + \sum_{l=5}^{7} gd\alpha_l + \sum_{l=6}^{7} gd\alpha_l)] \\ & + (d\alpha_3 + d\alpha_4 + d\alpha_5 + d\alpha_6 + d\alpha_7) * [(\sum_{l=1}^{7} gd\alpha_l * (l) + \sum_{l=2}^{7} gd\alpha_l + \sum_{l=3}^{7} gd\alpha_l + \\ & \sum_{l=4}^{7} gd\alpha_l + \sum_{l=5}^{7} gd\alpha_l + \sum_{l=6}^{7} gd\alpha_l + \sum_{l=7}^{7} gd\alpha_l)]\end{aligned}$$

Knowing that $gd\alpha_l = \mathrm{d}\alpha_l = \alpha_l$ and $\sum_{l=1}^{7} \alpha_l = 1$

$$\begin{aligned}\bar{\mathrm{m}}_{\mathrm{mgd}} = {} & [(\sum_{l=1}^{7} \alpha_l * (l) + \sum_{l=2}^{7} \alpha_l + \sum_{l=3}^{7} \alpha_l + \sum_{l=4}^{7} \alpha_l + \sum_{l=5}^{7} \alpha_l)] \\ & + [(\sum_{l=2}^{7} \alpha_l) * (\sum_{l=6}^{7} \alpha_l)] + [(\sum_{l=3}^{7} \alpha_l) * (\sum_{l=7}^{7} \alpha_l)]\end{aligned}$$

Over time, the number of grand-dams of females (TI) decreases and new records are cumulated for living grand-dams as the rank of target individual increases. The following table summarizes the increase in grand-dam records given the rank of the target individual.

TI Rank	New records for grand-dams
2	$\mathrm{d}\alpha_1 * (\mathrm{gd}\alpha_6 + \mathrm{gd}\alpha_7) + \mathrm{d}\alpha_2 * gd\alpha_7 = \alpha_1 * (\alpha_6 + \alpha_7) + \alpha_2 * \alpha_7$
3	$\mathrm{d}\alpha_1 * gd\alpha_7 = \alpha_1 * \alpha_7$
4	All grand-dams have been culled

We computed $\bar{\boldsymbol{m}}_{\boldsymbol{pgd}}(\boldsymbol{i})$ following the same reasoning. $\bar{\boldsymbol{m}}_{\boldsymbol{pgd}}(\boldsymbol{i})$ depends on sire category.
The average number of records for paternal grand-dams in 'unknown sire' categories was 0.
For other categories, $\bar{\boldsymbol{m}}_{\boldsymbol{pgd}}(\boldsymbol{i})$ can be deduced from maternal grand-dams formulae and Figure 1 which indicates the delay between sire birth and use for breeding.

Sire = natural mating sire (1st use), non-progeny-tested AI sires (1st use) and males being progeny tested.
The number of records for their dams (paternal grand-dams of TI) was identical to that for dams of rank_1-dams.

$\bar{\mathrm{m}}_{\mathrm{pgd}} = \sum_{l=1}^{7} gd\alpha_l * (l) + \sum_{l=2}^{7} gd\alpha_l + \sum_{l=3}^{7} gd\alpha_l + \sum_{l=4}^{7} gd\alpha_l + \sum_{l=5}^{7} gd\alpha_l = \sum_{l=1}^{7} \alpha_l * (l) + \sum_{l=2}^{7} \alpha_l + \sum_{l=3}^{7} \alpha_l + \sum_{l=4}^{7} \alpha_l + \sum_{l=5}^{7} \alpha_l$

TI Rank	New records for grand-dams
2	$\mathrm{gd}\alpha_6 + \mathrm{gd}\alpha_7 = \alpha_6 + \alpha_7$
3	$gd\alpha_7 = \alpha_7$
4	All grand-dams have been culled

Sire = natural mating sire (2nd use), non-progeny-tested AI sires (2nd use).
The number of records for their dams (paternal grand-dams of TI) was identical to that for dams of rank_2-dams.

$\bar{\mathrm{m}}_{\mathrm{pgd}} = \sum_{l=1}^{7} gd\alpha_l * (l) + \sum_{l=2}^{7} gd\alpha_l + \sum_{l=3}^{7} gd\alpha_l + \sum_{l=4}^{7} gd\alpha_l + \sum_{l=5}^{7} gd\alpha_l + \sum_{l=6}^{7} gd\alpha_l = \sum_{l=1}^{7} \alpha_l * (l) + \sum_{l=2}^{7} \alpha_l + \sum_{l=3}^{7} \alpha_l + \sum_{l=4}^{7} \alpha_l + \sum_{l=5}^{7} \alpha_l + \sum_{l=6}^{7} \alpha_l$

TI Rank	New records for grand-dams
2	$gd\alpha_7 = \alpha_7$
3	All grand-dams have been culled

Sire = natural mating sire (3rd use), proven sires (1st – 3rd use).
The number of records of their dams (paternal grand-dams of TI) was identical to that for dams of rank_3-dams/rank_5-dams.

$\bar{m}_{pgd} = \sum_{l=1}^{7} gd\alpha_l * (l) + \sum_{l=2}^{7} gd\alpha_l + \sum_{l=3}^{7} gd\alpha_l + \sum_{l=4}^{7} gd\alpha_l + \sum_{l=5}^{7} gd\alpha_l + \sum_{l=6}^{7} gd\alpha_l +$
$\sum_{l=7}^{7} gd\alpha_l = \sum_{l=1}^{7} \alpha_l * (l) + \sum_{l=2}^{7} \alpha_l + \sum_{l=3}^{7} \alpha_l + \sum_{l=4}^{7} \alpha_l + \sum_{l=5}^{7} \alpha_l + \sum_{l=6}^{7} \alpha_l + \sum_{l=7}^{7} \alpha_l +$

TI Rank	New records for grand-dams
2	All grand-dams have been culled

The average number of maternal half-sisters, $\bar{n}_{mhs}(i)$: Target individuals (TI) were born from α_1 rank_1-dams, α_2 rank_2-dams, α_3 rank_3-dams and so on. For a TI of rank 1, the only progeny of rank_1-dam is the TI. Contrary to rank_1-dams, rank_7-dams have 6 previous lambings/kiddings from birth(TI)-6 to birth(TI)-1. Computing $\bar{n}_{mhs}(i)$ takes into account the probability for a dam to produce a selected newborn female. By simplification, the number of daughters procreated by dam at each reproduction cycle is 0 or 1. As there was no selection along the dam-female path, the probability that a dam, whatever her category, produces a selected newborn female, was equal to the proportion of rank1-females (α_1).

Thus $\bar{n}_{mhs}(i)$ was equal to the number of dam lambings at the target individual's birth ($\sum_{l=1}^{7} d\alpha_l * l$), minus 1 (record that produced the target individual), multiplied by the probability to produce a selected newborn female:

($d\alpha_l$, the proportion of dams of target individuals of rank l)

$$\bar{n}_{mhs}(i) = \sum_{l=1}^{7} d\alpha_l * (l-1) * \alpha_1 = \sum_{l=1}^{7} \alpha_l * (l-1) * \alpha_1$$

Over time, the number of new maternal half-sisters depends on the probability that a TI dam is still alive and produces a selected daughter. The following table summarizes the increase in the number of maternal half-sisters given the rank of the target individual.

TI Rank	Number of new maternal half-sisters
2	$\sum_{l=2}^{7} d\alpha_l * \alpha_1 = \sum_{l=2}^{7} \alpha_l * \alpha_1$
3	$\sum_{l=3}^{7} d\alpha_l * \alpha_1 = \sum_{l=3}^{7} \alpha_l * \alpha_1$
4	$\sum_{l=4}^{7} d\alpha_l * \alpha_1 = \sum_{l=4}^{7} \alpha_l * \alpha_1$
5	$\sum_{l=5}^{7} d\alpha_l * \alpha_1 = \sum_{l=5}^{7} \alpha_l * \alpha_1$
6	$\sum_{l=6}^{7} d\alpha_l * \alpha_1 = \sum_{l=6}^{7} \alpha_l * \alpha_1$
7	$\sum_{l=7}^{7} d\alpha_l * \alpha_1 = \sum_{l=7}^{7} \alpha_l * \alpha_1$

The average number of records per maternal half-sister, $\bar{m}_{mhs}(i)$: Contrary to rank_1-dams, rank_7-dams have 6 previous lambings/kiddings from birth(TI)-6 to birth(TI)-1. At each time, the probability that a dam produces a selected newborn female is equal to α_1.The number of records for maternal half-sisters depends on their birth time. Half-sisters born at birth(TI)-6 potentially have 7 records when the TI is of rank 1 (time=birth(TI)+2), but only a proportion of half-sisters born at birth(TI)-6 (α_1) is still present at birth(TI)+2 (α_7).

The following table gives the total records for maternal half-sisters from rank_7-dams depending on the birth time of the half-sisters (column) and the record time (row). Remember that the delay between birth and rank 1 is 2 reproduction cycles.

		Birth time of half-sister						
		BTI-6	BTI-5	BTI-4	BTI-3	BTI-2	BTI-1	BTI
Half-sister Record Time	BTI-4	$d\alpha_7 * \alpha_1$						
	BTI-3	$d\alpha_7 * \alpha_2$	$d\alpha_7 * \alpha_1$					
	BTI-2	$d\alpha_7 * \alpha_3$	$d\alpha_7 * \alpha_2$	$d\alpha_7 * \alpha_1$				
	BTI-1	$d\alpha_7 * \alpha_4$	$d\alpha_7 * \alpha_3$	$d\alpha_7 * \alpha_2$	$d\alpha_7 * \alpha_1$			
	BTI	$d\alpha_7 * \alpha_5$	$d\alpha_7 * \alpha_4$	$d\alpha_7 * \alpha_3$	$d\alpha_7 * \alpha_2$	$d\alpha_7 * \alpha_1$		
	BTI+1	$d\alpha_7 * \alpha_6$	$d\alpha_7 * \alpha_5$	$d\alpha_7 * \alpha_4$	$d\alpha_7 * \alpha_3$	$d\alpha_7 * \alpha_2$	$d\alpha_7 * \alpha_1$	
	BTI+2	$d\alpha_7 * \alpha_7$	$d\alpha_7 * \alpha_6$	$d\alpha_7 * \alpha_5$	$d\alpha_7 * \alpha_4$	$d\alpha_7 * \alpha_3$	$d\alpha_7 * \alpha_2$	$d\alpha_7 * TI$

(BTI=Birth(TI))

As a second example, the following table gives the total records for maternal half-sisters from rank_6-dams.

		Birth time of half-sister						
		BTI-6	BTI-5	BTI-4	BTI-3	BTI-2	BTI-1	BTI
Half-sister Record Time	BTI-4							
	BTI-3		$d\alpha_6 * \alpha_1$					
	BTI-2		$d\alpha_6 * \alpha_2$	$d\alpha_6 * \alpha_1$				
	BTI-1		$d\alpha_6 * \alpha_3$	$d\alpha_6 * \alpha_2$	$d\alpha_6 * \alpha_1$			
	BTI		$d\alpha_6 * \alpha_4$	$d\alpha_6 * \alpha_3$	$d\alpha_6 * \alpha_2$	$d\alpha_6 * \alpha_1$		
	BTI+1		$d\alpha_6 * \alpha_5$	$d\alpha_6 * \alpha_4$	$d\alpha_6 * \alpha_3$	$d\alpha_6 * \alpha_2$	$d\alpha_6 * \alpha_1$	
	BTI+2		$d\alpha_6 * \alpha_6$	$d\alpha_6 * \alpha_5$	$d\alpha_6 * \alpha_4$	$d\alpha_6 * \alpha_3$	$d\alpha_6 * \alpha_2$	$d\alpha_6 * TI$

We can express the total number of progeny records (including TI = α_1) in terms of l, the dam rank, and l', the daughter rank, as $d\alpha_l * \sum_{l'=1}^{l}[(l - l' + 1) * \alpha_{l'}]$. Thus the total of records for maternal half-sisters was:

$$\bar{m}_{mhs}(i) = \sum_{l=1}^{7}\left(d\alpha_l * \left(\left(\sum_{l'=1}^{l}(l - l' + 1) * \alpha_{l'}\right) - \alpha_1\right)\right)$$

$$= \sum_{l=1}^{7}\left(\alpha_l * \left(\left(\sum_{l'=1}^{l}(l - l' + 1) * \alpha_{l'}\right) - \alpha_1\right)\right)$$

Over time, the number of new maternal half-sisters depends on (i) the probability that a TI dam is still alive and produces a selected daughter, (ii) a new record exists for previous maternal half-sisters. Both factors depend on the rank of the dam of the TI.

The following table gives the new records for maternal half-sisters from rank_7-dams, known at Birth(TI)+3.

		Birth time of half-sister							
		BTI-6	BTI-5	BTI-4	BTI-3	BTI-2	BTI-1	BTI	BTI+1
Half-sister Record Time	BTI+3		$d\alpha_7 * \alpha_7$	$d\alpha_7 * \alpha_6$	$d\alpha_7 * \alpha_5$	$d\alpha_7 * \alpha_4$	$d\alpha_7 * \alpha_3$	TI	

By definition, no new progeny were born from those dams at BTI+1. Half-sisters born at BTI-6 were of rank 7 at BTI+3, thus there were no new records for these females at BTI+3.

The following table gives the new records for maternal half-sisters from rank_6-dams, known at Birth(TI)+3. Because a proportion of rank_6-dams was culled at BTI+1, only $d\alpha_7 * \alpha_1$ daughters were of rank 1 at BTI+3.

		Birth time of half-sister							
		BTI-6	BTI-5	BTI-4	BTI-3	BTI-2	BTI-1	BTI	BTI+1
Half-sister Record Time	BTI+3		$d\alpha_6 * \alpha_7$	$d\alpha_6 * \alpha_6$	$d\alpha_6 * \alpha_5$	$d\alpha_6 * \alpha_4$	$d\alpha_6 * \alpha_3$	TI	$d\alpha_7 * \alpha_1$

We can express the number of new records for progeny born before the TI, in terms of l, the dam rank, and l', the daughter rank as $\sum_{l=2}^{7}\left(d\alpha_l * \left(\sum_{l'=3}^{\min(l+1,7)} \alpha_{l'}\right)\right)$. Records from new maternal half-sisters were those of new maternal half-sisters selected when TI reached rank 2 ($\sum_{l=2}^{7} d\alpha_l * \alpha_1$).

By summing the two components, the number of new records was obtained for each rank of the TI. The following table summarizes the increase in the number of records for maternal half-sisters given the rank of the target individual.

Rank	New records of maternal half-sisters $\bar{m}_{mhs}$
2	$\sum_{l=2}^{a}\left(d\alpha_l * \left(\sum_{l'=3}^{\min(l+1,7)} \alpha_{l'}\right)\right) + \sum_{l=2}^{7} d\alpha_l * \alpha_1 = \sum_{l=2}^{7}\left(\alpha_l * \left(\sum_{l'=3}^{\min(l+1,7)} \alpha_{l'}\right)\right) + \sum_{l=2}^{7} \alpha_l * \alpha_1$
3	$\sum_{l=2}^{7}\left(\alpha_l * \left(\sum_{l'=4}^{\min(l+2,7)} \alpha_{l'}\right)\right) + \sum_{l=2}^{7} \alpha_l * \alpha_2 + \sum_{l=3}^{7} \alpha_l * \alpha_1$
4	$\sum_{l=2}^{7}\left(\alpha_l * \left(\sum_{l'=5}^{\min(l+3,7)} \alpha_{l'}\right)\right) + \sum_{l=2}^{7} \alpha_l * \alpha_3 + \sum_{l=3}^{7} \alpha_l * \alpha_2 + \sum_{l=4}^{7} \alpha_l * \alpha_1$
5	$\sum_{l=2}^{7}\left(\alpha_l * \left(\sum_{l'=6}^{\min(l+4,7)} \alpha_{l'}\right)\right) + \sum_{l=2}^{7} \alpha_l * \alpha_4 + \sum_{l=3}^{7} \alpha_l * \alpha_3 + \sum_{l=4}^{7} \alpha_l * \alpha_2 + \sum_{l=5}^{7} \alpha_l * \alpha_1$
6	$\sum_{l=2}^{7}\left(\alpha_l * \left(\sum_{l'=7}^{\min(l+5,7)} \alpha_{l'}\right)\right) + \sum_{l=2}^{7} \alpha_l * \alpha_5 + \sum_{l=3}^{7} \alpha_l * \alpha_4 + \sum_{l=4}^{7} \alpha_l * \alpha_3 + \sum_{l=5}^{7} \alpha_l * \alpha_2 + \sum_{l=6}^{7} \alpha_l * \alpha_1$
7	$\sum_{l=2}^{7} \alpha_l * \alpha_6 + \sum_{l=3}^{7} \alpha_l * \alpha_5 + \sum_{l=4}^{7} \alpha_l * \alpha_4 + \sum_{l=5}^{7} \alpha_l * \alpha_3 + \sum_{l=6}^{7} \alpha_l * \alpha_2 + \sum_{l=7}^{7} \alpha_l * \alpha_1$

The average number of paternal half-sisters, $\bar{n}_{phs}(i)$ and the average number of records for paternal half-sisters, $\bar{m}_{phs}(i)$: The number of daughters per sire (and consequently the number of paternal half-sisters per target individual) depended on sire status. For AI sires, the number of daughters per sire depended on decisional variables (input parameters) related to breeding program organization (number of progeny needed for progeny testing, number of AI sires given the level of AI use in the breeding program and genetic variability management, and so on). For those sires, the KPR had no effect. For natural mating sires, the number of known daughters depended on both zootechnical considerations (such as the number of dams mated to a male) and the KPR (the proportion of sires of selected newborn females that was known). To provide an overview of the differences, we have detailed below how these numbers were computed for rank1 females given the sire status.

Rank1-females born from an unknown sire (NoAI, AI_NoPT and AI_PT programs): by definition, $\bar{n}_{phs}(i) = 0$ and $\bar{m}_{phs}(i) = 0$.

Rank1-females born from the first breeding of a natural mating (No_AI, AI_NoPT and AI_PT programs)

The number of paternal half-sisters is equal to the number of known daughters per sire minus one (the TI itself). For natural mating sires, we assumed that daughters were assigned using markers. Assignment was performed within a sub-flock (assigned sub-flock). The size of the assigned sub-flock allowed a minimum (resp. maximum) number of fertile dams to be mated to the sire equal to $Das_{min} = 15$ (resp. $Das_{max} = 36$). We assumed that there were no difference between the mean EBV of sires used within assigned sub-flocks and the mean EBV of sires used within unassigned sub-flocks. We also assumed that there were no differences between the mean EBV of the females belonging to each sub-flock. As KPR increased and the number of sires available for the entire flock remained constant, the number of daughters selected and assigned per sire tends to a maximum value (i.e. the number of dams belonging to the unassigned sub-flock decreased). We opted for this way of computing to avoid underestimating the accuracy of the EBV for females born from a known sire when KPR is low.

Given a KPR, $Das = \max(Das_{min}, \frac{KPR - PropF_{AI}}{1 - PropF_{AI}} * Das_{max})$

The number of known daughters per natural mating sire (ND_{NM}), depended on dam prolificacy ($prolif_{NM}$), survival of their progeny, sex-ratio and the proportion selected given functional abilities ($s_{nb} * sex\ ratio * r_{Farn}$), survival from birth+1 to birth +2 (first rank) and fertility of newborn female selected ($f * s_f$).

$ND_{NM} = Das * prolif_{NM} * s_{nb} * sex\ ratio * r_{Farn} * f * s_f$

Natural sires had a constant number of daughters per reproduction cycle, ND_{NM}, during their productive life and a maximum reproductive life of 3 cycles.

The average number of paternal half-sisters $\bar{n}_{phs}(i)$ for rank1 females was:

$\bar{n}_{phs} = ND_{NM} - 1$

By construction all of these paternal half-sisters had one record when TI was of rank 1.

$\bar{m}_{phs}(i) = 1$

To determine $\bar{n}_{phs}(i)$ for rank2 females, we included the probability that the sire was still present.

TI Rank	Number of new paternal half-sisters $\bar{n}_{phs}$
2	$ND_{NM} * s_m$
3	$ND_{NM} * s_m^2$
4	Sires have been culled

To determine $\bar{m}_{phs}$(i), we computed the total number of records divided by the total number of new paternal half-sisters. For example, when TI was of rank 2, we included the first records of newly selected paternal half-sisters ($ND_{NM} * s_m$) and second records of previously selected paternal half-sisters that were still present and fertile ($(ND_{NM} - 1) * f * s_f$).

TI Rank	Average number of records for paternal half-sisters: $\bar{m}_{phs}$
2	$\frac{(ND_{NM}-1)*(1+f*s_f)+ND_{NM}*s_m}{ND_{NM}*(1+s_m)-1}$
3	$\frac{(ND_{NM}-1)*\sum_{i=0}^{2}(f*s_f)^i+ND_{NM}*s_m*(1+f*s_f)+ND_{NM}*s_m^2}{ND_{NM}*(1+s_m+s_m^2)-1}$
4	$\frac{(ND_{NM}-1)*\sum_{i=0}^{3}(f*s_f)^i+ND_{NM}*s_m*\sum_{i=0}^{2}(f*s_f)^i+ND_{NM}*s_m^2*(1+f*s_f)}{ND_{NM}*(1+s_m+s_m^2)-1}$
5	$\frac{(ND_{NM}-1)*\sum_{i=0}^{4}(f*s_f)^i+ND_{NM}*s_m*\sum_{i=0}^{3}(f*s_f)^i+ND_{NM}*s_m^2*\sum_{i=0}^{2}(f*s_f)^i}{ND_{NM}*(1+s_m+s_m^2)-1}$
6	$\frac{(ND_{NM}-1)*\sum_{i=0}^{5}(f*s_f)^i+ND_{NM}*s_m*\sum_{i=0}^{4}(f*s_f)^i+ND_{NM}*s_m^2*\sum_{i=0}^{3}(f*s_f)^i}{ND_{NM}*(1+s_m+s_m^2)-1}$
7	$\frac{(ND_{NM}-1)*\sum_{i=0}^{6}(f*s_f)^i+ND_{NM}*s_m*\sum_{i=0}^{5}(f*s_f)^i+ND_{NM}*s_m^2*\sum_{i=0}^{4}(f*s_f)^i}{ND_{NM}*(1+s_m+s_m^2)-1}$

Rank1-females born from the second breeding of a natural mating (NoAI, AI_NoPT and AI_PT programs)

$\bar{n}_{phs} = 2ND_{NM} - 1$

$\bar{m}_{phs} = \frac{ND_{NM}*(1+f*s_f)+(ND_{NM}-1)}{2ND_{NM}-1}$

TI Rank	Number of new paternal half-sisters $\bar{n}_{phs}$
2	$ND_{NM} * s_m$
3	Sires have been culled

TI Rank	Average number of records for paternal half-sisters: $\bar{m}_{phs}$
2	$\frac{ND_{NM}*\sum_{i=0}^{2}(f*s_f)^i+(ND_{NM}-1)*(1+f*s_f)+ND_{NM}*s_m}{ND_{NM}*(2+s_m)-1}$
3	$\frac{ND_{NM}*\sum_{i=0}^{3}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{2}(f*s_f)^i+(ND_{NM}*s_m)*\sum_{i=0}^{1}(f*s_f)^i}{ND_{NM}*(2+s_m)-1}$
4	$\frac{ND_{NM}*\sum_{i=0}^{4}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{3}(f*s_f)^i+(ND_{NM}*s_m)*\sum_{i=0}^{2}(f*s_f)^i}{ND_{NM}*(2+s_m)-1}$
5	$\frac{ND_{NM}*\sum_{i=0}^{5}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{4}(f*s_f)^i+(ND_{NM}*s_m)*\sum_{i=0}^{3}(f*s_f)^i}{ND_{NM}*(2+s_m)-1}$
6	$\frac{ND_{NM}*\sum_{i=0}^{6}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{5}(f*s_f)^i+(ND_{NM}*s_m)*\sum_{i=0}^{4}(f*s_f)^i}{ND_{NM}*(2+s_m)-1}$
7	$\frac{ND_{NM}*\sum_{i=0}^{6}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{6}(f*s_f)^i+(ND_{NM}*s_m)*\sum_{i=0}^{5}(f*s_f)^i}{ND_{NM}*(2+s_m)-1}$

<u>Rank1-females born from the third breeding of a natural mating sire (NoAI, AI_NoPT and AI_PT programs)</u>

$\bar{n}_{phs} = 3 * ND_{NM} - 1$

$\bar{m}_{phs} = \frac{ND_{NM}*\sum_{i=0}^{2}(f*s_f)^i+ND_{NM}*(1+f*s_f)+(ND_{NM}-1)}{3ND_{NM}-1}$

TI Rank	Number of new paternal half-sisters $\bar{n}_{phs}$
2	Sires have been culled

TI Rank	Average number of records for paternal half-sisters: $\bar{m}_{phs}$
2	$\frac{ND_{NM}*(f*s_f)^3+ND_{NM}*\sum_{i=0}^{2}(f*s_f)^i+(ND_{NM}-1)*f*s_f}{3ND_{NM}-1}$
3	$\frac{ND_{NM}*\sum_{i=0}^{4}(f*s_f)^i+ND_{NM}*\sum_{i=0}^{3}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{2}(f*s_f)^i}{3ND_{NM}-1}$
4	$\frac{ND_{NM}*\sum_{i=0}^{5}(f*s_f)^i+ND_{NM}*\sum_{i=0}^{4}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{3}(f*s_f)^i}{3ND_{NM}-1}$
5	$\frac{ND_{NM}*\sum_{i=0}^{6}(f*s_f)^i+ND_{NM}*\sum_{i=0}^{5}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{4}(f*s_f)^i}{3ND_{NM}-1}$
6	$\frac{2*ND_{NM}*\sum_{i=0}^{6}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{5}(f*s_f)^i}{3ND_{NM}-1}$
7	$\frac{2*ND_{NM}*\sum_{i=0}^{6}(f*s_f)^i+(ND_{NM}-1)*\sum_{i=0}^{6}(f*s_f)^i}{3ND_{NM}-1}$

Rank1-females born from the first breeding of an AI sire (AI_NoPT).

The number of daughters selected per sire at each breeding, ND_{AI}, was an internal parameter which depended on the number of AI sires selected per year, M_{AI}, and the rate of selected newborn females born from an AI sire, $PropF_{AI}$. Both M_{AI} and $PropF_{AI}$. were input parameters. Because AI sires were bred twice in the AI_NoPT program, the total number of AI sires was $M_{AI}(1 + s_m)$.

$ND_{AI} = (F * PropF_{AI} * \alpha_1)/(M_{AI}(1 + s_m))$

$\bar{n}_{phs}(i) = ND_{AI} - 1$

$\bar{m}_{phs}(i) = 1$

TI Rank	Number of new paternal half-sisters $\bar{n}_{phs}$
2	$ND_{AI} * s_m$
3	Sires have been culled

To determine $\bar{m}_{phs}$(i), we computed the total number of records divided by the total number of new paternal half-sisters. For example, for a TI of rank 2, we included the first records of newly selected paternal half-sisters ($ND_{AI} * s_m$) and the second records of previously selected paternal half-sisters that were still present and fertile ($(ND_{AI} - 1) * f * s_f$).

TI Rank	Average number of records for paternal half-sisters: $\bar{m}_{phs}$
2	$\frac{(ND_{AI}-1)*(1+f*s_f)+ND_{AI}*s_m}{ND_{AI}(1+s_m)-1}$
3	$\frac{(ND_{AI}-1)*\sum_{i=0}^{2}(f*s_f)^i+(ND_{AI}*s_m)*(1+f*s_f)}{ND_{AI}(1+s_m)-1}$
4	$\frac{(ND_{AI}-1)*\sum_{i=0}^{3}(f*s_f)^i+(ND_{AI}*s_m)*\sum_{i=0}^{2}(f*s_f)^i}{ND_{AI}(1+s_m)-1}$
5	$\frac{(ND_{AI}-1)*\sum_{i=0}^{4}(f*s_f)^i+(ND_{AI}*s_m)*\sum_{i=0}^{3}(f*s_f)^i}{ND_{AI}(1+s_m)-1}$
6	$\frac{(ND_{AI}-1)*\sum_{i=0}^{5}(f*s_f)^i+(ND_{AI}*s_m)*\sum_{i=0}^{4}(f*s_f)^i}{ND_{AI}(1+s_m)-1}$
7	$\frac{(ND_{AI}-1)*\sum_{i=0}^{6}(f*s_f)^i+(ND_{AI}*s_m)*\sum_{i=0}^{5}(f*s_f)^i}{ND_{AI}(1+s_m)-1}$

Rank1-females born from the second breeding of an AI sire (AI_NoPT).

$\bar{n}_{phs} = 2 * ND_{AI} - 1$

$\bar{m}_{phs} = \frac{ND_{AI}*(1+f*s_f)+(ND_{AI}-1)}{2*ND_{AI}-1}$

TI Rank	Number of new paternal half-sisters $\bar{n}_{phs}$
2	Sires have been culled

TI Rank	Average number of records for paternal half-sisters: $\bar{m}_{phs}$
2	$\frac{ND_{AI}*\sum_{i=0}^{2}(f*s_f)^i+(ND_{AI}-1)*(1+f*s_f)}{2*ND_{AI}-1}$
3	$\frac{ND_{AI}*\sum_{i=0}^{3}(f*s_f)^i+(ND_{AI}-1)*\sum_{i=0}^{2}(f*s_f)^i}{2*ND_{AI}-1}$
4	$\frac{ND_{AI}*\sum_{i=0}^{4}(f*s_f)^i+(ND_{AI}-1)*\sum_{i=0}^{3}(f*s_f)^i}{2*ND_{AI}-1}$
5	$\frac{ND_{AI}*\sum_{i=0}^{5}(f*s_f)^i+(ND_{AI}-1)*\sum_{i=0}^{4}(f*s_f)^i}{2*ND_{AI}-1}$
6	$\frac{ND_{AI}*\sum_{i=0}^{6}(f*s_f)^i+(ND_{AI}-1)*\sum_{i=0}^{5}(f*s_f)^i}{2*ND_{AI}-1}$
7	$\frac{ND_{AI}*\sum_{i=0}^{6}(f*s_f)^i+(ND_{AI}-1)*\sum_{i=0}^{6}(f*s_f)^i}{2*ND_{AI}-1}$

Rank1-females born from a male being progeny tested (AI_PT)

The number of daughters selected per male being progeny testes, ND_{PT}, was an input parameter.

$\bar{n}_{phs}(i) = ND_{PT} - 1$

$\bar{m}_{phs}(i) = 1$

For higher rank females born from a male being progeny tested, there was no increase in $\bar{n}_{phs}(i)$ until some of the males being progeny tested became proven sires. $\bar{m}_{phs}(i)$ depended on the new records available for paternal half-sisters that were still present in the population. From rank 4, the number of new paternal half-daughters depended on the selection intensity after progeny testing and the survival rate of rams.

TI Rank	Number of new paternal half-sisters n_{phs}
2	0
3	0
4	$ND_{PV} * tst * s_m$
5	$ND_{PV} * tst * s_m^2$
6	$ND_{PV} * tst * s_m^3$
7	0

With ND_{PV} the number of daughters selected per proven sire (see details in following section)

To determine $\bar{m}_{phs}$(i), we computed the total number of records divided by the total number of new paternal half-sisters. For example, for a TI of rank 4, we included the first records of newly selected paternal half-sisters ($ND_{PV} * tst * s_m$) and the total records (from birth(TI) to birth(TI)+4) of previously selected paternal half-sisters (($ND_{PT} - 1) * \sum_{i=0}^{3}(f * s_f)^i$).

TI Rank	Average number of records for paternal half-sisters m_{phs}
2	$1 + f * s_f$
3	$\sum_{i=0}^{2}(f * s_f)^i$
4	$\frac{(ND_{PT}-1)*\sum_{i=0}^{3}(f*s_f)^i + ND_{PV}*tst*s_m}{(ND_{PT}-1)+ND_{PV}*tst*s_m}$
5	$\frac{(ND_{PT}-1)*\sum_{i=0}^{4}(f*s_f)^i + (ND_{PV}*tst*s_m)*(1+f*s_f)+(ND_{PV}*tst*s_m^2)}{(ND_{PT}-1)+ND_{PV}*tst*(s_m+s_m^2)}$
6	$\frac{(ND_{PT}-1)*\sum_{i=0}^{5}(f*s_f)^i + (ND_{PV}*tst*s_m)*\sum_{i=0}^{2}(f*s_f)^i+(ND_{PV}*tst*s_m^2)*(1+f*s_f)+(ND_{PV}*tst*s_m^3)}{(ND_{PT}-1)+ND_{PV}*tst*(s_m+s_m^2+s_m^3)}$
7	$\frac{(ND_{PT}-1)*\sum_{i=0}^{6}(f*s_f)^i + (ND_{PV}*tst*s_m)*\sum_{i=0}^{3}(f*s_f)^i+(ND_{PV}*tst*s_m^2)*\sum_{i=0}^{2}(f*s_f)^i+(ND_{PV}*tst*s_m^3)*(1+f*s_f)}{(ND_{PT}-1)+ND_{PV}*tst*(s_m+s_m^2+s_m^3)}$

Rank1-females born from the first breeding of a proven AI sire (AI_PT)

The number of daughters selected per proven sire, ND_{PV}, depends on five input parameters: the proportion of selected newborn females (nucleus) born from an AI sire, $PropF_{AI}$, the number of progeny-tested males per cycle, M_{PT}; the number of daughters selected per progeny-tested sire, ND_{PT}; the selection intensity of progeny-tested sires, tst; and the survival rate of sires, s_m.

We assumed that the total number of selected newborn females born from an AI sire was constant ($F * \alpha_1 * PropF_{AI}$). A proportion of these AI females was allocated to progeny testing ($PropF_{AIPT} = (ND_{PT} * M_{PT})/(F * \alpha_1 * PropF_{AI})$). We assumed that proven sires were bred three times at the most and produced a constant number of doses during their productive life. Given tst, we inferred the number of proven sires being bred for the 1st time as ($M_{PT} * tst * s_m$). The number of daughters selected per proven sire was:

$$ND_{PV} = \lambda_{(1)} * F * PropF_{AI} * (1 - PropF_{AIPT}) * \alpha_1/(M_{PT} * tst * s_m)$$

With $\lambda_{(1)}$ the proportion of newborn females from proven sires being bred for the first time.

($\lambda_{(y)}$=, the proportion of sires being bred for the yth time).

For proven sires; the already known daughters born from progeny-testing had a maximum of 4 records. New daughters had one record:

$$\bar{n}_{phs}(i) = ND_{PT} + (ND_{PV} - 1)$$

$$\bar{m}_{phs}(i) = \frac{ND_{PT} * \sum_{i=0}^{3} (f * s_f)^i + (ND_{PV} - 1)}{ND_{PT} + (ND_{PV} - 1)}$$

For higher rank females born from progeny-tested sires, $\bar{n}_{phs}(i)$ depended on the number of females per proven sire (ND_{PV}), which was constant during the reproductive life of the sire, and the survival rate of sires between two cycles (s_m).

TI Rank	Number of new paternal half-sisters n_{phs}
2	$ND_{PV} * s_m$
3	$ND_{PV} * s_m^2$
4	Sires have been culled

To determine $\bar{m}_{phs}$(i), we computed the total number of records divided by the total number of new paternal half-sisters. For example, for a TI of rank 2, we included the first records for newly selected paternal half-sisters ($s_m * ND_{PV}$), the second records for previously selected paternal half-sisters that were still present and fertile ($(ND_{PV} - 1) * f * s_f$), and the new records for females born during the progeny-testing of their sire ($ND_{PT} * (f * s_f)^4$).

TI Rank	Average number of records for paternal half-sisters: m_{phs}
2	$\frac{ND_{PT} * \sum_{i=0}^{4} (f * s_f)^i + (ND_{PV} - 1) * (1 + f * s_f) + s_m * ND_{PV}}{ND_{PT} + (1 + s_m) * ND_{PV} - 1}$
3	$\frac{ND_{PT} * \sum_{i=0}^{5} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{2} (f * s_f)^i + (s_m * ND_{PV}) * (1 + f * s_f) + (s_m^2 * ND_{PV})}{ND_{PT} + (1 + s_m + s_m^2) * ND_{PV} - 1}$
4	$\frac{ND_{PT} * \sum_{i=0}^{6} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{3} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{2} (f * s_f)^i + (s_m^2 * ND_{PV}) * (1 + f * s_f)}{ND_{PT} + (1 + s_m + s_m^2) * ND_{PV} - 1}$
5	$\frac{ND_{PT} * \sum_{i=0}^{6} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{4} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{3} (f * s_f)^i + (s_m^2 * ND_{PV}) * \sum_{i=0}^{2} (f * s_f)^i}{ND_{PT} + (1 + s_m + s_m^2) * ND_{PV} - 1}$
6	$\frac{ND_{PT} * \sum_{i=0}^{6} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{5} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{4} (f * s_f)^i + (s_m^2 * ND_{PV}) * \sum_{i=0}^{3} (f * s_f)^i}{ND_{PT} + (1 + s_m + s_m^2) * ND_{PV} - 1}$
7	$\frac{(ND_{PT} + ND_{PV} - 1)) * \sum_{i=0}^{6} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{5} (f * s_f)^i + (s_m^2 * ND_{PV}) * \sum_{i=0}^{4} (f * s_f)^i}{ND_{PT} + (1 + s_m + s_m^2) * ND_{PV} - 1}$

Rank1-females born from the second breeding of a proven AI sire (AI_PT)

$$\bar{n}_{phs} = ND_{PT} + 2 * ND_{PV} - 1$$

$$\bar{m}_{phs} = \frac{ND_{PT} * \sum_{i=0}^{4} (f * s_f)^i + ND_{PV} * (1 + f * s_f) + (ND_{PV} - 1)}{ND_{PT} + 2 * ND_{PV} - 1}$$

TI Rank	Number of new paternal half-sisters $\bar{n}_{phs}$
2	$ND_{PV} * s_m$
3	Sires have been culled

TI Rank	Average number of records for paternal half-sisters: $\bar{m}_{phs}$
2	$\frac{ND_{PT} * \sum_{i=0}^{5} (f * s_f)^i + ND_{PV} * \sum_{i=0}^{2} (f * s_f)^i + (ND_{PV} - 1) * (1 + f * s_f) + (s_m * ND_{PV})}{ND_{PT} + (2 + s_m) * ND_{PV} - 1}$
3	$\frac{ND_{PT} * \sum_{i=0}^{6} (f * s_f)^i + (ND_{PV}) * \sum_{i=0}^{3} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{2} (f * s_f)^i + (s_m * ND_{PV}) * (1 + f * s_f)}{ND_{PT} + (2 + s_m) * ND_{PV} - 1}$
4	$\frac{ND_{PT} * \sum_{i=0}^{6} (f * s_f)^i + (ND_{PV}) * \sum_{i=0}^{4} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{3} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{2} (f * s_f)^i}{ND_{PT} + (2 + s_m) * ND_{PV} - 1}$
5	$\frac{ND_{PT} * \sum_{i=0}^{6} (f * s_f)^i + (ND_{PV}) * \sum_{i=0}^{5} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{4} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{3} (f * s_f)^i}{ND_{PT} + (2 + s_m) * ND_{PV} - 1}$
6	$\frac{(ND_{PT} + ND_{PV}) * \sum_{i=0}^{6} (f * s_f)^i + (ND_{PV} - 1) * \sum_{i=0}^{5} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{4} (f * s_f)^i}{ND_{PT} + (2 + s_m) * ND_{PV} - 1}$
7	$\frac{(ND_{PT} + 2ND_{PV} - 1) * \sum_{i=0}^{6} (f * s_f)^i + (s_m * ND_{PV}) * \sum_{i=0}^{5} (f * s_f)^i}{ND_{PT} + (2 + s_m) * ND_{PV} - 1}$

Rank1-females born from the third breeding of a proven AI sire (AI_PT)

$$\bar{n}_{phs} = ND_{PT} + 3 * ND_{PV} - 1$$

$$\bar{m}_{phs} = \frac{ND_{PT} * \sum_{i=0}^{5} (f * s_f)^i + ND_{PV} * \sum_{i=0}^{2} (f * s_f)^i + (ND_{PV}) * (1 + f * s_f) + (ND_{PV} - 1)}{ND_{PT} + 3 * ND_{PV} - 1}$$

Rank	Number of new paternal half-sisters $\bar{n}_{phs}$
2	Sires have been culled

Rank	New records of paternal half-sisters $\bar{m}_{phs}$
2	$\frac{ND_{PT}*\sum_{i=0}^{6}(f*s_f)^i + ND_{PV}*\sum_{i=0}^{3}(f*s_f)^i + ND_{PV}*\sum_{i=0}^{2}(f*s_f)^i + (ND_{PV}-1)*(1+f*s_f)}{ND_{PT}+3*ND_{PV}-1}$
3	$\frac{ND_{PT}*\sum_{i=0}^{6}(f*s_f)^i + ND_{PV}*\sum_{i=0}^{4}(f*s_f)^i + ND_{PV}*\sum_{i=0}^{3}(f*s_f)^i + (ND_{PV}-1)*\sum_{i=0}^{2}(f*s_f)^i}{ND_{PT}+3*ND_{PV}-1}$
4	$\frac{ND_{PT}*\sum_{i=0}^{6}(f*s_f)^i + ND_{PV}*\sum_{i=0}^{5}(f*s_f)^i + ND_{PV}*\sum_{i=0}^{4}(f*s_f)^i + (ND_{PV}-1)*\sum_{i=0}^{3}(f*s_f)^i}{ND_{PT}+3*ND_{PV}-1}$
5	$\frac{(ND_{PT}+ND_{PV})*\sum_{i=0}^{6}(f*s_f)^i + ND_{PV}*\sum_{i=0}^{5}(f*s_f)^i + (ND_{PV}-1)*\sum_{i=0}^{4}(f*s_f)^i}{ND_{PT}+3*ND_{PV}-1}$
6	$\frac{(ND_{PT}+2ND_{PV})*\sum_{i=0}^{6}(f*s_f)^i + (ND_{PV}-1)*\sum_{i=0}^{5}(f*s_f)^i}{ND_{PT}+3*ND_{PV}-1}$
7	$\frac{(ND_{PT}+3*ND_{PV}-1)*\sum_{i=0}^{6}(f*s_f)^i}{ND_{PT}+3*ND_{PV}-1}$

The average number of daughters, $\bar{n}_{da}(i)$, and their average number of records, $\bar{m}_f(i)$:

Given the lag between birth and the first parity (rank) the average number of daughters per reproductive female was low. Because we did not implement a selection differential along the dam-female path, there was no difference in contribution between the female categories. Hence, the average number of daughters per female was neglected.

INFORMATION CONSIDERED FOR A MALE TARGET INDIVIDUAL (TI).

In the No_AI and AI_NoPT breeding programs, the sires used for breeding were not selected. In AI_PT programs, sires were selected at time = birth + 4. In addition to progeny records, other information on close relatives known at time = birth + 4 were used to compute the accuracy of the EBV for males. Except for paternal half-sisters and daughters, the amount of information considered depended on the contribution of each dam category to selected newborn males. The contribution of dam categories to selected newborn females depended on demographic parameters (α_l). The contribution of dam categories to selected newborn males depended on selection. $\chi_j(t)$ was the proportion of selected newborn males born from a dam belonging to the category j at time t. Computing of $\chi_j(t)$ is explained in the description of the genetic model (Dam male path). We denoted by $\chi_l = \sum_{j\in_l}\chi_j(t)$ the proportion of selected newborn males born from rank l dam categories.

We used the same reasoning as when computing female information. Consequently, details will not be provided. They are available on request to the corresponding authors.

The average number of records for dams, $\bar{m}_d(i)$:

As described previously for females, $\bar{\mathrm{m}}_{\mathrm{d}}(i)$, depended on the records available at the males' birth and new records that became available between their birth and the birth of their progeny, i.e. 4 reproduction cycles later.

$\bar{\mathrm{m}}_{\mathrm{d}}(i) = \sum_{l=1}^{7} \chi_l * (l) + \sum_{l=2}^{7} \chi_l + \sum_{l=3}^{7} \chi_l + \sum_{l=4}^{7} \chi_l + \sum_{l=5}^{7} \chi_l$

The average number of records for maternal grand-dams, $\bar{m}_{mgd}(i)$, and paternal grand-dams, $\bar{m}_{pgd}(i)$, were equal to the maximum average number of records per female because all maternal and paternal grand-dams were culled when males were being progeny tested.

$\bar{m}_{mgd}(i) = \bar{m}_{pgd}(i) = [(\sum_{l=1}^{7} \alpha_l * (l) + \sum_{l=2}^{7} \alpha_l + \sum_{l=3}^{7} \alpha_l + \sum_{l=4}^{7} \alpha_l + \sum_{l=5}^{7} \alpha_l + \sum_{l=6}^{7} \alpha_l + \sum_{l=7}^{7} \alpha_l)]$

The average number of maternal half-sisters, $\bar{n}_{mhs}(i)$, was equal to the number of records for their dam multiplied by the probability of selecting a newborn female (α_1). We assumed that newborn females and males were selected independently.

The last records of dams (time = birth+3 and birth+4) were not included because their daughters had yet to reach rank 1.

$\bar{n}_{mhs}(i) = \sum_{l=1}^{7} \chi_l * (l) * \alpha_1 + (\sum_{l=2}^{7} \chi_l + \sum_{l=3}^{7} \chi_l) * \alpha_1$

The average number of records for maternal half-sisters, $\bar{m}_{mhs}(i)$: As for females, this quantity depended on the proportion of rank l dams (χ_l), the difference between the rank of the dam l and that of their daughter l' ($l - l' + 3$) and the probability that a maternal half-sister had a record given its rank ($\alpha_{l'}$).

$\bar{\mathrm{m}}_{mhs}(\mathrm{i}) = \sum_{l=1}^{7} \left(\chi_l * \left(\sum_{l'=1}^{min(l-l'+3,7)} \alpha_{l'} \right) \right)$

The average number of paternal half-sisters, $\bar{n}_{phs}(i)$, and the average number of records for paternal half-sisters, $\bar{m}_{phs}(i)$: selected newborn males were born from the 3 categories of proven sires. The relative contributions of sire categories to selected newborn males were input parameters as expressed in Table 2, section "Paternal origins of male candidates" (1st (40%), 2nd (30%) and 3rd breeding (30%)). $\bar{n}_{phs}(i)$ and $\bar{\mathrm{m}}_{phs}(\mathrm{i})$ corresponded to the weighted average given the contribution of these 3 sire categories. When the results were received for the males being progeny-tested, their sire was culled

whatever the sire category (proven sires were bred three times). Thus, all their paternal-half sisters were born and had at least one record.

In consequence, $\bar{n}_{phs}(i)$ depended on the number of daughters selected per progeny-tested sire (ND_{PT}), the number of daughters selected by proven sire (ND_{PV}) and the survival rate of males (s_m).

For example, for males being progeny tested born from the 1st breeding of a proven sire, $\bar{n}_{phs}(i) = ND_{PT} + ND_{PV} * (1 + s_m + s_m^2)$

Thus we computed $\bar{m}_{phs}(I)$ taking into account the potential number of records of each group of paternal half-sisters given their age: ND_{PT} females had a maximum of 6 records, ND_{PV}, born at time = birth (TI) have 3 records, $ND_{PV} * s_m$ born at time = birth(TI)+1 had a maximum of two records and so on.

Thus $\bar{m}_{phs}(I) = \frac{ND_{PT}*\left(1+f*s_f+(f*s_f)^2+(f*s_f)^3+(f*s_f)^4+(f*s_f)^5\right)+ND_{PV}*\left[\left(1+f*s_f+(f*s_f)^2\right)+s_m*(1+f*s_f)+s_m^2\right]}{ND_{PV}*(1+s_m+s_m^2)+ND_{PT}}$

Let $PROV_i$ denote proven sires in their i^{th} breeding cycle (as a proven male)

IT Sire status	$\bar{n}_{phs}$: Number of paternal half-sisters
$PROV_1$	$ND_{PV} * (1 + s_m + s_m^2) + ND_{PT}$
$PROV_2$	$ND_{PV} * (2 + s_m) + ND_{PT}$
$PROV_3$	$3 * ND_{PV} + ND_{PT}$

Sire status	$\bar{m}_{phs}$: average number of records for paternal half-sisters
$PROV_1$	$\frac{ND_{PT}*\left(1+f*s_f+(f*s_f)^2+(f*s_f)^3+(f*s_f)^4+(f*s_f)^5\right)+ND_{PV}*\left[\left(1+f*s+(f*s_f)^2\right)+s_m*(1+f*s_f)+s_m^2\right]}{ND_{PV}*(1+s_m+s_m^2)+ND_{PT}}$
$PROV_2$	$\frac{ND_{PT}*\left(1+f*s_f+(f*s_f)^2+(f*s_f)^3+(f*s_f)^4+(f*s_f)^5+(f*s_f)^6\right)+ND_{PV}*\left[\left(2+2f*s_f+2(f*s_f)^2+(f*s_f)^3\right)+s_m*\left(1+f*s_f\right)\right]}{ND_{PV}*(2+s_m)+ND_{PT}}$
$PROV_3$	$\frac{ND_{PT}*\left(1+f*s_f+\left(f*s_f\right)^2+\left(f*s_f\right)^3+\left(f*s_f\right)^4+\left(f*s_f\right)^5+\left(f*s_f\right)^6\right)+ND_{PV}*\left(3+3f*s_f+3\left(f*s_f\right)^2+2\left(f*s_f\right)^3+\left(f*s_f\right)^4\right)}{3*ND_{PV}+ND_{PT}}$

The general formulae used to compute the average known information is:

$\bar{n}_{phs} = \sum_{nm} p_{1,j} * \bar{n}_{phs}(j)$

$\bar{m}_{phs} = \sum_{nm} p_{1,j} * \bar{m}_{phs}(j)$

With $p_{1,j}$ $j \in [1, nm]$ the proportion of newborn males born from the sire category nm, and $\bar{n}_{phs}(j)$ and $\bar{m}_{phs}(j)$ the number of paternal half-sisters and the average number of records for paternal half-

sisters estimated for newborn males born from the sire category j. Note that $p_{1,j}$ $j \in [1, nm]$ corresponds to the paternal origins of male candidates indicated in Table 2.

The average number of daughters, $\bar{n}_{da}(i)$, and the average number of records for daughters, $\bar{m}_{da}(i)$: the number of daughters selected per male being progeny tested, ND_{PT}, was an input parameter.

$\bar{n}_{da}(i) = ND_{PT}$

$\bar{m}_{da}(i) = 1$

SUPPLEMENTARY MATERIALS-2: COMPUTATION OF THE AVERAGE ACCURACY OF EBV

The reliability was computed using:

$$r^2_{g\hat{g}} = \frac{cov\,(g,y) * \ var(y)^{-1} * cov\,(y,g)}{var(g)}$$

Where $cov\,(g,y)$ (and $cov\,(y,g)$) is the vector containing the covariance between the genetic value of individuals and known records of individuals and their close relatives, and $var(y)$ is the covariance-variance matrix. The inverse of the $var(y)$ matrix was numerically computed using the NAG FORTRAN library routine F01ADF which gives the approximate inverse of a real symmetric positive definite matrix using a Cholesky decomposition. Given our deterministic model, we assumed that all individuals belonging to a category have the mean number of records.

Elements of vectors y and $cov\,(g,y')$ for females

$y = (\overline{P_{mgd}}, \overline{P_{pgd}}, \overline{P_d}, \overline{P_p}, \frac{\sum_{phs}^{n_{phs}} \bar{P}_{phs}}{n_{phs}}, \frac{\sum_{mhs}^{n_{mhs}} \bar{P}_{mhs}}{n_{mhs}})$ with $\overline{P_i}$ the average performance of i.

$cov\,(g,\ y') = (\frac{1}{4}\sigma_g^2, \frac{1}{4}\sigma_g^2, \frac{1}{2}\sigma_g^2, \sigma_g^2, \frac{1}{4}\sigma_g^2, \frac{1}{4}\sigma_g^2)$

Elements of vectors y and $cov\,(g,y')$ for males

$y = (\overline{P_{mgd}}, \overline{P_{pgd}}, \overline{P_d}, \frac{\sum_{mhs}^{n_{mhs}} \bar{P}_{mhs}}{n_{mhs}}, \frac{\sum_{phs}^{n_{phs}} \bar{P}_{phs}}{n_{phs}}, \frac{\sum_{da}^{n_{da}} \bar{P}_{da}}{n_{da}})$ with $\overline{P_i}$ the average record for i.

$cov\,(g,\ y') = (\frac{1}{4}\sigma_g^2, \frac{1}{4}\sigma_g^2, \frac{1}{2}\sigma_g^2, \frac{1}{4}\sigma_g^2, \frac{1}{4}\sigma_g^2, \frac{1}{2}\sigma_g^2)$

Elements of the variance / covariance matrix (note that some elements are specific to males or females)

$$var\left(\overline{P_{mgd}}\right) = \frac{\sigma_p^2(1+(m_{mgd}-1)r)}{m_{mgd}}$$

$$var\left(\overline{P_{pgd}}\right) = \frac{\sigma_p^2(1+(m_{pgd}-1)r)}{m_{pgd}}$$

$$var\left(\overline{P_d}\right) = \frac{\sigma_p^2(1+(m_d-1)r)}{m_d}$$

$$var\left(\overline{P_p}\right) = \frac{\sigma_p^2(1+(m_p-1)r)}{m_p}$$

$$var\left(\frac{\sum_{phs}^{n_{phs}} \bar{P}_{phs}}{n_{phs}}\right) = \frac{\sigma_p^2}{n_{phs}}\left[\frac{((1+(m_{phs}-1)r))}{m_{phs}} + \frac{1}{4}(n_{phs}-1)h^2)\right]$$

$$var\left(\frac{\sum_{mhs}^{n_{mhs}} \bar{P}_{mhs}}{n_{mhs}}\right) = \frac{\sigma_p^2}{n_{mhs}}\left[\frac{((1+(m_{mhs}-1)r))}{m_{mhs}} + \frac{1}{4}(n_{mhs}-1)h^2)\right]$$

$$var\left(\frac{\sum_{da}^{n_{da}}\bar{P}_{da}}{n_{da}}\right)=\frac{\sigma_p^2}{n_{da}}\left[\frac{((1+(m_{da}-1)r))}{m_{da}}+\frac{1}{4}(n_{da}-1)h^2)\right]$$

$$cov(\overline{P_{mgd}},\overline{P_{pgd}})=cov\left(\overline{P_{mgd}},\frac{\sum_{phs}^{n_{phs}}\bar{P}_{phs}}{n_{phs}}\right)=cov(\overline{P_{pgd}},\overline{P_d})=cov\left(\overline{P_{pgd}},\frac{\sum_{mhs}^{n_{mhs}}\bar{P}_{mhs}}{n_{mhs}}\right)=$$

$$cov\left(\overline{P_d},\frac{\sum_{phs}^{n_{phs}}\bar{P}_{phs}}{n_{phs}}\right)=cov\left(\frac{\sum_{phs}^{n_{phs}}\bar{P}_{phs}}{n_{phs}},\frac{\sum_{mhs}^{n_{mhs}}\bar{P}_{mhs}}{n_{mhs}}\right)=0$$

$$cov(\overline{P_{mgd}},\overline{P_d})=cov\left(\overline{P_d},\bar{P}_p\right)=cov\left(\overline{P_d},\frac{\sum_{mhs}^{n_{mhs}}\bar{P}_{mhs}}{n_{mhs}}\right)=\frac{1}{2}\sigma_g^2$$

$$cov(\overline{P_{mgd}},\bar{P}_p)=cov\left(\overline{P_{mgd}},\frac{\sum_{mhs}^{n_{mhs}}\bar{P}_{mhs}}{n_{mhs}}\right)=cov(\overline{P_{pgd}},\bar{P}_p)=cov\left(\overline{P_{pgd}},\frac{\sum_{phs}^{n_{phs}}\bar{P}_{phs}}{n_{phs}}\right)=$$

$$cov\left(\bar{P}_p,\frac{\sum_{phs}^{n_{phs}}\bar{P}_{phs}}{n_{phs}}\right)=cov\left(\bar{P}_p,\frac{\sum_{mhs}^{n_{mhs}}\bar{P}_{mhs}}{n_{mhs}}\right)=\frac{1}{4}\sigma_g^2$$

$$cov\left(\overline{P_{mgd}},\frac{\sum_{da}^{n_{da}}\bar{P}_{da}}{n_{da}}\right)=cov\left(\overline{P_{pgd}},\frac{\sum_{da}^{n_{da}}\bar{P}_{da}}{n_{da}}\right)=cov\left(\frac{\sum_{phs}^{n_{phs}}\bar{P}_{phs}}{n_{phs}},\frac{\sum_{da}^{n_{da}}\bar{P}_{da}}{n_{da}}\right)=$$

$$cov\left(\frac{\sum_{mhs}^{n_{mhs}}\bar{P}_{mhs}}{n_{mhs}},\frac{\sum_{da}^{n_{da}}\bar{P}_{da}}{n_{da}}\right)=\frac{1}{8}\sigma_g^2$$

SUPPLEMENTARY MATERIALS-3: ELEMENTS OF THE TRANSITION MATRIX P

The transition matrix P describes the gene flow between categories after one reproduction cycle. P can be partitioned into 6 submatrices as:

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{31} & P_{33} \end{pmatrix}$$

The index 1 corresponded to male categories in the nucleus, index 2 to female categories in the nucleus and index 3 to male and female categories in the commercial population.
nm is the number of male categories and nf the number of female categories.

The block matrix $\boldsymbol{P_{11}}$ (nm, nm) describes the genetic contribution of male categories at t-1 to male categories at t. The elements $p_{1,j}$ j∈ $[1, nm]$ are the contributions of sire categories to the newborn male category. For example, if 70% of young males (row 1) were born from natural mating sires in their 1st use (column 3) and 30% from their 2nd use (column 4) all elements $p_{1,j}$ j∈ $[1, nm]$ would be equal to 0 except $p_{1,3} = 0.7 * 0.5 = 0.35$ and $p_{1,4} = 0.3 * 0.5 = 0.15$. The factor 0.5 means that only 50% of genes are inherited from sires ($\sum_{j=1}^{nm} p_{1,j} = 0.5$). The value of each genetic contribution is an input data of the model: $2 * p_{1,j}$ j∈ $[1, nm]$, the proportion of newborn males born from the sire category j corresponds to the paternal origins of male candidates indicated in Table 2. This value depends on the breeding program assessed. Other lines of this block matrix describe the time-dependent flow of genes between categories (population ageing). For example, males are progeny tested at cycle birth +3. All males that are in progeny testing (row 6) at time t were aged (birth +2) at time t-1. Thus $p_{6,2} = 1$.

$$P_{11} = \begin{pmatrix} 0 & 0 & p_{1,3} & p_{1,4} & p_{1,5} & p_{1,6} & 0 & p_{1,8} & p_{1,9} & p_{1,10} & p_{1,11} & p_{1,12} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Equivalence between category j and male categories (Fig. 1)	
1	Newborn male candidates
2	Young selected males
3	Natural mating sire, 1st breeding
4	Natural mating sire, 2nd breeding
5	Natural mating sire, 3rd breeding
6	Male being progeny tested (Birth+2)
7	Male being progeny tested (Birth+3)
8	Proven sire, 1st breeding
9	Proven sire, 2nd breeding
10	Proven sire, 3rd breeding
11	AI sire (not progeny tested), 1st breeding
12	AI sire (not progeny tested), 2nd breeding

The block matrix $\boldsymbol{P_{12}}$ (nm, nf) describes the genetic contribution of female categories at t-1 to male categories at t. Only the elements of the first row may be different from 0. The elements $p_{1,j} \quad j \in [nm+1, nm+nf]$ are the genetic contributions of dam categories to the newborn male category ($\sum_{j=nm+1}^{nm+nf} p_{1,j} = 0.5$). For a given class j, its genetic contribution is $0.5 * \chi_j(t)$, with $\chi_j(t) = \frac{q_{j,K}(t)}{r_{DM}}$ the proportion of selected newborn males born from a dam belonging to category j at time t ($\sum_{j=nm+1}^{nm+nf} \chi_j(t) = 1$), $q_{j,K}(t)$ the proportion of females (among females of the nucleus population) from category j selected as dams of sires given the single truncation threshold $K(t)$ at time t and r_{DM} the proportion of females selected as dams of sires.

The block matrix $\boldsymbol{P_{21}}$ (nf, nm) describes the genetic contribution of male categories at t-1 to female categories at t. This matrix block may be divided into 10 sub-matrices $(7, nm)$ corresponding to the 10 paternal origins used to define female categories. 7 is the maximum rank (parity) observed in the population. Only the elements of the first row of sub-matrices may be different from 0. Newborn male candidates, young selected males and males waiting for progeny testing results (at time = birth+3) were not used for reproduction. Thus, for $i > nm$, $p_{i,1} = p_{i,2} = p_{i,7} = 0$. Except for the 'unknown sire' female category, one element of the first line nb_c of sub-matrices is equal to 0.5: $p_{nb_c, j=c} = 0.5$, $p_{nb_c, j \neq c} = 0$. nb_c corresponds to newborn females born from a sire (= category) c. For the sub-matrix corresponding to unknown sires, we assumed that females were born from all available natural mating sires. Let λ (1), λ (2), λ (3) be the proportion of natural mating sires belonging to the categories NM

sires-1st breeding ($\lambda(1) = \frac{1}{\sum_{v=1}^{3}\lambda(v)}$), NM sires-2nd breeding ($\lambda(2) = \frac{s_m}{\sum_{v=1}^{3}\lambda(v)}$) and NM sires-3rd breeding ($\lambda(3) = \frac{{s_m}^2}{\sum_{v=1}^{3}\lambda(v)}$): $p_{nb_c,3} = 0.5 * \lambda(1)$, $p_{nb_c,4} = 0.5 * \lambda(2)$, $p_{nb_c,5} = 0.5 * \lambda(3)$ with c = 'unknown sire' category.

$$P_{21} = \begin{pmatrix} p_{(nm+1,1)} & p_{(nm+1,2)} & \cdots & \cdots & \cdots & p_{(nm+1,12)} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ p_{(nm+1+7,1)} & p_{(nm+1+7,2)} & \cdots & \cdots & \cdots & p_{(nm+1+7,12)} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ p_{(nm+1+2*7,1)} & p_{(nm+1+2*7,2)} & \cdots & \cdots & \cdots & p_{(nm+1+2*7,12)} \\ \vdots & \vdots & 0 & 0 & 0 & \vdots \\ \vdots & \vdots & 0 & 0 & 0 & \vdots \end{pmatrix}$$

As an example, the submatrix below describes the submatrix corresponding to the 'proven sire' category (c = 8)

$$(7, nm)_{c=8} = \begin{pmatrix} p_{(nb_8,1)} = 0 & 0 & \ldots & 0 & p_{(nb_8,8)} = 0.5 & 0 & \cdots \\ \vdots & 0 & \ldots & 0 & 0 & 0 & \cdots \\ p_{(nb_8+7,1)} = 0 & 0 & \ldots & 0 & 0 & 0 & \cdots \end{pmatrix}$$

The block matrix P_{22} (nf, nf) describes the genetic contributions of female categories at $t-1$ to female categories at t. We assumed that all female categories j contribute to newborn female categories depending on the number of females belonging to category j at time $e_{j,t}$. Thus, for any row nb_c corresponding to a newborn female category (${P^{22}}_{cc',c\neq c'}$ and ${P^{22}}_{cc',c=c'}$), the contribution to the newborn female category was equal to: $p_{i,j} = 0.5 * e_{j,t}/F$ with F the total female population ($\sum_j e_{j,t} = F$). ${P^{22}}_{cc'}$ is the sub-matrix describing the genetic contribution of female category c' to female category c. Elements of rows $nb_c + 1$ to $nb_c + a$ are equal to 0 for sub-matrices ${P^{22}}_{cc',c\neq c'}$. Elements equal to 1 into sub-matrices ${P^{22}}_{cc',c=c'}$ correspond to the ageing of the population.

$$P_{22} = \begin{bmatrix} {p^{22}}_{c_1c_1} & \cdots & {p^{22}}_{c_1c_{10}} \\ \vdots & \ddots & \vdots \\ {p^{22}}_{c_{10}c_1} & \cdots & {p^{22}}_{c_{10}c_{10}} \end{bmatrix}$$

$${p^{22}}_{cc',c=c'} = \begin{pmatrix} p_{nb_c,j} & p_{nb_c,j+1} & \cdots & \cdots & \cdots & \cdots & p_{nb_c,j+a-1} \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

$${p^{22}}_{cc',c\neq c'} = \begin{pmatrix} p_{nb_c,j} & p_{nb_c,j+1} & \cdots & \cdots & \cdots & \cdots & p_{nb_c,j+a-1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Block matrices P_{13} and P_{23} describe the genetic contributions of commercial population categories at t-1 to male and female categories of the nucleus at t. As we assumed these contributions were null, the elements of these two matrices were equal to 0.

Block matrix P_{33} describes the gene flow within the commercial population. The commercial population was divided into 5 male categories (as in the No_AI program) and 7 female categories. The genetic contribution of parental categories depended on male (tdm) and female (tdf) transfer from the nucleus to the commercial population.

Contributions of male and female categories (commercial population) to newborn females were affected by a coefficient $(1 - tdf)$
Contributions of male and female categories (commercial population) to newborn males were affected by a coefficient $(1 - tdm)$
Following a classical approach, some elements of this matrix were equal to 1 to describe the ageing of the population.

Block matrices P_{31} and P_{32} describe the genetic contributions of male and female categories of the nucleus at t-1 to male and female categories of the commercial population at t. Only elements of the first row of each matrix could be different from 0 and were affected by a coefficient tdm. Other elements were equal to 0.

Computing of coefficients tdm and tdf

The number of newborn females produced by the nucleus and available for transfer to the industry depended on the number of females mated to a natural mating sire ($\frac{F*\beta_{NM}}{f}$), their fertility (f), their prolificacy, the survival rate of lambs, the sex ratio (0.5), the losses due to non-genetic factors such as functional abilities and fertility (r_{Farc}), and the number of newborn females selected for the nucleus ($F\alpha_1 * (1 - PropF_{AI})$):

$\frac{F*\beta_{NM}}{f} * f * prolif_{NM} * s_{nb} * 0.5 * r_{Farc} - F\alpha_1 * (1 - PropF_{AI})$

$= F * \left(\beta_{NM} * prolif_{NM} * s_{nb} * 0.5 * r_{Farc} - \alpha_1 * (1 - PropF_{AI})\right)$

Note that F was the fertile female population, so $\frac{F}{f}$ was the total population.

A portion of the ewes mated to a natural mating sire were first unsuccessfully inseminated giving: $\beta_{mn} = 1 - \beta_{AI} * f_{AI}$.

The number of selected newborn females born from AI ($F * PropF_{AI} * \alpha_1$) and daughters of $\frac{\beta_{AI}*F*f_{AI}}{f}$ successfully inseminated females, resulted in the average birth of $prolif_{AI} * s_{nb} * 0.5 * r_{Farn}$ replacement ewe lambs for the nucleus.

Thus: $\frac{\beta_{AI}*F*f_{AI}*prolif_{AI}*s_{nb}*0.5*r_{Farn}}{f} = F * PropF_{AI} * \alpha_1$ giving $\beta_{AI} = \frac{PropF_{AI}*\alpha_1*f}{f_{AI}*prolif_{AI}*s_{nb}*0.5*r_{Farn}}$

As we assumed that the proportion of females of parity l in the commercial population and nucleus were the same, the number of newborn females needed for the commercial population was:

$F * \alpha_1 * ps$

with ps the relative size of the commercial population.

It was assumed that no differential selection (positive or negative) was applied to the newborn females transferred from the nucleus to the industry. When the potential of the nucleus was higher than the needs of the commercial population, the proportion of commercial newborn females born in the nucleus, tdf, was equal to 1, giving

$tdf = MIN\left(\frac{\beta_{NM}*f*prolif_{NM}*s_{nb}*0.5*r_{Farc}-\alpha_1(1-PropF_{AI})}{\alpha_1*ps}, 1\right)$

When genetic progress was transferred via males, the proportion of new reproducers in commercial flocks born in the nucleus, tdm, was the sum of newborn males born from AI sires (tdm_{AI}) and natural mating sires (tdm_{NM}). It was supposed that, for transfer purposes, preference was given to newborn males born from AI sires. Following the approach developed for the transfer of females (ratio of candidates by needs), we computed tdm_{AI} and tdm_{NM} .

The number of newborn males needed by the commercial population was $\frac{\mathrm{F}*ps}{f*r_{F/M}*dum_c}$,

with $r_{F/M}$ the number of females mated per male and dum_c, the average number of reproduction cycles per ram in the commercial population.

The number of newborn males born from AI sires available for transfer to commercial flocks was:

$\frac{F*\beta_{AI}}{f} * f_{AI} * prolif_{AI} * s_{nb} * 0.5 * r_{Marc}$

With r_{Marc} the proportion of newborn males available for transfer. We neglected the proportion of newborn males selected for the nucleus. As for females, we considered that no differential selection (positive or negative) was applied to newborn males transferred to the industry. Thus we computed tdm_{AI} as:

$$tdm_{AI} = MIN\left(\frac{r_{F/M}*\beta_{AI}*f_{AI}*prolif_{AI}*s_{nb}*r_{Marc}*0.5*dum_c}{ps}, 1\right)$$

And following the same reasoning:

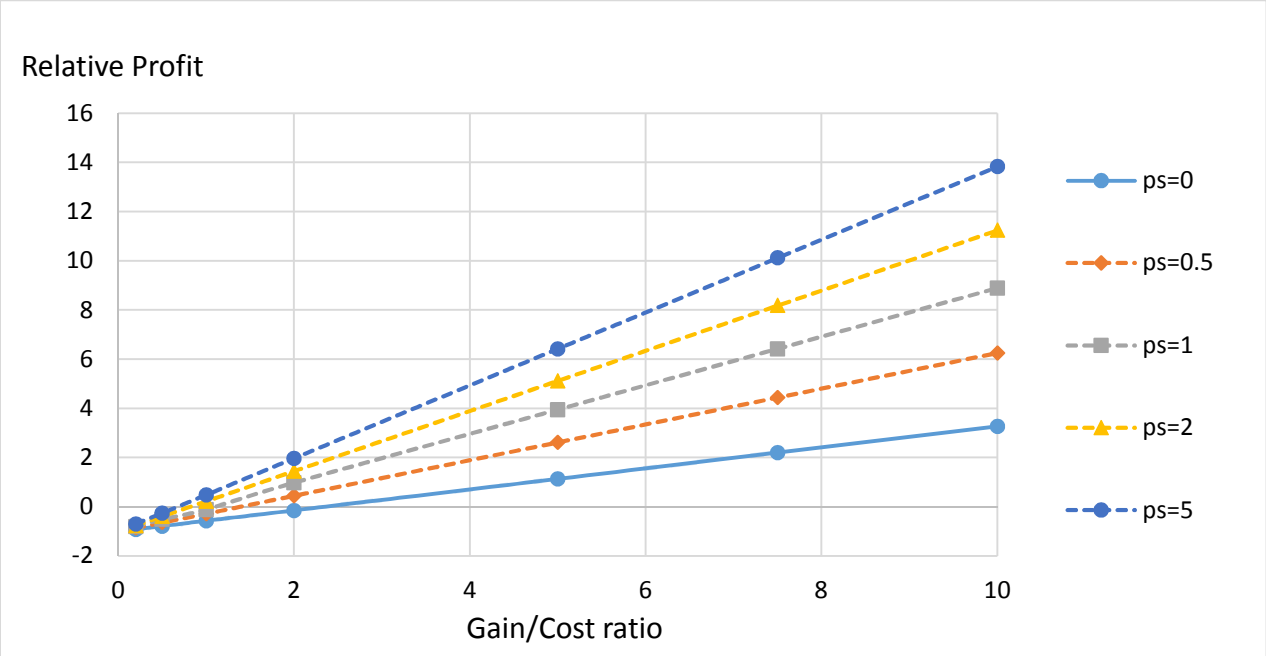
$$tdm_{NM} = MIN\left(\frac{r_{F/M}*\beta_{NM}*f*prolif_{NM}*s_{nb}*r_{Marc}*0.5*dum_c}{ps}, 1 - tdm_{AI}\right)$$

SUPPLEMENTARY RESULTS:

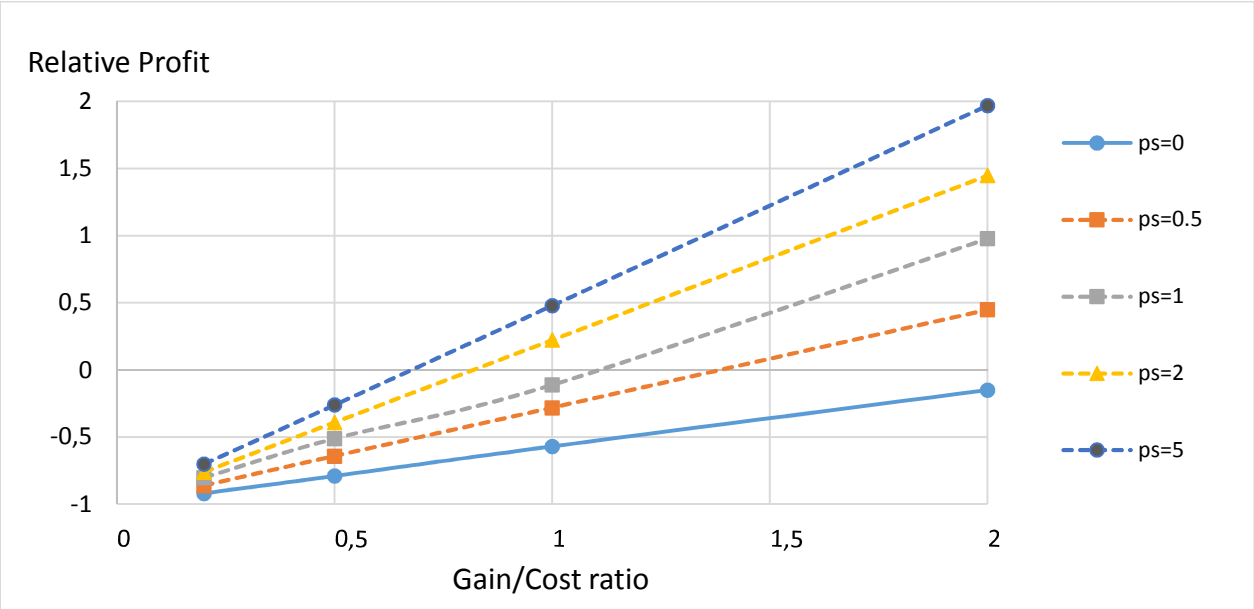
No_AI program

Figure SR1: Relative profit resulting from an increase of the female known paternity rate for various gain/cost ratios and relative sizes of the commercial population. Results for the No_AI program (males born from natural mating sires) when females are used to transfer the genetic progress. a: Full range of gain/cost ratios considered. b: Limited range of gain/cost ratios [0.2,2]. We considered an investment period of 60 reproduction cycles. Each curve corresponds to a given population size. ps is the relative size of the commercial population compared to the nucleus size (nucleus size $= 1$). The gain/cost ratio is defined as the monetary gain due to an extra genetic standard deviation (MG) divided by the cost of one assignment (Cp).

(a)



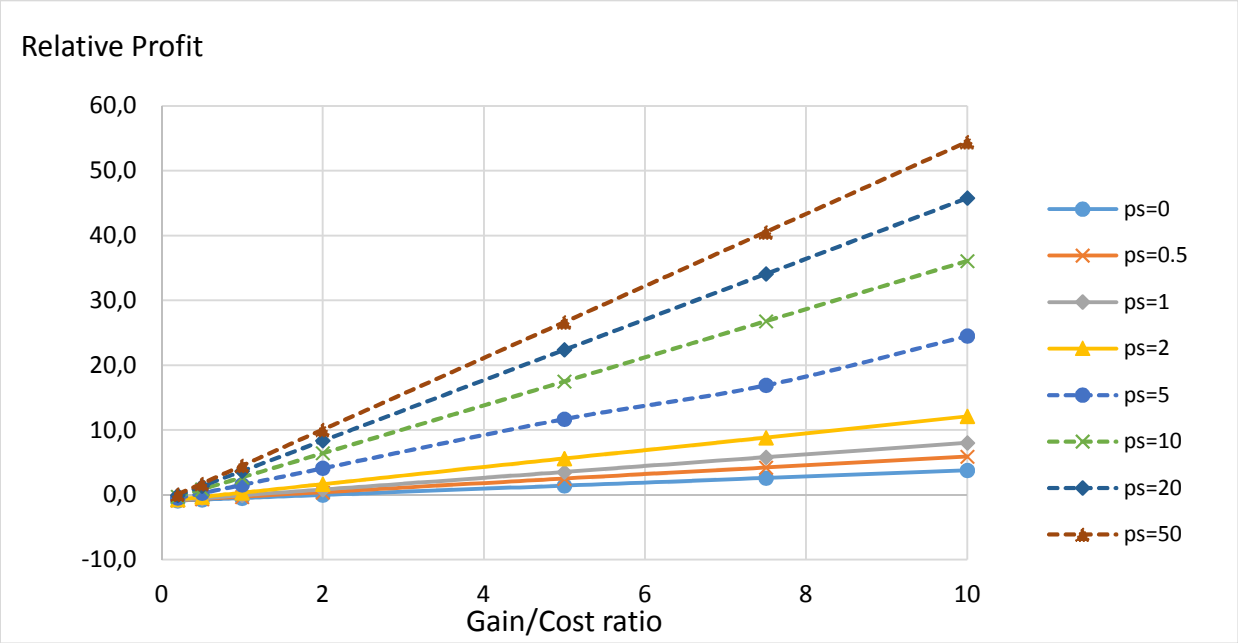
(b)



AI_NoPT program

Figure SR2: Relative profit resulting from an increase of the female known paternity rate for various gain/cost ratios and relative sizes of the commercial population. Results for the AI_NoPT program (males born from non-progeny-tested AI sires) when males are used to transfer the genetic progress. a: Full range of gain/cost ratios considered. b: Limited range of gain/cost ratios [0.2,2]. We considered an investment period of 60 reproduction cycles. Each curve corresponds to a given population size. ps is the relative size of the commercial population compared to the nucleus size (nucleus size = 1). The gain/cost ratio is defined as the monetary gain due to an extra genetic standard deviation (MG) divided by the cost of one assignment (Cp).

(a)



(b)

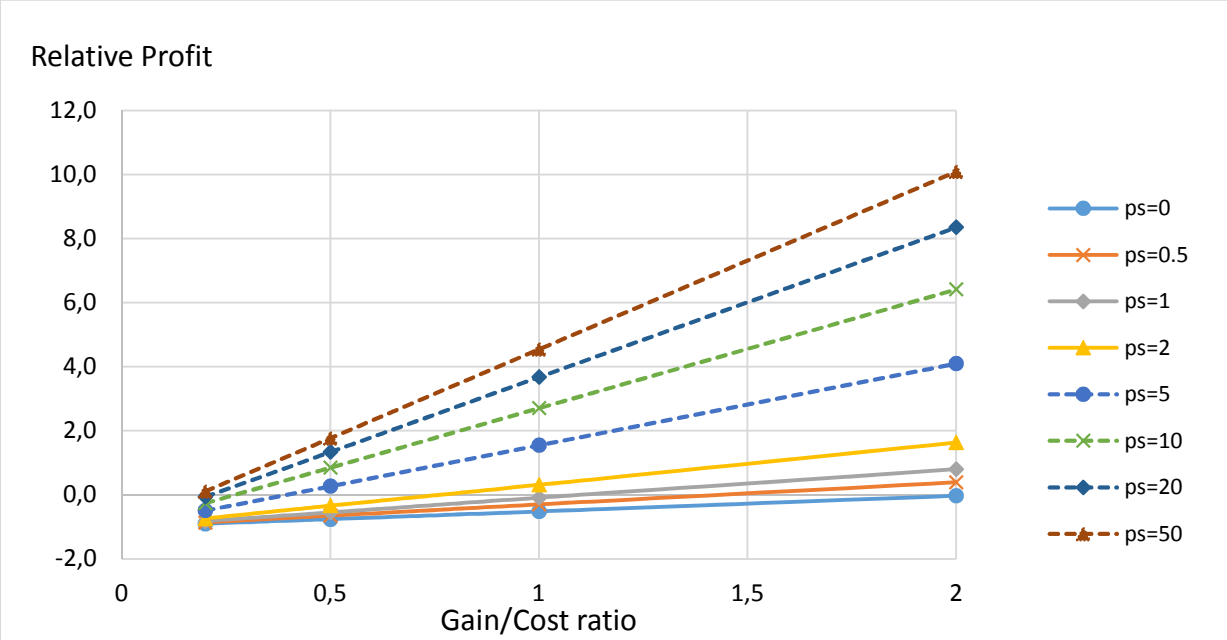
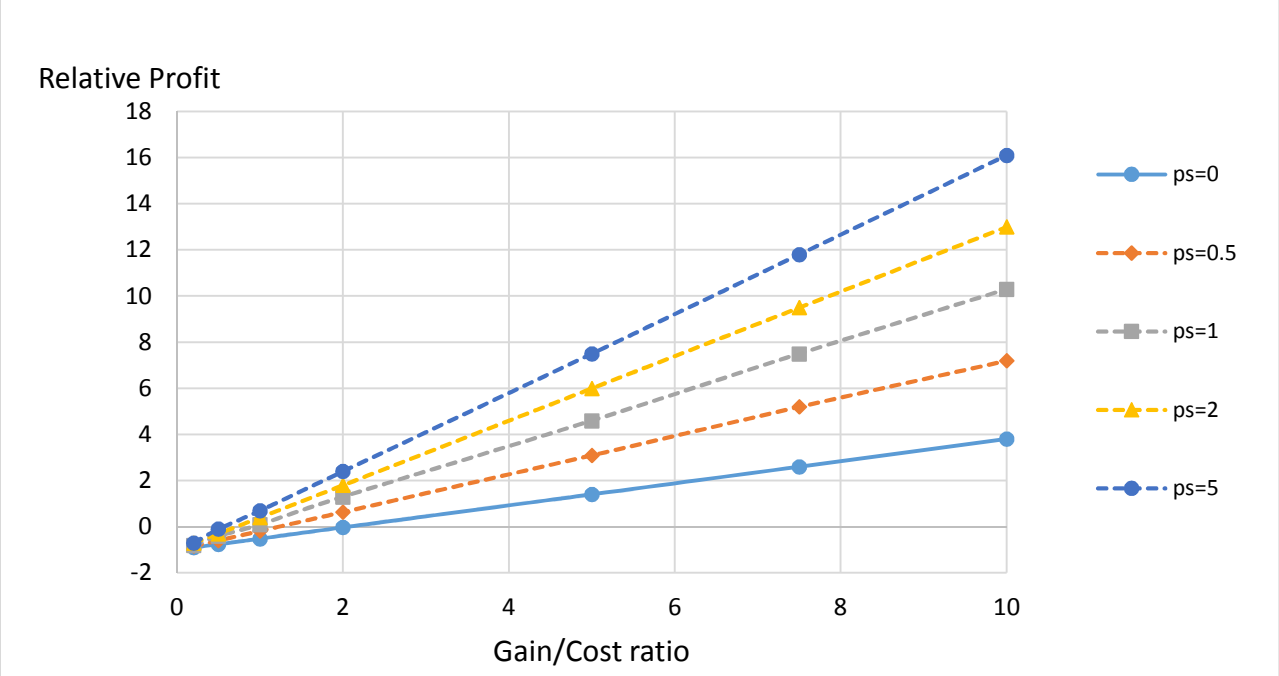


Figure SR3: Relative profit resulting from an increase of the female known paternity rate for various gain/cost ratios and relative sizes of the commercial population. Results for the AI_NoPT program (males born from non-progeny-tested AI sires) when females are used to transfer the genetic progress. a: Full range of gain/cost ratios considered. b: Limited range of gain/cost ratios [0.2,2]. We considered an investment period of 60 reproduction cycles. Each curve corresponds to a given population size. ps is the relative size of the commercial population compared to the nucleus size (nucleus size $= 1$). The gain/cost ratio is defined as the monetary gain due to an extra genetic standard deviation (MG) divided by the cost of one assignment (Cp).

(a)



(b)

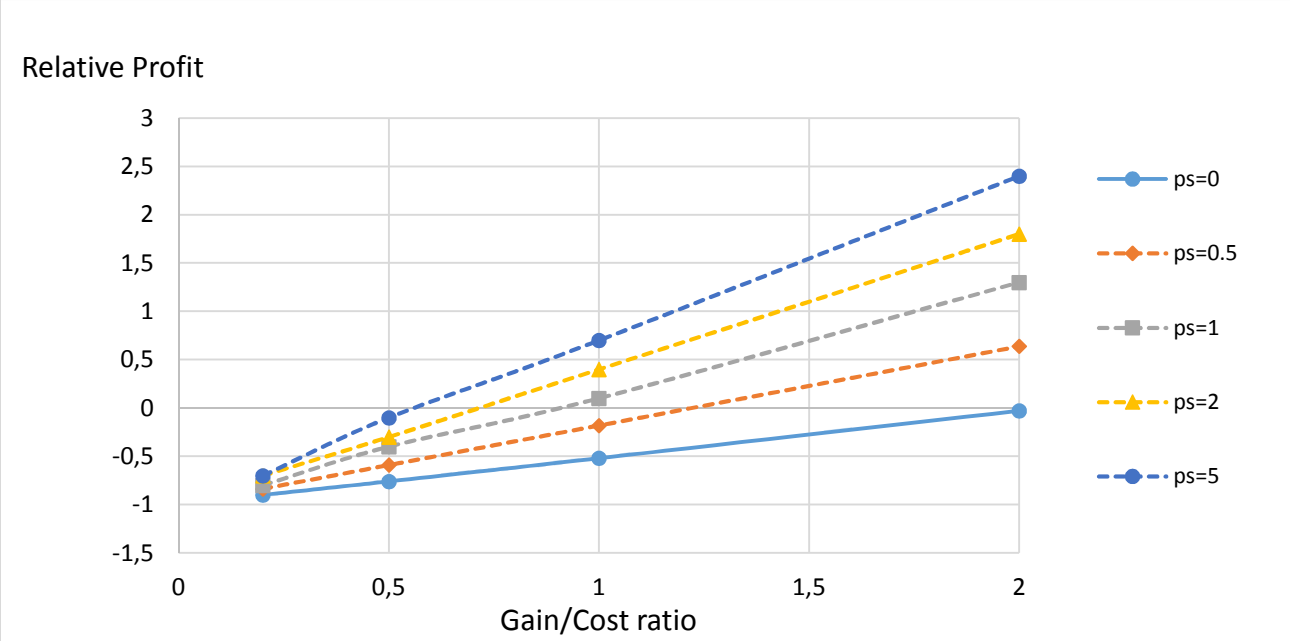
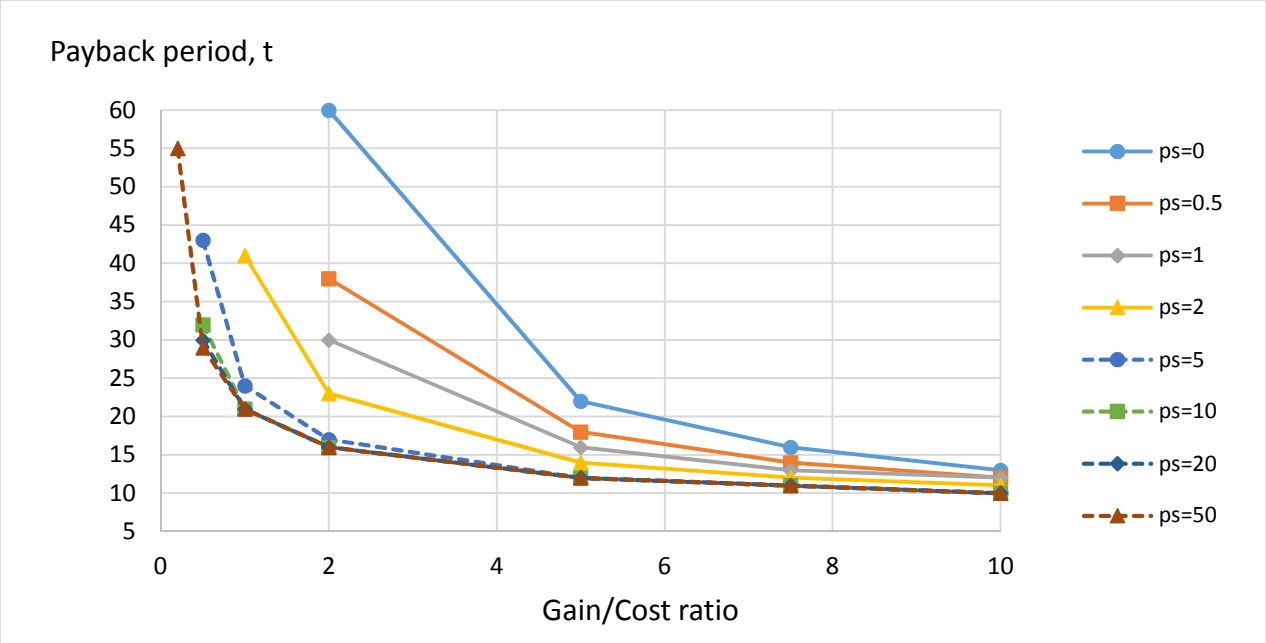
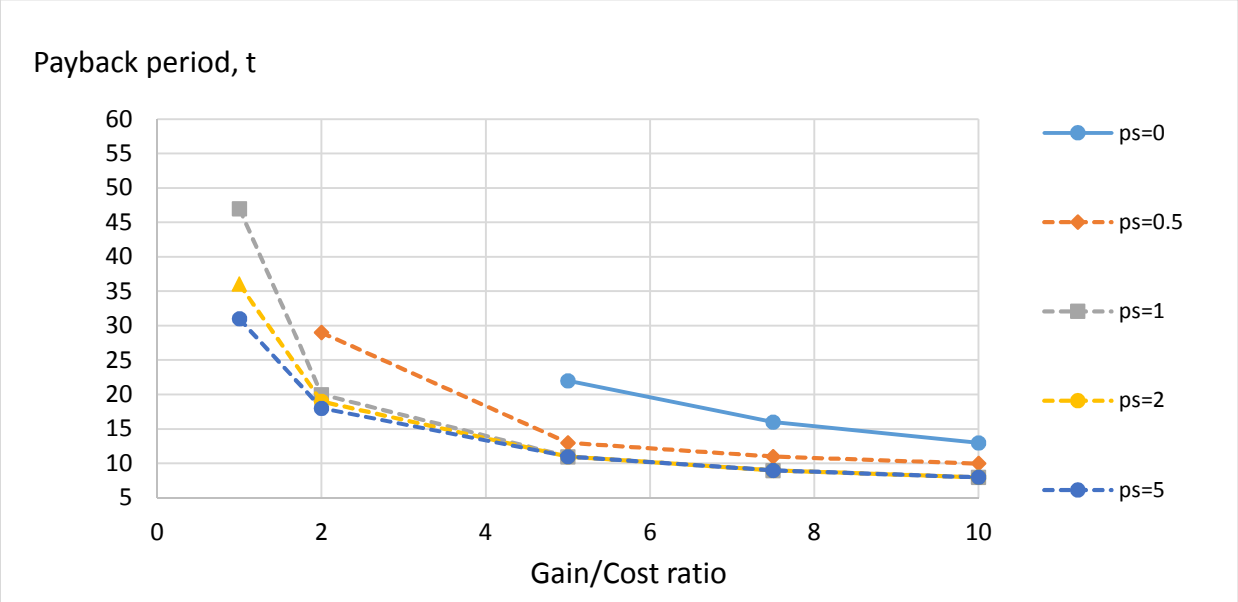


Figure SR4: Payback period (number of reproductive cycles) for various relative sizes of the commercial population as a function of the gain/cost ratio when males or females are used to transfer the genetic gain. Results for the AI_NoPT program (males born from non-progeny-tested AI sires). a: Transfer via males. b: Transfer via females. Each curve corresponds to a given population size. ps is the relative size of the commercial population compared to the nucleus size (nucleus size $= 1$). The gain/cost ratio is defined as the monetary gain due to an extra genetic standard deviation (MG) divided by the cost of one assignment (Cp).

(a)



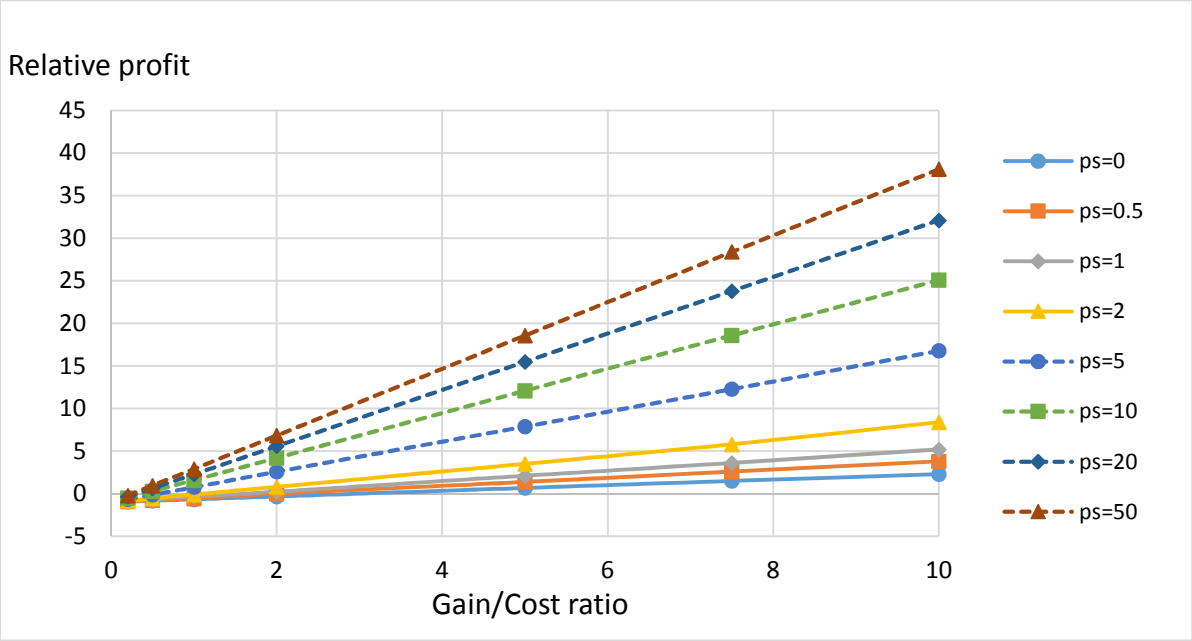
(b)



AI_PT program

Figure SR5: Relative profit resulting from an increase of the female known paternity rate for various gain/cost ratios and relative sizes of the commercial population. Results for the AI_PT program (males born from progeny-tested AI sires) when males are used to transfer the genetic progress. a: Full range of gain/cost ratios considered. b: Limited range of gain/cost ratios [0.2,2]. We considered an investment period of 60 reproduction cycles. Each curve corresponds to a given population size. ps is the relative size of the commercial population compared to the nucleus size (nucleus size = 1). The gain/cost ratio is defined as the monetary gain due to an extra genetic standard deviation (MG) divided by the cost of one assignment (Cp).

(a)



(b)

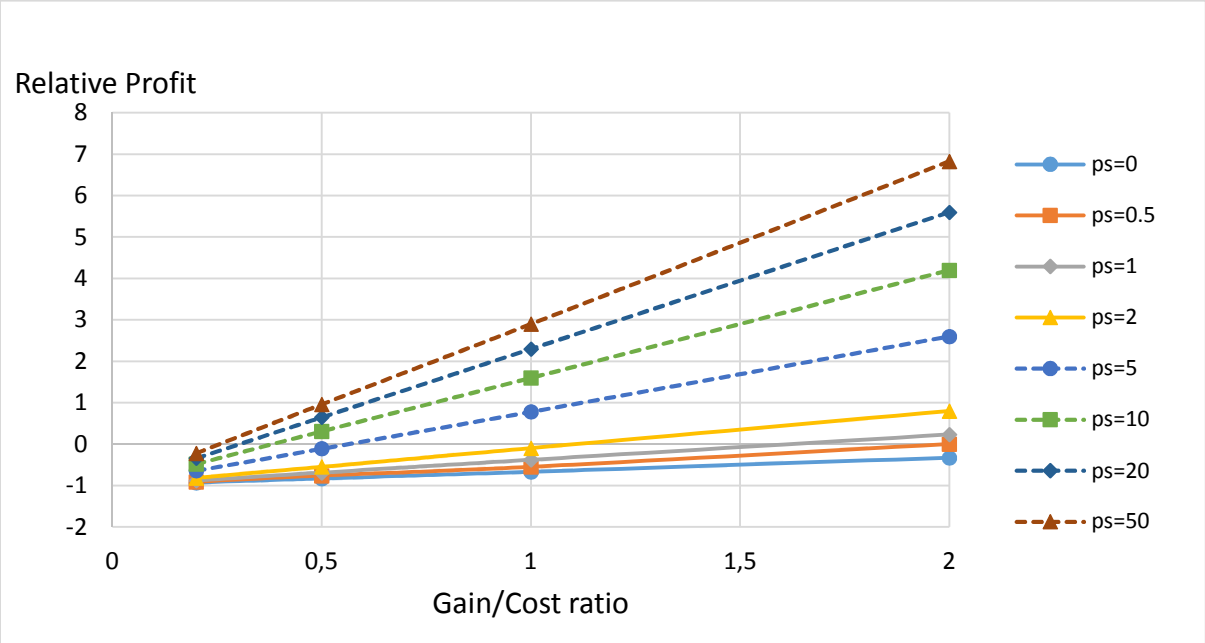
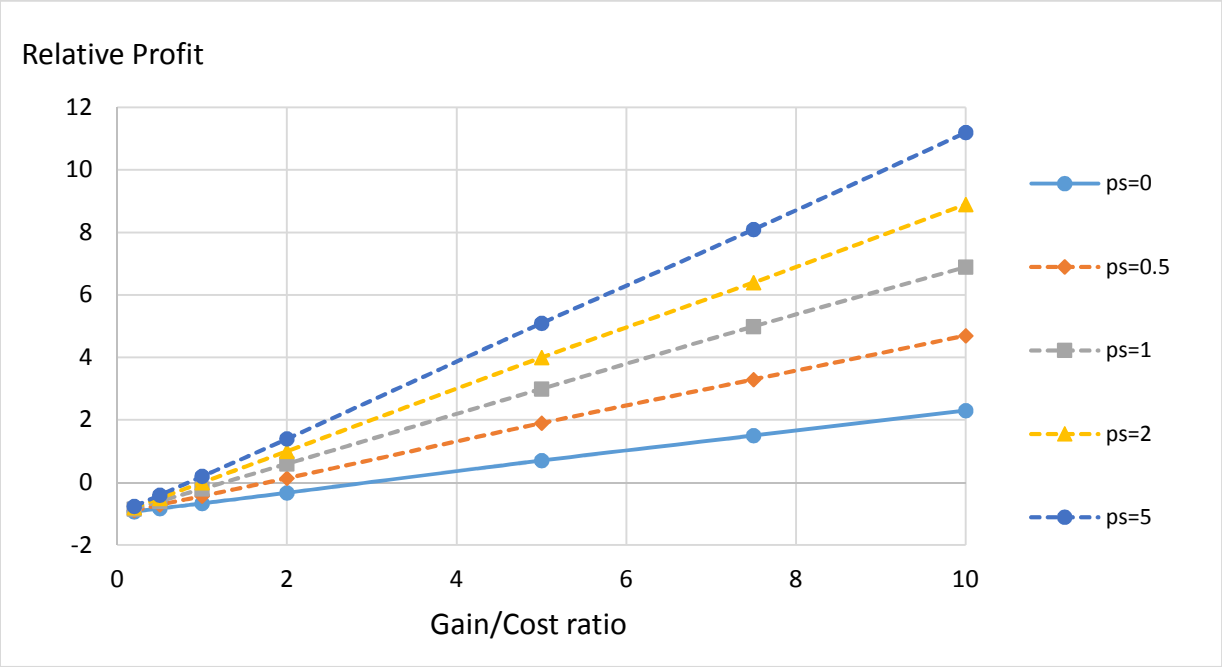


Figure SR6: Relative profit resulting from an increase of the female known paternity rate for various gain/cost ratios and relative sizes of the commercial population. Results for the AI_PT program (males born from progeny-tested AI sires) when females are used to transfer the genetic progress. a: Full range of gain/cost ratios considered. b: Limited range of gain/cost ratios [0.2,2]. We considered an investment period of 60 reproduction cycles. Each curve corresponds to a given population size. ps is the relative size of the commercial population compared to the nucleus size (nucleus size $= 1$). The gain/cost ratio is defined as the monetary gain due to an extra genetic standard deviation (MG) divided by the cost of one assignment (Cp).

(a)



(b)

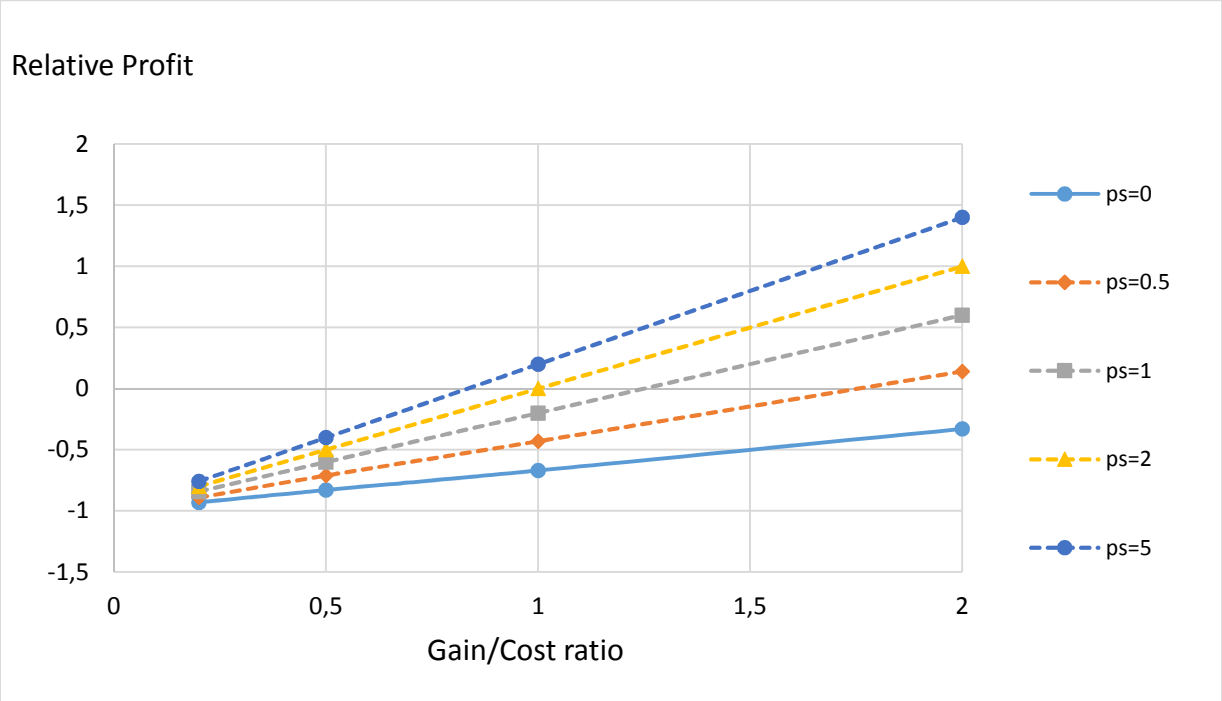
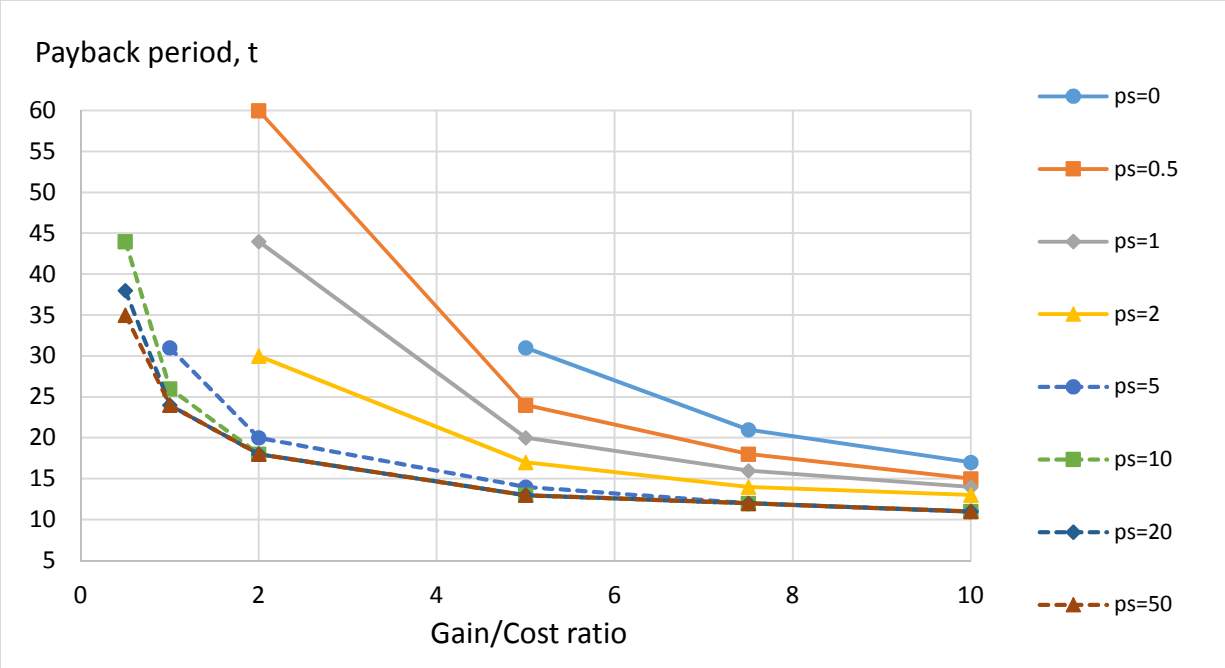
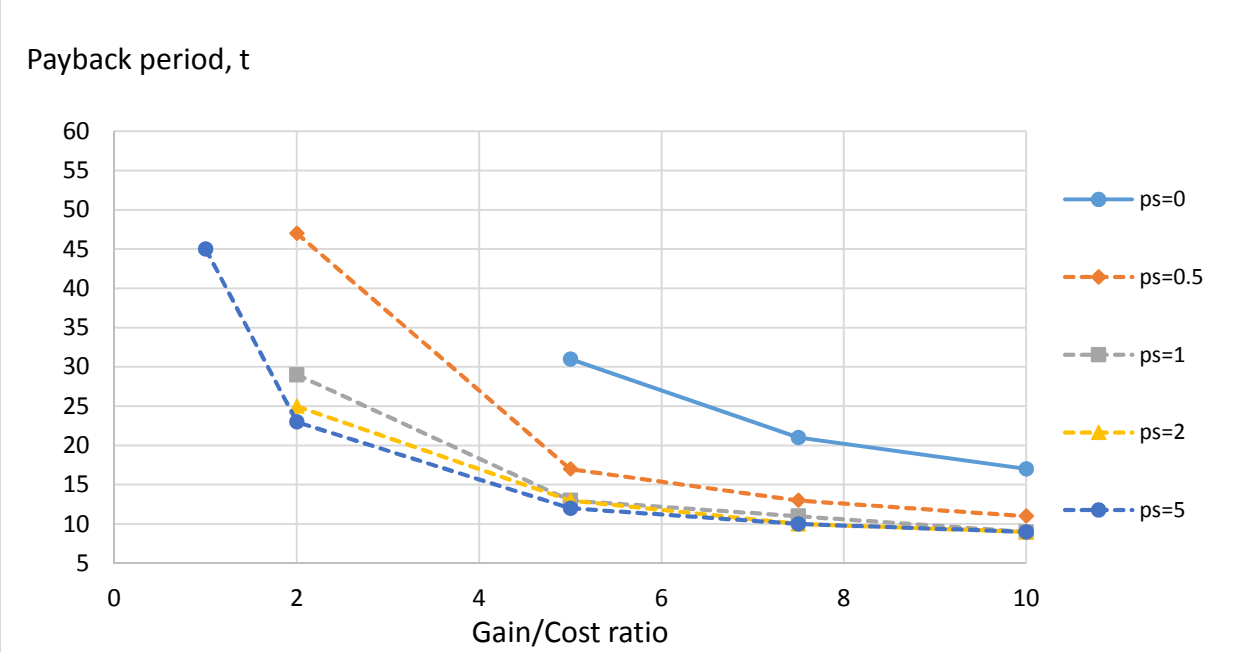


Figure RS7: Payback period for various relative sizes of the commercial population as a function of the gain/cost ratio when males or females are used to transfer the genetic gain. Results for the AI_PT program (males born from progeny-tested AI sires). a: Transfer via males. b: Transfer via females. Each curve corresponds to a given population size. ps is the relative size of the commercial population compared to the nucleus size (nucleus size $= 1$). The gain/cost ratio is defined as the monetary gain due to an extra genetic standard deviation (MG) divided by the cost of one assignment (Cp).

(a)



(b)



Chapitre 3 – optimisation de la gestion des gènes d'ovulation dans les populations ovines allaitantes

Résumé des articles 2 et 3

Le nombre d'agneaux sevrés par brebis mise en lutte, ou productivité numérique, affecte considérablement l'efficacité économique de la production ovine allaitante. L'amélioration génétique de la prolificité et de la viabilité des agneaux, concourant à l'amélioration de la productivité numérique, sont incluses dans l'objectif de sélection de la plupart des races ovines allaitantes françaises. En général on suppose que la variabilité des caractères est déterminée par l'action d'une multitude de gènes avec chacun un faible effet. Cependant, certains gènes avec des effets importants, appelés « gènes majeurs », ont été identifiés pour certains caractères. Dans les populations ovines, plusieurs gènes majeurs affectant le taux d'ovulation ont été identifiés. Chacun de ces gènes ont leurs propres caractéristiques : effet sur la prolificité, localisation sur le génome (chromosome sexuel ou non), induction ou non de stérilité chez les femelles porteuses homozygotes de l'allèle hyper-prolifique. Dans de nombreux cas, le génotype hétérozygote est le plus favorable tandis que le génotype porteur homozygote est très défavorable. L'élimination de l'allèle hyper-prolifique pourrait être contre-productive étant donné l'influence du caractère sur la rentabilité de la production ovine. Il s'agit donc, pour ce type de gènes majeurs, de déterminer leur gestion optimale c'est-à-dire la fréquence de chacun des génotypes (non-porteur, hétérozygote et porteur homozygote) chez les mâles et les femelles et l'organisation des accouplements en tenant compte des génotypes. La stratégie de gestion adoptée par le programme de sélection a des

conséquences sur la fréquence des génotypes des brebis du noyau (notamment des brebis hétérozygotes, les plus productives) et des agnelles disponibles pour la diffusion, sur le nombre d'individus à génotyper et sur le progrès génétique réalisé par le noyau.

L'objectif du premier article de ce chapitre était de déterminer, pour un gène majeur localisé sur un autosome, la proportion optimale de chaque type d'accouplement qui maximise la rentabilité économique au niveau de la population totale (élevages du noyau de sélection et élevages commerciaux). Un modèle déterministe a été conçu afin de calculer le profit pour différentes proportions de chaque type d'accouplements. Un premier modèle en génération discrète a été décrit. Le gène majeur a été supposé bi-allélique. Les femelles homozygotes porteuses de l'allèle d'hyper-prolificité sont supposées être écartées de la reproduction. A une génération donnée, la population de candidats était subdivisée en catégories en fonction du sexe des individus, de leur génotype et des génotypes de leurs parents (*i.e.* le type d'accouplement dont ils sont issus). Chaque catégorie, dont l'effectif dépendait de la proportion de chaque type d'accouplement, était caractérisée par un niveau génétique moyen (et précision moyenne) pour la prolificité, seul caractère sélectionné. Pour un sexe et un génotype donné les reproducteurs étaient sélectionnés par troncature au-dessus d'un seuil de sélection unique appliqué aux catégories comportant des candidats. Par exemple les mâles non porteurs étaient sélectionnés parmi les candidats non porteurs issus des accouplements entre i) deux parents non porteurs, ii) un parent non porteur et un parent hétérozygote (2 catégories) et iii) deux parents hétérozygotes, soit au total quatre catégories de candidats. L'évolution de la valeur génétique moyenne par catégorie au cours du temps a été décrite à partir d'un modèle de flux de gènes basé sur le calcul des contributions génétiques de chaque catégorie au renouvellement et des différentielles de sélection. Une fonction objective a ensuite été définie : elle incluait les recettes liées au nombre d'agneaux abattus, dépendant du niveau génétique

de la prolificité et de la fréquence des hétérozygotes, et les coûts liés aux besoins en génotypages. En utilisant une méthode de programmation quadratique séquentielle, les stratégies optimales (proportion de chaque type d'accouplement) qui maximisaient la fonction objective ont été déterminées pour des effets faibles à importants de l'allèle hyper-prolifique et une gamme de coûts de génotypage. Pour un effet de l'allèle et un coût de génotypage donné, la stratégie optimale a été comparée à des stratégies basées sur un nombre limité de types d'accouplement, plus simple à implémenter dans les programmes de sélection. En effet la stratégie optimale correspondait parfois à une combinaison de nombreux types d'accouplements entre femelles et mâles (*e.g.* [♀/♂] = 11% [++/m+], 25% [++/mm], 60% [m+/++], 5% [m+/mm]) qu'il est tentant de simplifier en choisissant une configuration plus facile à mettre en place (*e.g.* [♀/♂] = [++/mm], [m+/++], [m+/mm] ou bien [♀/♂] = [++/m+], [m+/++]).

Les résultats obtenus indiquent que la stratégie optimale était variable selon les situations évaluées (cinq niveaux pour l'effet de l'allèle et six niveaux pour le coût unitaire du génotypage) et l'horizon pris en compte (court terme et/ou long terme). Pour un horizon économique de 20 cycles de reproduction, un effet fort de l'allèle hyper-prolifique favorisait une utilisation élevée des femelles hétérozygotes au niveau du noyau, d'autant plus élevée que le coût du génotypage était bas. Pour des effets plus faibles de l'allèle hyper-prolifique et un coût de génotypage bas à modéré, la stratégie optimale tendait à maximiser le progrès génétique en se basant sur des utilisations plus équilibrées des femelles non porteuses et hétérozygotes. Pour des coûts de génotypage plus important, la solution optimale consistait à éliminer l'allèle hyper-prolifique en se basant uniquement sur des femelles non porteuses. La rentabilité économique obtenue pour certains designs basés sur deux ou trois accouplements était proche de celle obtenue pour les solutions optimales voire correspondait à la solution optimale pour certains situations (effet du gène x coût du génotypage).

Un modèle en génération chevauchante, adapté du modèle précédent, a ensuite été utilisé pour décrire le programme de sélection Lacaune viande Ovi-Test basé sur le testage sur descendance des mâles d'insémination. De nouvelles catégories ont été définies pour les reproductrices, selon leur parité, statut du père et génotype, et les reproducteurs en fonction de leur propre statut (mâle d'IA en testage, mâles d'IA améliorateur, mâle de monte naturelle) et génotype. Seuls les accouplements par insémination étaient organisés en fonction des génotypes parentaux. Le génotype des femelles, toutes conservées sur insémination, était supposé connu (génotypage ou déduction). La reproduction en monte naturelle étant réalisée avec des lots de lutte comportant plusieurs mâles, ceux-ci étaient tous non porteurs pour éviter la production de femelles porteuses homozygotes. Issus des accouplements entre les mâles d'IA améliorateurs et les meilleures brebis pour le caractère sélectionné, les mâles étaient présélectionnés sur ascendance en fonction de leur génotype, tel que décrit pour le modèle à génération discrète. Une fraction de ces candidats était évaluée sur descendance et la moitié supérieure sélectionnée comme mâles améliorateurs. Les mâles de monte naturelle étaient sélectionnés parmi les mâles présélectionnés non retenus pour le testage. Les mâles non sélectionnés étaient abattus. Les agnelles non retenues pour la reproduction dans le noyau étaient soit abattues soit diffusées comme reproductrice dans la population commerciale en accord avec les pratiques du programme de sélection Ovi-Test. De façon analogue à la démarche présentée précédemment, les solutions optimales correspondant à l'effet observé de l'allèle en ségrégation dans la population Lacaune (+0.5 agneau par mise-bas) pour différents coûts de génotypage ont été déterminées puis comparées à des stratégies basées sur deux ou trois types d'accouplements. Les résultats montrent que les solutions optimales, qui maximisaient la rentabilité économique à l'échelle de la population, étaient basées sur cinq types d'accouplement ([♀/♂] = [++/++], [++/m+], [++/mm], [m+/++], [m+/mm]). La stratégie

basée sur des accouplements entre non-porteurs et hétérozygotes ([♀/♂] = [++/m+], [m+/++]) a permis, en fonction du coût unitaire du génotypage, d'obtenir entre 92 et 97% de la rentabilité économique de la solution optimale. Basée sur seulement deux types d'accouplements, cette stratégie proche de l'optimum économique serait plus facile à implémenter.

Dans le second article, le modèle à génération chevauchante a été adapté pour étudier la situation d'un gène majeur localisé sur le chromosome sexuel X dont l'allèle hyper-prolifique aurait un effet de +0.5 agneau par mise-bas. Le caractère considéré dans cette étude, analogue à la « valeur laitière », est exprimé par les femelles mais indépendant de la prolificité. Les candidats mâles étaient uniquement sélectionnés sur ascendance, en fonction de leur génotype, tel que décrit pour le modèle à génération discrète. Les femelles porteuses homozygotes n'étant pas retenues pour la reproduction, seul deux génotypes ont été considérés par sexe (non-porteur et porteur pour les mâles, non-porteur et porteur hétérozygote pour les femelles) soit quatre types d'accouplement. La fonction objective a été modifiée pour prendre en compte l'effet de la sélection du caractère maternel sur les gains économiques puis la même démarche que précédemment a été adoptée pour déterminer les solutions optimales pour différents coûts unitaires du génotypage. Le progrès génétique pour le caractère sélectionné était plus faible dans une stratégie où l'allèle hyper-prolifique était maintenu et sa gestion optimisée que dans une stratégie ou l'allèle aurait été éradiqué. Les résultats montrent cependant que l'utilisation de cet allèle, compte tenu de son effet sur la prolificité, était économiquement avantageuse. Pour le coût actuel du génotypage, la stratégie optimale était basée sur une combinaison des quatre types d'accouplement ([♀/♂] = [++/+], [++/m], [m+/+], [m+/m]). La stratégie basée sur l'accouplement des femelles hétérozygotes à des mâles non-porteurs ([♀/♂] = [m+/+]), soit un seul type d'accouplement, a donné des

résultats proches de la solution optimale en termes de rentabilité économique et progrès génétique.

Article 2 : Optimal mating strategies to manage a heterozygous advantage major gene in sheep

J. Raoul, I. Palhière, J.M. Astruc, A. Swan and J.M. Elsen. 2017, Animal.

Animal, page 1 of 10 © The Animal Consortium 2017
doi:10.1017/S1751731117001835

animal

Optimal mating strategies to manage a heterozygous advantage major gene in sheep

J. Raoul[1,2†], I. Palhière[2], J. M. Astruc[1], A. Swan[3] and J. M. Elsen[2]

[1]*Institut de l'Elevage, BP 42118 - 31321 Castanet-Tolosan cedex, France;* [2]*GenPhySE, INRA, 24, chemin de Borde-Rouge - Auzeville Tolosane 31326 Castanet-Tolosan, France;* [3]*Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia*

(Received 6 December 2016; Accepted 5 June 2017)

Some mutations (or 'major genes') have a desirable effect in heterozygous carriers but an undesirable effect in homozygous carriers. When these mutations affect a trait of significant economic importance, their eradication, depending on their effect and frequency, may be counterproductive. This is especially the case of major genes affecting the ovulation rate and thus the prolificacy in meat sheep populations. To manage such situations, a mating design based on the major genotypes of reproducers has to be optimized. Both the effect of the major gene and the cost of genotyping candidates at this locus influence the expected genetic progress and profitability of the breeding plan. The aim of this study was to determine the optimal combination of matings that maximizes profitability at the level of the whole population (nucleus + commercial flocks). A deterministic model was developed and, using sequential quadratic programming methodology, the optimal strategy (optimal combination of matings) that maximized the economic gain achieved by the population across a range of genotype effects and genotyping costs was determined. The optimal strategy was compared with simpler and more practical strategies based on a limited number of parental genotype mating types. Depending on the genotype effect and genotyping costs, the optimal strategy varied, such that either the heterozygous frequency and/or polygenic gain was maximized with a large number of animals genotyped, or when genotyping costs were higher, the optimization led to lower heterozygous frequency and/or polygenic gain with fewer animals genotyped. Comparisons showed that some simpler strategies were close to the optimal strategy. An overlapping model was then derived as an application of the real case of the French Lacaune meat sheep OVI-TEST breeding program. Results showed that a practical strategy based on mating non-carriers to heterozygous carriers was only slightly less effective than the optimal strategy, with a reduction in efficiency from 3% to 8%, depending on the genotyping costs. Based on only two different parental genotype mating types, this strategy would be easy to implement.

Keywords: major gene, breeding program, deterministic model, economics

Implications

The number of lambs weaned per ewe joined considerably increases the economic efficiency of meat sheep production. This trait is included in the selection objective of most French breeds. Numerous major genes affecting the ovulation rate, which contributes to the number of lambs weaned, have been identified in sheep populations. This paper aims to understand the consequences of various strategies of major gene management on the genetic trends of prolificacy in a nucleus breeding program and their economic impact at both the nucleus and commercial population levels.

Introduction

Major genes are managed in various livestock breeding programs including pigs (de Vries *et al.*, 1998), poultry (Sodhi *et al.*, 2013), goats (Sánchez *et al.*, 2005) and sheep (McEwan, 1995; Palhière *et al.*, 2003; Martin *et al.*, 2014). Standard BLUP (Best Linear Unbiased Prediction) selection tends to increase favorable allele frequencies in the population if no specific management practices are implemented (Fontanesi *et al.*, 2015). However genetic prediction based on major gene information can increase the efficiency of selection, especially for low heritability traits, recessive alleles, low initial frequencies and strong allele effects (Larzul *et al.*, 1997; Gomez-Raya and Klemetsdal, 1999; Sánchez *et al.*, 2005; Carillier-Jacquin *et al.*, 2016). Manfredi *et al.* (1998) and Dekkers and Van Arendonk (1998) found that including genotype information, by dynamic optimization of within-genotype selection and planned genotype-based mating, increased the cumulated and discounted genetic progress due to both a higher polygenic gain and an optimal increase in favorable allele frequency across generations. A simulation study highlighted the importance of the major

† E-mail: Jerome.raoul@inra.fr

gene management on long-term genetic progress and the inbreeding level (Villanueva *et al.*, 1999) and reported that including major gene information for breeding value prediction had only a limited effect.

The number of lambs produced per female has a large impact on profitability in meat sheep production (Cheype *et al.*, 2013) and several major genes affecting this trait (indirectly through the ovulation rate) have been identified since the initial discovery in Booroola Merinos (Davis *et al.*, 1982; Piper and Bindon, 1982). Most often, these polymorphisms have a positive effect on heterozygous carriers but, in homozygous ewes, the effect is either too large leading to high rates of neonatal lamb mortality, or there are fertility issues. Consequently, the optimal management of such genes is still a challenge (Bodin *et al.*, 2011). For mutations localized on the X chromosome, such as *Inverdale*, several strategies can be implemented as pointed out by McEwan (1995) and Amer *et al.* (1998). However, the optimal management of a mutation localized on an autosomal chromosome, such as the *FecL* mutation (Drouilhet *et al.*, 2009), might be different. A planned mating scheme based on sire and dam genotypes (say *mm*, *m*+ and ++ for a biallelic *m*/+ locus) is one of the most important components for the optimal management of such a mutation. A maximum of nine mating types are possible with a biallelic locus. The selection of the parents within each genotype based on their polygenic value (Estimated Breeding Value, EBV) is the other component of this management strategy. With the exception of matings between homozygous carriers and non-carriers (*mm* by ++), all mating combinations are self-sufficient: parental genotypes are present in the progeny born from their mating (e.g. *m*+ females mated to ++ males produce both *m*+ and ++ progeny). Each mating type has its own characteristics: the proportion of a given genotype in their progeny varies (e.g. *mm* by ++ will produce only *m*+ progeny, whereas *mm* by *m*+ produce 50% of *m*+ and *mm*) as well as the number of progeny produced (higher for *m*+ compared with ++ females due to the allele effect). Matings between homozygous carriers *mm* and non-carriers ++ cannot be implemented alone and need at least one additional type of mating to produce replacements. In addition, some mating combinations increase the selection intensity and result in a higher genetic gain. For each mating combination, the heterozygous frequencies in both the nucleus and the commercial population, the genetic gain and the amount of genotyping involved can be determined.

The aim of this study was to determine the optimal strategy (combination of matings) that maximizes the benefit of the breeding scheme when homozygous carrier females are systematically culled. Complex breeding organization would typically be required to implement the optimal strategy, so the latter was also compared with simpler strategy that breeders would find easier to implement. First, a simple discrete generation deterministic model was developed to determine the optimal strategy according a range of genotype effects (GEs) and genotyping costs (GCs). Then, an overlapping generation deterministic model was developed to illustrate such an approach with the real example of the French Lacaune meat sheep breeding program OVI-TEST (2016).

Material and methods

Discrete generation model

Variables, parameters and input values of parameters are defined in Table 1. At each time t, $nb_F = 1000$ females were mated to $nb_M = 200$ males to produce a new candidate generation. A steady state was supposed and the proportions of each possible mating, that is, parental genotype combinations, were constant across generations. Homozygous carrier females were not used due to the negative phenotypic effects on reproduction associated with this genotype, leaving six possible matings: heterozygous (φ_{m+}) and non-carrier (φ_{++}) females could be mated with non-carrier (σ_{++}), heterozygous (σ_{m+}) or homozygous carrier (σ_{mm}) males. As described in Figure 1, the candidate population was divided into categories according to their sex a (*m* or *f*), their genotype $g = ++$, *m*+, *mm* and their dam × sire genotype combination (i.e. mating $[g_d g_s]$). The number of weaned lambs per female was the only selected trait. Parents of the next generation were selected based on their EBVs within each genotype and the mating combinations (i.e. dam × sire genotype combination) from which those genotypes where produced (e.g. *mm* progeny can be derived from $\varphi_{m+} \times \sigma_{m+}$ and $\varphi_{m+} \times \sigma_{mm}$). It was assumed that polygenes affecting the selected trait and the major gene genotypes were genetically independent. The mean EBVs of the categories in the next generation were computed in accordance with the mating plan and selection intensities. An objective function which included discounted incomes (from slaughtered lambs) and costs (due to genotyping) was used to assess the economic value of the mating design assessed.

Categories. The number of animals per candidate category at time t, $c_{[g_d g_s]ga(t)}$, depended on $x_{[g_d g_s]}$, the number of g_d dams mated to g_s sires, $\beta_{[g_d g_s]g}$, the proportion of candidates of genotype g born from parents of genotype $[g_d g_s]$, $NLFN_{g_d(t-1)}$ the number of progeny per nucleus dams with genotype g_d at $t-1$ and the proportion of sex a (0.5) among progeny: $c_{[g_d g_s]ga(t)} = x_{[g_d g_s]} \times \beta_{[g_d g_s]g} \times NLFN_{g_d(t-1)} \times 0.5$. As some genotypes were not produced for a given mating (e.g. $\beta_{[++/++]m+} = \beta_{[++/++]mm} = 0$), only 11 categories corresponding to $\beta_{[g_d g_s]g} \neq 0$ are reported in Figure 1. A mating design was fully described by the set x of $x_{[g_d g_s]}$ values ($g_d = ++$, *m*+ and $g_s = ++$, *m*+, *mm*). Without loss of generality, the GE was defined as the phenotypic difference between *m*+ and ++ females for the selected trait due to the major gene. Different values of GE were assessed (Table 1). The EBV average reliability of parents was assumed to be 0.35 for all candidate categories.

Genetic model. A genetic model was implemented to describe the genetic evolution of the selected trait over time. The classical gene flow approach described by Hill, Elsen and

Table 1 *Parameters of the models*

a: State, decisional and linked variables

Name	Parameter
$[g_d g_s]$	Mating given the dam g_d and sire g_s genotypes
g	Genotype (non-carrier: ++, heterozygous: *m*+, homozygous carrier: *mm*)
l[1]	Parity, $l = [1,5]$
a	Sex ($a = m$ for male, $a = f$ for female)
t	Time unit
z	Subscript use to define a category: $[g_d g_s]ga$: animals with genotype g and sex a originating from mating between a dam with genotype g_d and a sire with genotype g_s. For the overlapping model subscript l indicates the dam parity
α_l[1]	Proportion of dams in parity l
$x_{[g_d g_s]}$	Number of mating females $g_d \times$ male g_s
$c_{z(t)}$	Candidate number of category z at t
$NLFN_{g_d(t)}$[2]	Number of lambs per nucleus female of genotype g_d at t
$NLFC_{g_d(t)}$[2]	Number of lambs per commercial female of genotype g_d at t
$NLFE_{g_d l(t)}$[1]	Number of lambs per nucleus female of genotype g_d and parity l at t on natural estrus
nb_{ga}	Number of selected parents of genotype g and sex a
nb_{NBg}[1]	Number of newborn animals of genotype g born from natural mating sire
$\beta_{[g_d g_s]g}$	Proportion of genotype g progeny born from dam–sire genotypes $[g_d g_s]$
$EBV_{z(t)}$	Average polygenic value of category z at t
ρ^2_z	Average reliability of polygenic breeding value of category z
$p_{z',z(t)}$[2]	Genetic contributions of category z at $t-1$ to category z' at t
$p_{[g_d g_s]lm,j}$[1]	Genetic contributions of parental category j at t to category $[g_d g_s]lm$ at $t+1$
$d_{z(t)}$	Selection differential applied to category z to select parents at t
$K_{ga,t}$	Single truncation point across sex-genotype categories
$\omega_{z(t)}$	Probability for a candidate of category z at t to be selected as a parent
$q_{z(t)}$	Proportion of parents ga selected among candidate category z at t
$Geno_t$	Number of genotyped animals at t
NS_t	Number of slaughtered animals related to t
$ObfFun$	Discounted revenue at time horizon plan T

b: Input parameters

Name	Parameter	Discrete model[2]	Overlapping model[1]
h^2	Heritability	0.1	0.1
rep	Repeatability	0.2	0.2
σ_g	Genetic SD	1	1
GE[3]	Genotype effect – phenotypic difference due to the major gene between *m*+ and ++ females	0.5, 1, 2, 3 and 4	0.5
nb_M[2]	Number of sires	200	
nb_{ym}[1]	Number of young males selected per time		150
nb_F	Number of nucleus females	1000	12 000
$NLF_{0,g}$	Number of lambs per ++ females with genotype g at $t=0$ $(NLF_{0,g=m+} = NLF_{0,g=++} + \mathrm{GE})$	6.0	1.5
$\nu_{jag_{s'}}$[1]	Sire category j of genotype $g_{s'}$ used to produce candidates of sex a	–	See additional material
GC	Genotyping cost (expressed as a proportion of marginal value per slaughtered lamb)	0/0.25/0.5/1/1.5/2	0/0.15/0.25/0.4/0.5/0.6/0.75/1/1.25/1.5/2
d	Discounting rate	0.05	0.05

[1]Variable only implemented in the overlapping generation model.
[2]Variable only implemented in the discrete generation model.
[3]Per productive life for the discrete model and per lambing for the overlapping model.

Mocquot (Elsen and Mocquot, 1974; Hill, 1974; Elsen, 1993) was adapted to candidate categories (but not parents) using: $\mathbf{ebv}_{t+1} = \boldsymbol{P} \times (\mathbf{ebv}_t + \boldsymbol{d}_t)$, where $\boldsymbol{P}$ denotes the transition matrix describing the gene flow between candidate categories after one generation, $\mathbf{ebv}_t$ and $\boldsymbol{d}_t$ the vectors of the average EBV and selection differential per candidate category at time t.

New parents were selected within sex × genotype by truncation selection on their EBVs: five single truncation thresholds (two for females and three for males) were determined across candidate EBV distributions: ++ parents were selected among candidates produced from matings [++/++], [++/*m*+], [*m*+/ ++] and [*m*+/*m*+], *m*+ parents from [++/*m*+], [++/*mm*], [*m*+/ ++], [*m*+/*m*+] and [*m*+/*mm*]

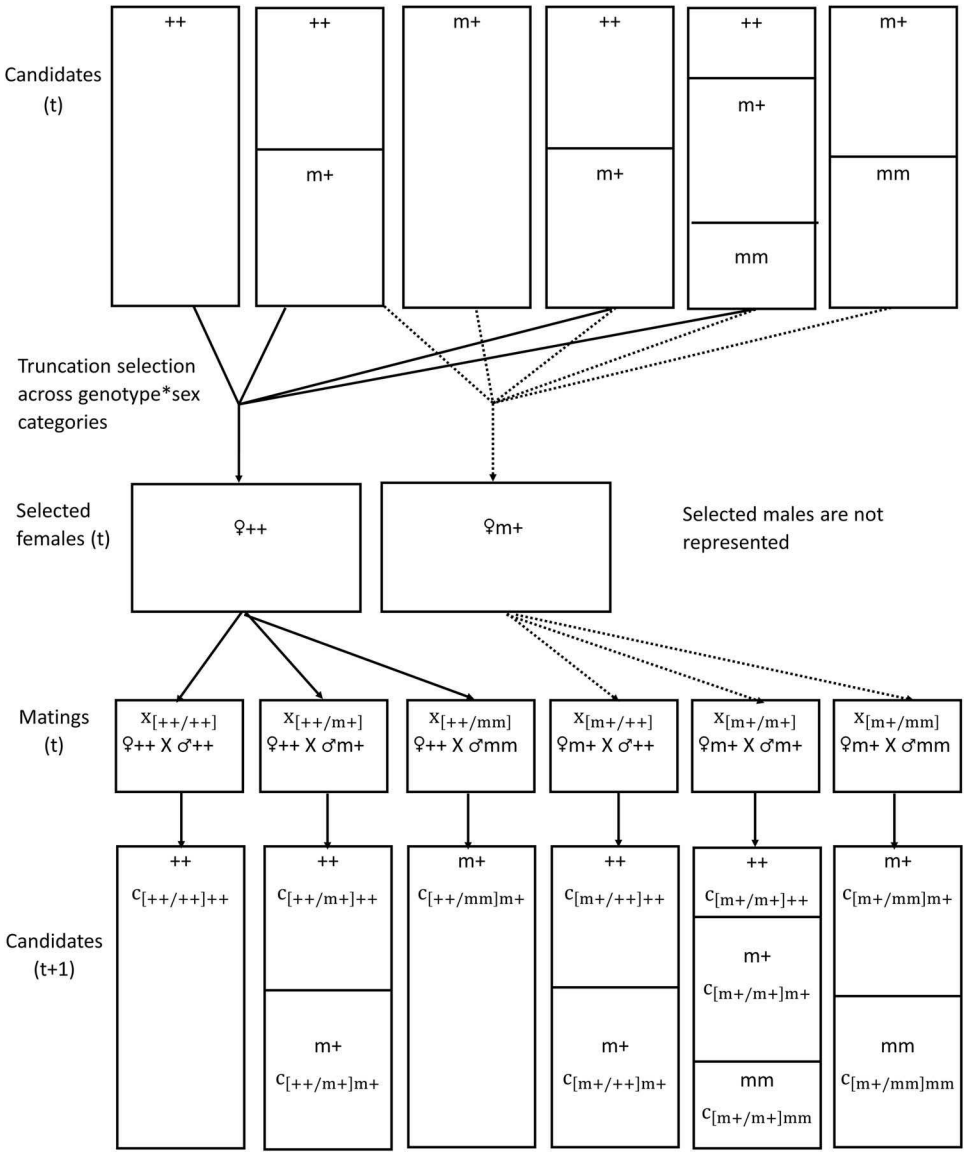


Figure 1 Overview of the discrete generation model. To simplify the subscript *a* (sex) has been removed. $c_{[g_d g_s]g(t)}$ = animal number belonging to $[g_d g_s]$ *ga* category at *t*; $[g_d g_s]$ = parental genotype (mating type); *g* = candidate genotype.

and *mm* males (only) from [*m*+/*m*+] and [*m*+/*mm*]. Selected parents were randomly assigned to a type of mating to produce the next generation.

To minimize boundary effects (i.e. a very low number of selected animals within a category that would lead to extreme selection differentials), we computed selection differentials for categories that represented at least 2% (within sex) of the parent population. Applying the gene flow equation, we computed the mean EBV per category at $t+1$ and updated the average phenotypic value of parents selected for the nucleus or transferred to the commercial population. Computational details are provided in Supplementary Material S1.

Objective function. Incomes were modeled as the number of lambs slaughtered multiplied by the marginal value per animal slaughtered (independent of genotypes), whereas costs were proportional to the genotyping needs. As marginal value per animal and cost per genotyping are subjected to evolve, we expressed the GC as a proportion of marginal value. A low GC means that the cost per genotyping is low regarding to the marginal value per slaughtered animal. A large range of values was considered (from 0 to 2, see Table 1) to cover all possible situations. The genotyping needs at time *t*, $Geno_t$, depended on the number of parents selected from each type of mating. The genotypes of animals born from [++/++] and [++/*mm*] were known without genotyping. For other matings, genotyping needs, computed by sex, depended on the relative number of animals selected of each genotype and originating from the same mating. For example, the number of genotyping related to the mating [++/*m*+] for females depended on the number of ++ and *m*+ females selected from this specific mating. In this case the number of genotype is computed as the maximum

of ++ *v.* *m*+ females selected. The number of slaughtered lambs related to time *t*, NS_t, included both animals slaughtered at *t* in the nucleus and $t+1$ in the commercial population. Non-selected males were slaughtered at time *t*, whereas non-selected nucleus females born from the mating types [++/++], [++/*m*+], [++/*mm*] and [*m*+/++] were transferred to the commercial population and produced offspring slaughtered at time $t+1$. For non-selected females born from [*m*+/*m*+] and [*m*+/*mm*], to avoid the transfer of *mm* females to the commercial population, we chose to either genotype all females and transfer only ++ or *m*+, or cull all females, depending on which option resulted in the best economic value. To distinguish the short- and long-term effects of mating management, three profit functions were computed for the short-term ($t=\{2,6\}$), for the long-term ($t=\{16,20\}$) and for the whole time horizon ($t=\{2,20\}$). Details on how $Geno_t$, NS_t and the profit function were computed are provided in Supplementary Material S2.

Overlapping generations

To deal with the real case of the French Lacaune meat sheep breeding program run by the OVI-TEST breeding society, we extended the previous model to an overlapping generation deterministic model taking into account dam parity and the aging of males. The number of weaned lambs was expressed per lambing and the GE was derived from the effect of the *FecL* allele, that is, +0.5 lamb born per lambing, as estimated by Martin *et al.* (2014) in the OVI-TEST population. The full description of the overlapping model is provided in Supplementary Material S3.

Optimal and simple strategy comparison

Schemes were modeled to optimize the number of each mating (**x**), which were the decision variables. The values of state variables (number of animals and average EBV by category) at time *t* depended on decision variables at *t* and values of state variables at $t-1$. Output variables (incomes and costs) depended on decision variables and the values of state variables at *t*. Using the NAG routine E05UCF (The Numerical Algorithms Group (NAG), Oxford, United Kingdom) based on sequential quadratic programming methodology, we obtained the **x** that maximized the profit function subject to constraints: $\boldsymbol{lb} \leq \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{Ax} \end{pmatrix} \geq \boldsymbol{ub}$, where $\boldsymbol{lb}$ and $\boldsymbol{ub}$ were the lower and upper bounds and $\boldsymbol{A}$ the matrix that described linear constraints between variables. Details of linear constraints are given in Supplementary Material S4. With this approach, an optimal mating scheme (OPT) was obtained, fully defined by the optimum values of the vector **x** elements. Simple strategies were compared to OPT. In strategy S2, two matings were organized: [++/*m*+] and [*m*+/++] such that only matings between non-carriers (++) and heterozygotes were allowed. This strategy required systematic genotyping of candidates but did not produce *mm* individuals. In strategy S3+, three matings were organized: [++/++], [++/*mm*] and [*m*+/*mm*]. The strategy S3+ required homozygous carrier males ($\male_{mm}$) and homozygous females ($\female_{++}$) to produce heterozygous females ($\female_{m+}$) by [++/*mm*]. This type of mating produced heterozygous females that did not need to be genotyped but required at least one additional type of mating to replace the parents ($\female_{++}$ and $\male_{mm}$). Homozygous carrier males ($\male_{mm}$) were produced by a small proportion of [*m*+/*mm*], whereas non-carrier females ($\female_{++}$) were produced by [++/++]. In strategy S3m, three matings were organized: [++/*mm*], [*m*+/++] and [*m*+/*mm*].This strategy, based on the mating between homozygous carrier males ($\male_{mm}$) and non-carrier females ($\female_{++}$) was similar to S3+ but non-carrier females ($\female_{++}$) were produced by [*m*+/++].

Results

Using a discrete generation model, this study assesses the consequences of the proportion of mating types (combination of parental genotype) on a profit function computed at different time horizons. The optimal strategy (OPT) corresponded to the strategy that maximizes the profit function and was assessed for various combinations of GE and GC. The profit obtained for the optimal strategy was then compared with simpler strategies based on either only two (S2) or three mating types (S3+, S3m). An overlapping model was then used to illustrate such an approach with the real example of the French Lacaune meat sheep breeding program OVI-TEST.

Discrete generation model:

The mating combinations that yielded the maximum return depending on GE and GC in the short-term, the long-term and over the whole time horizon are reported in Figure 2a to c. In the short term, the optimal scenarios were diverse, and depended on GC. The [*m*+/++] mating, that is, the use of heterozygous females ($\female_{m+}$), was generally predominant in optimized designs. Heterozygous females ($\female_{m+}$) were replaced by genotyping the progeny of [*m*+/++] when the GC was low, but increasingly from [++/*mm*], which does not need genotyping, when the cost was higher. In this situation a small proportion of [*m*+/*mm*] was required to produce homozygous carrier rams ($\male_{mm}$). This situation was very close to the S3m strategy. When the gene effect was very low and the GCs high, homozygous females ($\female_{++}$) were used either with homozygous non-carrier rams ([++/++]), approaching 100% in the extreme scenario where GC = 2 and GE = 0.5, or with homozygous carrier rams ([++/*mm*]), which produced *m*+ ewes to be mostly used in commercial flocks. A proportion of [++/*m*+] was observed when both GE and GC were low, in combination with [*m*+/++], a scenario which optimized the use of genotyping as all females, ++ and *m*+, were candidates for replacement. This situation was very close to the S2 strategy. The latter scenario was predominant for the long-term time horizon. It was only when GE was very low and GC high that [++/++] dominated, as in the short term. The relative proportion of [++/*m*+] increased with GC and decreased with GE. Optimization over the whole time horizon gave scenarios clearly intermediate between short- and long-term optimizations.

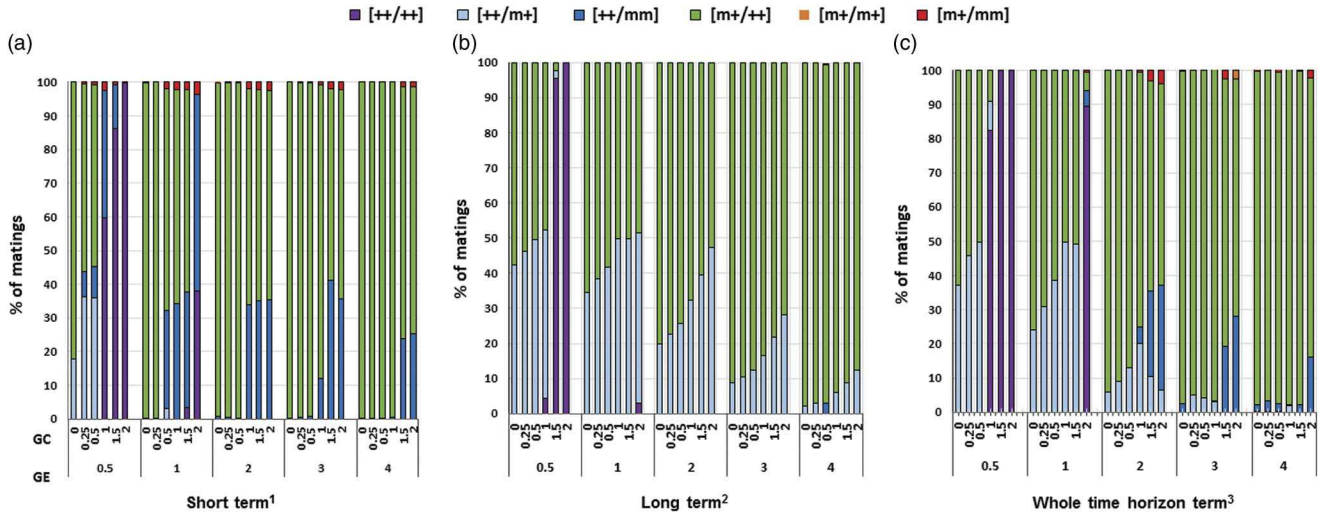


Figure 2 Mating combinations that gave the maximal return as a function of genotype effect and genotyping cost – discrete generations in (a) short-term, (b) long-term, and (c) whole time horizon term. Matings: [dam genotype × sire genotype]. GE = genotype effect – phenotypic difference due to the major gene between *m+* and ++ females. GC = genotyping cost (expressed as a proportion of marginal value per slaughtered lamb). [1] $t = \{2,6\}$; [2] $t = \{16,20\}$; [3] $t = \{2,20\}$.

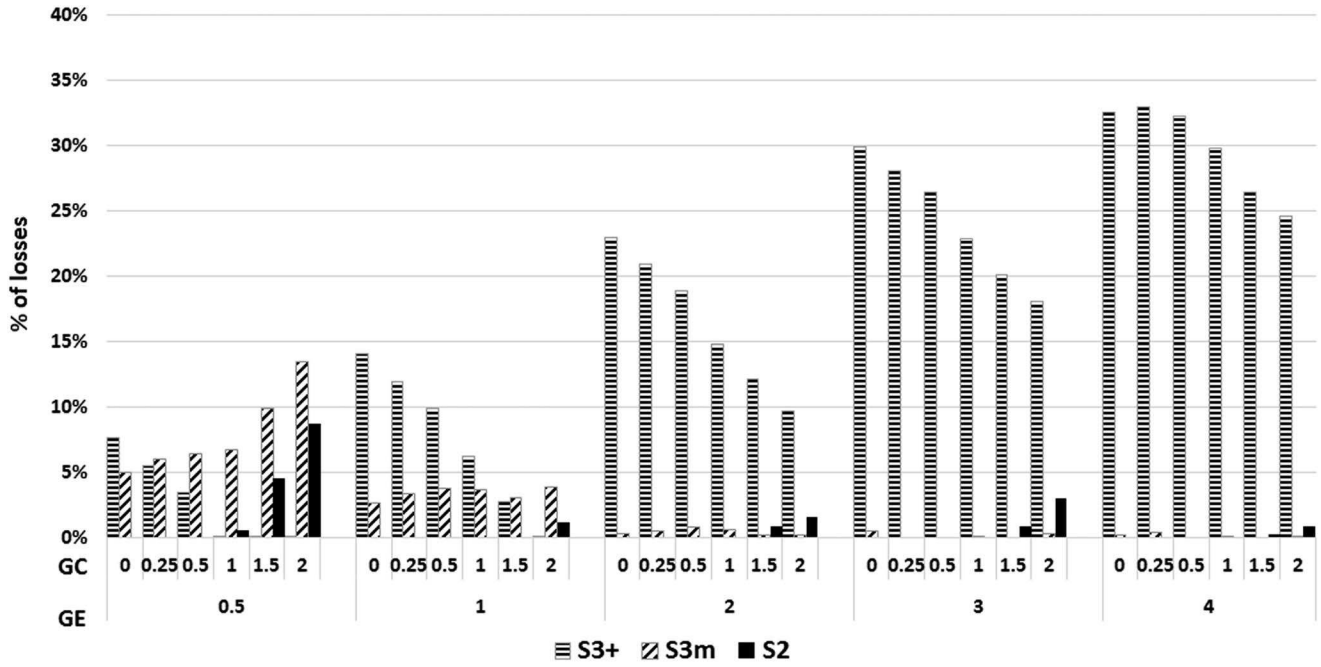


Figure 3 Losses of economic returns for various strategies compared with the optimal strategy as a function of the genotype effect and genotyping cost – discrete generations. Matings: [dam genotype × sire genotype]. S2 = strategy based on [++/*m*+] and [*m*+/++]; S3*m* = strategy based on [++/*mm*], [*m*+/++] and [*m*+/*mm*]; and S3+= strategy based on [++/ ++], [++/*mm*] and [*m*+/*mm*]. GE = genotype effect – phenotypic difference due to the major gene between *m+* and ++ females. GC = genotyping cost (expressed as a proportion of marginal value per slaughtered lamb). Time horizon, $t = \{2,20\}$.

The percentages of losses over the whole time horizon of S2, S3+ and S3m strategies compared with the OPT strategy, for a range of GE and GC, are reported in Figure 3. Except at low GE and high GC (e.g. GE = 0.5 and GC = 1.5), the losses for the S3+ strategy were higher than losses for S2 and S3m strategies. The percentage of losses of S3+ increased with GE and decreased with GC. At a low GE (⩽1), S2 outperformed S3m, whatever the GC. For a low GE and high GC (GE = 0.5 and GC = 2), the loss for S2 was ~10%. Otherwise, for higher GE, the S2 and S3m strategies gave similar results and were close or equal to the OPT strategy. Only at high GE and GC, S3m slightly outperformed S2.

Overlapping generation model

The optimal combination of matings as a function of GC is reported in Figure 4. Two optimal combinations can be described. First, when GC was null, OPT was based on [*m*+/++] (60%), [++/*m*+] (33%) and [++/*mm*] (6%) (plus [*m*+/*mm*] to produce δ_{mm}). Second, for the range of GC assessed, OPT was nearly stable and based on the same

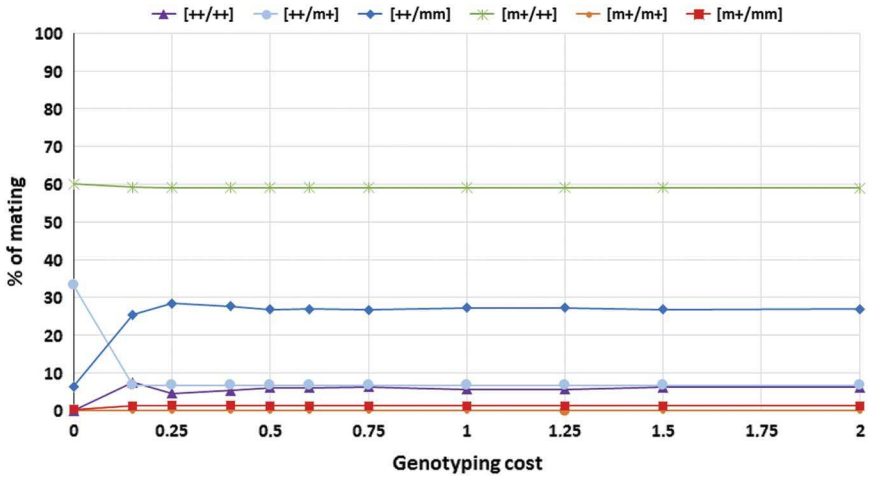


Figure 4 Optimal combination of matings as a function of genotyping cost – overlapping generations. Matings: [dam genotype × sire genotype]. GE = genotype effect – phenotypic difference due to the major gene between *m*+ and ++ females; GC = genotyping cost (expressed as a proportion of marginal value per slaughtered lamb).

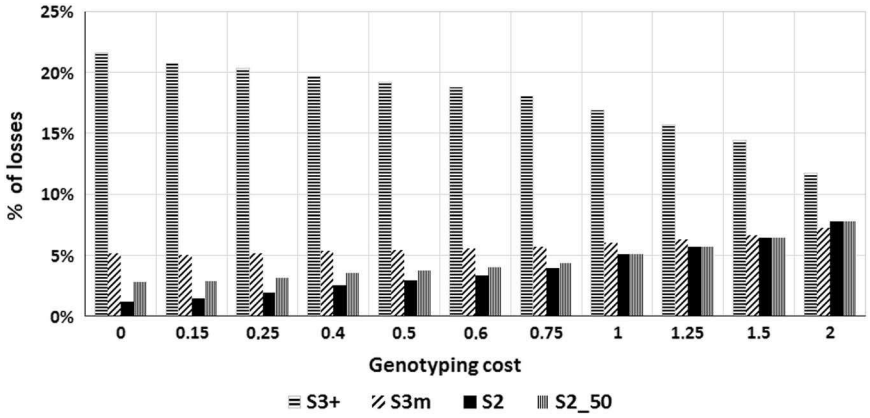


Figure 5 Percentage of losses of simpler strategies compared with the optimal strategy as a function of genotyping cost – overlapping generations. Matings: [dam genotype × sire genotype]. S2 = strategy based on [++/*m*+] and [*m*+/++]; S3*m* = strategy based on [++/*mm*], [*m*+/++] and [*m*+/*mm*]; and S3 + = strategy based on [++/++], [++/*mm*] and [*m*+/*mm*]. GE = genotype effect – phenotypic difference due to the major gene between *m*+ and ++ females. GC = genotyping cost (expressed as a proportion of marginal value per slaughtered lamb).

matings plus [++/++]: [*m*+/++] remained high at 59%, whereas [++/*mm*] increased (to ~27%), [++/*m*+] decreased (~7%) and [++/++] reached 6% ([*m*+/*mm*] remained low at 1%). The OPT strategy consisted of a combination of five types of mating. The results of the comparison of OPT with simple strategies are reported in Figure 5. An additional strategy, S2_50, identical to S2 but with a balanced proportion of [++/*m*+] and [*m*+/++] was tested. Figure 5 shows that S3+ displayed the highest losses (from 12% to 22%) irrespective of GC. The losses with the S3m strategy were higher than for the S2 and S2_50 strategies at low GC values but the gap was reduced at higher GC. At the highest GC, the losses of S3m were slightly lower than for S2 and S2_50. For all cases, for S3m, S2 and S2_50, the losses were limited and ranged from 1% to 8%. The [++/*m*+]/[*m*+/++] ratio obtained for the S2 strategy was of 40/60 up to GC = 0.75, and about 50/50 above (results not shown). The percentage of heterozygous females ($♀_{m+}$) mated in the nucleus and transferred to the commercial population, as well as annual genetic gain (AGG) and genotyping needs obtained for the current estimated GC (GC = 1) are reported in Table 2 for each strategy. The AGG and genotyping needs were expressed as relative values compared with the idealistic situation (OPT, GC = 0). Optimal mating scheme gave both the highest AGG (98% to 100%) and the highest $♀_{m+}$ frequency in the nucleus (around 60%) for a moderate level of genotyping (66% to 69%). The S2 and S2_50 strategies gave a similar AGG (91% to 95%) but were associated with higher genotyping needs (93% to 112%), whereas S3m gave the highest $♀_{m+}$ frequency in the commercial population (62%) but a lower AGG (45%). S3+ was associated very low genotyping needs (4%) but gave both a lower AGG (60%) and $♀_{m+}$ frequency (14%) at the nucleus level.

Discussion

In this study we modeled a discrete generation breeding program to determine the optimal combination of matings when a major mutation affects a selected trait and

Table 2 *Heterozygous frequency of dams in nucleus and commercial populations, annual genetic gain and genotyping needs given the strategy – overlapping generations*

Strategy	Economic return (% of OPT, GC = 1)	Nucleus (% of ♀$_{m+}$)	Com. Pop. (% of ♀$_{m+}$)	AGG (% of OPT, GC = 0)	Genotyping needs (% of OPT, GC = 0)
OPT, GC = 0	–	60	42	100	100
OPT, GC = 1	100	60	51	98	67
S3*m*, GC = 1	94	60	62	45	66
S2, S2_50, GC = 1	95	50	43	95	93
S3 +, GC = 1	83	14	61	60	4

OPT = optimal combination that gave the maximal economic value; GC = genotyping cost (expressed as a proportion of marginal value per slaughtered lamb), GC = 1 corresponds to the current estimated genotyping cost and GC = 0 corresponds to the idealistic situation (no genotyping costs); S2 = strategy based on [++/*m*+] and [*m*+/++]; S2_50 = strategy based on a balanced proportion of [++/*m*+] and [*m*+/++]; S3*m* = strategy based on [++/*mm*]; [*m*+/++] and [*m*+/*mm*]; S3 + = strategy based on [++/++], [++/*mm*] and [*m*+/*mm*]; Com. Pop. = commercial population; AGG = annual genetic gain.
Mating: [dam genotype × sire genotype].

heterozygous carriers are desirable, whereas homozygous carriers are not. In a second step, an overlapping generation model was adapted to the OVI-TEST Lacaune meat sheep breeding program as a real case study. Using sequential quadratic programming methodology, we determined the optimal combination of matings (OPT) that maximized the economic value achieved by the population in the short-term, in the long-term and over the whole time horizon for a range of GEs and GCs. Optimal mating scheme strategy was then compared with simpler strategies based on a limited number of matings; two matings for S2 or three matings for S3+ and S3m.

Parameter values and optimal designs

The optimal combination of matings varied according to the time horizon, GE and GC values assessed. Given the values of parameters, it was more profitable to either increase the AGG, increase the heterozygous frequency of females in both the nucleus and the commercial population, or decrease genotyping needs. Values of each variable were reported for the overlapping model (Table 2) but not for the discrete model as results were consistent between models. At high GE and low GC values, OPT put more emphasis on maximizing the heterozygous frequency. At very low GE and high GC values, the AGG was the unique goal and the mutation eradicated. In between, depending on the combination of the GE and GC parameters and the time horizon considered, OPT varied. The strategy either led to increased emphasis on maximizing the heterozygous frequency in both the nucleus and commercial population or on maximizing the AGG. As expected, strategies that put more emphasis on heterozygous frequencies were superior in the short term, whereas strategies that put more emphasis on AGG were superior in the long term. Over the whole time horizon, for a low GE, OPT was mainly based on [++/*m*+] and [*m*+/++], a combination that favored the AGG as all progeny were considered as candidates for replacement. As the GC increased, the proportion of each type of mating became more balanced. A higher AGG was obtained for a balanced proportion of [++/*m*+] and [*m*+/++] (50% of each type of mating) than those obtained for unbalanced proportions (results from simulations not shown). A balanced proportion decreased the frequency of heterozygous carriers in the nucleus but maintained the frequency in the commercial population, resulting in a higher AGG and decreased genotyping needs. For high GE values, OPT was mainly based on [*m*+/++] and led to a high frequency of heterozygous carriers at the nucleus level. In this case, as non-carrier females and heterozygous males were not candidates for replacement, lower selection intensities resulted in a lower AGG. For higher GC values, [*m*+/++] was associated with [++/*mm*] (and [*m*+/*mm*]): the frequency of heterozygous females decreased at the nucleus level but increased at the commercial population level, whereas genotyping needs decreased. [*m*+/*mm*] was always superior to [*m*+/*m*+] for the production of homozygous carrier males (♂$_{mm}$).

With the overlapping model, the OPT strategy obtained across a range of GC values was based on five different types of mating. Except when GC was null, the proportion of each mating remained stable irrespective of GC. The OPT strategy observed for the overlapping model (time unit = year) was close to the short-term OPT strategy with the discrete model (time unit = generation): at low GC values, OPT was mainly based on [++/*m*+] and [*m*+/++] and then at higher GC values, on [++/*mm*] and [*m*+/++]. At high GC values, [*m*+/++] was replaced by both [++/++] and [++/*mm*] for the discrete model, whereas the combination of matings remained stable for the overlapping model across the range of GC values assessed. Differences in the breeding scheme modeling might explain this result. In the overlapping model, after estrus induction non-pregnant ewes were mated by natural mating with non-carrier males in multiple-sire matings. Due to the use of non-carrier males, the optimal balance between AGG and heterozygous frequency would be different in the overlapping and discrete models, with a higher emphasis on AGG in the overlapping model. To test this hypothesis, we assessed the OPT strategy for a higher GE, using the overlapping model and assuming that a higher GE would rebalance the relative importance of AGG and heterozygous frequency. The optimal combination

we obtained (result not shown) was based on the same types of mating as those obtained for the discrete model for a lower GE. Another difference is due to the intrinsic proprieties of discrete and overlapping generation models: the discrete model quickly reached a steady state compared with the overlapping model in which genetic values fluctuated more in the early stages of the breeding program. Consequently, we also studied OPT for the overlapping model and simpler strategies when the first 15 reproduction cycles were removed; the results (not shown) obtained were consistent with those presented in this paper.

Optimal design and simpler strategy comparison

Optimal mating scheme was compared with simpler mating strategies. The results obtained for both discrete and overlapping models were consistent and were of the same magnitude: compared with OPT, the economic losses were limited for S2 and S3m and higher for S3+. The S2 strategy favored a high AGG, whereas S3m favored a high heterozygous frequency. Both strategies gave similar results in terms of economic value and the differences with OPT were limited. More significant was the increased loss of efficiency for the S3+ strategy as GE increased (GE $\geqslant$1). In S3+, a proportion of the nucleus females was used to produce heterozygous females ([++/*mm*]) and only parents involved in [++/++] matings contributed to the genetic progress of non-carrier females. This strategy led to a reduction in the effective size of the nucleus as a proportion of the female nucleus population exclusively produced newborn females for transfer. An improvement might be obtained by assigning the non-carrier females to [++/++] and [++/*mm*] according to their EBV as most of the female progeny born from [++/*mm*] were transferred to the commercial population.

At the estimated current GC (approximately GC=1), the S2_50 strategy was the best simple strategy for the French Lacaune meat sheep population OVI-TEST. The breeding program is currently close to S2_50 (Martin *et al.*, 2014) and it would be straightforward to implement this strategy based on only two different types of mating. Although the GC considered did not take into account cost of breeder labor, the results were identical for higher GC values. For lower GC values, the best simple strategy was S2, based on the same mating types than S2_50 strategy but with a higher frequency of heterozygous females (60% *v.* 50%). Single-nucleotide polymorphism panels containing both parentage and major gene mutations may become available at lower cost in the coming years. The S3m strategy, as well as S3+ and OPT strategies, require matings [*m*+/*mm*] to produce homozygous carriers. Even if the proportion of [*m*+/*mm*] is low, they could be an issue for flocks of limited size (e.g. [*m*+/*mm*] = 5 for a flock of 500 ewes). These strategies would be based on specialized flocks producing homozygous carrier males (δ_{mm}) and would therefore be more sensitive to animal health problems or nucleus breeder turn-over.

Transition phase and multi-trait breeding goal

The initial allele frequency may be quite different to the optimal or desired frequency. In this case, the transition phase from the initial to the stationary state should be described and optimized using a dynamic model as developed in other studies (Larzul *et al.*, 1997; Dekkers and Van Arendonk, 1998; Manfredi *et al.*, 1998). Contrary to the OVI-TEST breeding program, where parents were born from artificial insemination, managing natural matings would be more complicated and should be taken into account when comparing strategies. In our study, the major gene has an effect on the number of weaned lambs, which was also the selected trait. The genetic and phenotypic evolution of the trait under selection tended to increase the AGG over time: the number of progeny (candidates) increased, whereas the number of selected parents was constant. This phenomenon was not considered in our modeling, assuming limited consequences in the course of the selection scheme. If we included another selected trait, the emphasis on both the heterozygous frequency and the AGG would depend on the relative economic value of the number of weaned lambs and other traits as well as the objective function modeling (number of expressions and relationships among traits): strategies that allow for a higher genetic gain might be promoted. Selection indexes that combine all traits would probably be useful in this situation, and the weight of each trait in the selection index and the mating combination must be included as variables to be optimized. A strategy based on the eradication of the *m* allele with polygenic selection of prolificacy only outperformed others when the GE is low (GE = 0.5) and GC high (GC $\geqslant$ 1.5). Such a strategy could be competitive over a very long time horizon as the wild genotype could reach the desired level of number of weaned lambs by polygenic selection. However, in the case of a substantial GE, the period to reach the desired level might be very long with profitability issues in the short term. In addition, we did not address the potential consequences of the different strategies on inbreeding. Strategies that limit the number of candidates among the progeny produced, such as S3+ or [*m*+/++] only strategies, could have deleterious effects.

Conclusion

In this study, a procedure based on a deterministic genetic model and an optimization algorithm was proposed to determine the optimal mating strategies, in terms of economic returns, to manage heterozygous carriers of a major gene in purebred meat sheep breeding programs. Our results show that the optimal strategy, that is, the optimal combination of matings, varied and put more emphasis either on the AGG, the frequencies of heterozygous females or the genotyping needs depending on the GE, GC and the horizon term considered. In the short term, strategies that favored heterozygous female frequency tend to give higher profit except when the GE was quite low. In the long term, strategies that favored the AGG performed better. Over the whole time horizon, the optimal strategy favored heterozygous frequency when the GE was high and favored the AGG when the GE was lower. Using our model, we computed the economic return of simpler mating strategies.

Comparisons showed that some simpler strategies gave economic returns close to the optimal strategy but were considerably different in terms of AGG.

Acknowledgments

This study was funded by Institut de l'Elevage, INRA and Apis-gène. The authors gratefully acknowledge the funding.

Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731117001835

References

Amer PR, McEwan JC, Dodds KG and Davis GH 1998. Cost benefit analysis of commercial use of the Inverdale prolificacy gene. Proceeding of the New Zealand Society of Animal Production 58, 157–160.

Bodin L, Raoul J, Demars J, Drouilhet L, Mulsant P, Sarry J, Tabet C, Tosser-Klopp G, Fabre S, Boscher MY, Tiphine L, Bertrand C, Bouquet PM, Maton C, Teyssier J, Jouannaux C, Hallauer J, Cathalan D, Gueux J and Pocachard M 2011. Etat des lieux et gestion pratique des gènes d'ovulation détectés dans les races ovines françaises. In 18èmes Rencontres Recherches Ruminants, 7–8 December, Paris, France, pp. 393–400.

Carillier-Jacquin C, Larroque H and Robert-Granié C 2016. Including α *s1 casein* gene information in genomic evaluations of French dairy goats. Genetic Selection Evolution 48, 54.

Cheype A, Guerrier J, Tortereau F, François D, Poivey JP, Chile K and Raoul J 2013. Economical weighting of breeding objectives and definition of total merit indexes in BMC sheep breed. In 64th EAAP Annual Meeting of the European Association for Animal Production, 26–30 August, Nantes, France, Book of Abstracts No. 19, p. 319.

Davis GH, Montgomery GW, Allison AJ, Kelly RW and Bray AR 1982. Segregation of a major gene influencing fecundity in progeny of Booroola sheep. New Zealand Journal of Agriculture Research 25, 525–529.

Dekkers JCM and van Arendonk JAM 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations. Genetical Research 71, 257–275.

de Vries AG, Sosnicki A, Garnier JP and Plastow GS 1998. The role of major genes and DNA technology in selection for meat quality in pigs. Meat Science 49, 245–255.

Drouilhet L, Lecerf F, Bodin L, Fabre S and Mulsant P 2009. Fine mapping of the FecL locus influencing prolificacy in Lacaune sheep. Animal Genetics 40, 804–812.

Elsen J and Mocquot JC 1974. Méthode de prévision de l'évolution du niveau génétique d'une population soumise à une opération de sélection et dont les générations se chevauchent. INRA Bulletin Technique du Département de Génétique Animale 17, 30–54.

Elsen JM 1993. Prediction of annual genetic gain and improvement lag between populations. Genetic Selection Evolution 25, 75–82.

Fontanesi L, Schiavo G, Scotti E, Galimberti G, Calò DG, Samorè AB, Gallo M, Russo V and Buttazzoni L 2015. A retrospective analysis of allele frequency changes of major genes during 20 years of selection in the Italian Large White pig breed. Journal of Animal Breeding and Genetics 132, 239–246.

Gomez-Raya L and Klemetsdal G 1999. Two-stage selection strategies utilizing marker-quantitative trait locus information and individual performance. Journal of Animal Science 77, 2008–2018.

Hill WG 1974. Prediction and evaluation of response to selection with overlapping generations. Animal Production 18, 117–139.

Larzul C, Manfredi E and Elsen JM 1997. Potential gain from including major gene information in breeding value estimation. Genetic Selection Evolution 29, 161–184.

Manfredi E, Barbieri M, Fournet F and Elsen JM 1998. A dynamic deterministic model to evaluate breeding strategies under mixed inheritance. Genetics Selection Evolution 30, 127–148.

Martin P, Raoul J and Bodin L 2014. Effects of the FecL major gene in the Lacaune meat sheep population. Genetics Selection Evolution 46, 48.

McEwan JC 1995. Industry use of the Inverdale gene (Fec X). Proceeding of the New Zealand Society of Animal Production 55, 298–300.

OVI-TEST 2016. 1er centre français d'insémination ovine. Official website. Retrieved on 23 September 2016 from http://www.lacaune-ovitest.com/fr/lacaune-viande-ovi-test.php.

Palhière I, Elsen JM, Astruc JM, Barillet F, Bed'Hom B, Bibé B, Bouix B, Brochard M, Catrou O, Dion F, François D, Griffon L, Jullien E, Orlianges M, Perret G and Tribon P 2003. Breeding for scrapie resistance in France. In Proceeding of the International Workshop on Major Genes and QTL in Sheep and Goat, 8–11 December 2003, Toulouse, France, CD-ROM Communication No. 3-03.

Piper LR and Bindon BM 1982. The Booroola Merino and the performance of medium non-Peppin crosses at Armidale (ed. L. R. Piper, B. M. Bindon and R. D. Nethery). In The Booroola Merino, pp. 9–20. CSIRO, Melbourne, VIC, Australia.

Sánchez A, Ilahi H, Manfredi E and Serradilla JM 2005. Potential benefit from using the αs1-casein genotype information in a selection scheme for dairy goats. Journal of Animal Breeding and Genetics 122, 21–29.

Sodhi SS, Jeong DK, Sharma N, Lee JH, Kim JH, Kim SH, Kim SW and Oh SJ 2013. Marker assisted selection-applications and evaluation for commercial poultry breeding. Korean Journal of Poultry Science 40, 223–234.

Villanueva B, Pong-Wong R, Grundy B and Woolliams JA 1999. Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genetics Selection Evolution 31, 115–133.

Optimal mating strategies to manage a heterozygous advantage major gene in sheep.

J. Raoul[1,2], I. Palhière[2], J.M. Astruc[1], A. Swan[3] and J.M. Elsen[2]

Supplementary material S1

Computation of the number of selected parents per sex and genotype:

The number of selected parents with genotype g and sex a, nb_{ga}, depended on the mating plan: for females, $nb_{g_df} = \sum_{g_s} x_{[g_dg_s]}$, and for males, $nb_{g_sm} = \sum_{g_d} x_{[g_dg_s]} * \frac{nb_m}{nb_f}$ where $\frac{nb_m}{nb_f}$ denotes the ratio of the number of males per female and $x_{[g_dg_s]}$ the number of females with genotype g_d mated to a sire with genotype g_s.

Determination of the single truncation threshold and per sex and genotype and computation of the probability to be selected as a parent for each candidate category.

Using a numerical approach based on the NAG routines C05ADF and G01EAF (The Numerical Algorithms Group (NAG), Oxford, United Kingdom), the single truncation point $K_{ga(t)}$ for selecting parents with genotype g and sex a across matings but within sex-genotype categories was determined numerically solving the equation $nb_{ga} = \sum_{g_dg_s} \omega_{[g_dg_s]ga(t)} * c_{[g_dg_s]ga,t}$ where $\omega_{[g_dg_s]ga(t)}$ denotes the probability of a candidate belonging to category $[g_dg_s]ga$ at time t to be selected as a parent with genotype g and sex a. Assuming a normal distribution of EBVs in each category (for simplification, $z = [g_dg_s]ga$):

$$\omega_{z(t)} = prob\left(\hat{g}_{z(t)} > K_{ga(t)}\right) = \int_{K_{ga(t)}}^{+\infty} \frac{1}{\sqrt{2\pi}*\rho} * e^{-0.5*\left(\frac{u-ebv_{z(t)}}{\rho}\right)^2} du$$

Where $prob\left(\hat{g}_{z(t)} > K_{ga(t)}\right)$ denotes the probability for an animal belonging to the category $[g_dg_s]ga$ to have an EBV at time t above the threshold $K_{ga(t)}$, $ebv_{z(t)}$ and ρ the mean EBV and mean accuracy of EBVs of animals belonging to the category $[g_dg_s]ga$.

Computation of the selection differentials and the proportion of selected parents originating from each candidate category (for simplification, $z = [g_dg_s]ga$)**.**

The selection differential was derived as $d_{z(t)} = \frac{\frac{1}{\sqrt{2\pi}}*e^{-0.5*\left(\frac{K_{ga(t)}-ebv_{z(t)}}{\rho}\right)^2}}{\omega_{z(t)}} * \rho * \sigma_g$ where $K_{ga(t)}$ denotes the threshold to select parents with genotype g and sex a, $ebv_{z(t)}$ and ρ the mean

EBV and mean accuracy of EBVs of animals belonging to the category $[g_d g_s]ga$ and σ_g the genetic standard deviation (polygenic component).

The proportion of selected parents ga at t originating from each candidate category was defined as $q_{z(t)} = \frac{\omega_{z(t)} c_{z(t)}}{nb_{ga}}$.

Computation of the elements of the transition matrix ***P***

By definition, for a given candidate category, all maternal and paternal genetic contribution came from respectively one maternal and one paternal category. Thus the element of ***P*** was derived according to the genetic contribution of candidate categories at t-1 to parental categories. The genetic contribution of the candidate category $[g_d g_s]ga$ at time $t-1$ to the candidate category $[g_d{}' g_s{}']g'a'$ at time t was derived as: $p_{[g_d{}' g_s{}']g'a',[g_d g_s]ga(t)}$=0.5*$\sum_{[g_d g_s]} q_{[g_d g_s]g'a'(t)}$ where $q_{[g_d g_s]g'a'(t)}$denotes the proportion of parents with genotype g' and sex a' at time t originating from mating $[g_d g_s]$ at time $t-1$.

Computation of the number of lambs per female (phenotypic value of the selected trait) at the nucleus, NLFN, and commercial population level, NLFC, according to the evolution of genetic breeding value

$$NLFN_{g(t)} = NLF_{0,g} + \sum_g \sum_{[g_d g_s]} q_{[g_d g_s]gf(t)*} \left(ebv_{[g_d g_s]gf(t)} + d_{[g_d g_s]gf(t)}\right)$$

$$NLFC_{g(t)} = NLF_{0,g} + \sum_g \sum_{[g_d g_s]} \left[\frac{\left(1-\omega_{[g_d g_s]gf(t)}\right)*c_{[g_d g_s]gf(t)}}{\sum_{[g_d g_s]} c_{[g_d g_s]gf(t)} - nb_{gf}} * \left(ebv_{[g_d g_s]gf(t)} - \frac{d_{[g_d g_s]gf(t)}*\omega_{[g_d g_s]gf(t)}}{1-\omega_{[g_d g_s]gf(t)}}\right)\right]$$

where $NLF_{0,g}$ denotes the number of weaned lambs per female prior to selection given the genotype g, $\omega_{z(t)}$ the probability of a candidate belonging to the category z at time t to be selected as a parent, $c_{z(t)}$ the number of animals belonging to the category z at time t, nb_{gf} the number of selected females with genotype g, $ebv_{z(t)}$ the mean EBVs of candidates belonging to the category z at time t and $d_{z(t)}$ the selection differential applied to category z. In this equation, $\frac{\left(1-\omega_{z(t)}\right)*c_{z(t)}}{\sum_i c_{z(t)} - nb_{ga}}$ was the relative proportion of category z among transferred females of genotype g and $\left(ebv_{z(t)} - \frac{d_{z(t)}*\omega_{z(t)}}{1-\omega_{z(t)}}\right)$ their mean EBV. Note that the Genotype effect, GE, was equal to $NLF_{0,g=m+} - NLF_{0,g=++}$.

Supplementary material S2

Details of $Geno_t$ computing

The genotyping needs depended on the number of parents selected from each type of mating. The genotypes of animals born from [++/++] and [++/*mm*] were known without genotyping. For other matings, genotyping needs, computed by sex, depended on the relative number of animals selected of each genotype and originating from the same mating. For example, $\omega_{[++/\mathrm{m}+]++f} * c_{[++/\mathrm{m}+]++f}$ the number of ++ females born from [++/*m*+] to be genotyped depended both of the number of ++ and *m*+ newborns selected. From [++/*m*+], a number of ++ females equal to $\omega_{[++/\mathrm{m}+]++f} * c_{[++/\mathrm{m}+]++f}$ is selected (ω_z is the proportion selected and c_z the number of candidate available). To obtain this number, $\frac{\omega_{[++/\mathrm{m}+]++f} * c_{[++/\mathrm{m}+]++f}}{\beta_{[++/\mathrm{m}+]++}} = 2 * \omega_{[++/\mathrm{m}+]++f} * c_{[++/\mathrm{m}+]++f}$ had to be genotyped. Genotyping needs for females were then equal to $max\left(\frac{\omega_{[++/\mathrm{m}+]++f} * c_{[++/\mathrm{m}+]++f}}{\beta_{[++/\mathrm{m}+]++}}, \frac{\omega_{[++/\mathrm{m}+]\mathrm{m}+f} * c_{[++/\mathrm{m}+]m+f}}{\beta_{[++/\mathrm{m}+]\mathrm{m}+}}\right)$. Genotyping needs at time t were therefore computed as:

$$Geno_t = \sum_a \sum_{[g_d g_s] \neq [++/++];[++/\mathrm{mm}]} \max\left(\frac{\omega_{[g_d g_s]ga(t)} * c_{[g_d g_s]ga(t)}}{\beta_{[g_d g_s]g}}\right).$$

Details of NS_t computing

Two modalities for computing $\boldsymbol{NS_t}$ were retained. Either ++ and *m*+ females born from matings [*m*+/*m*+] and [*m*+/*mm*] were (i) genotyped and transferred or (ii) culled. The choice of the modality depended on their relative economic interest.

If i) outperformed ii) then $NS_t =$

$$\sum_g \sum_{[g_d g_s]} c_{[g_d g_s]gm(t)} - nb_m + \sum_{[g_d g_s]} c_{[g_d g_s]mmf(t)} + \left[\sum_{g=++,m+} \sum_{[g_d g_s]} \left(1 - \omega_{[g_d g_s]gf(t)}\right) * c_{[g_d g_s]gf(t)} * NLFC_{g(t)}\right] * \frac{1}{1+d}$$

In this case, additional genotypes were computed.

If ii) outperformed i) then $NS_t =$

$$\sum_g \sum_{[g_d g_s]} c_{[g_d g_s]gm(t)} - nb_m + \sum_g \sum_{[g_d g_s]=[m+/m+],[m+/mm]} \left(1 - \omega_{[g_d g_s]gf(t)}\right) * c_{[g_d g_s]gf(t)} + \left[\sum_g \sum_{[g_d g_s] \neq [m+/m+],[m+/mm]} \left(1 - \omega_{[g_d g_s]gf(t)}\right) * c_{[g_d g_s]gf(t)} * NLFC_{g(t)}\right] * \frac{1}{1+d}.$$

where $c_{[g_d g_s]ga(t)}$ denotes the number of candidates with sex a, genotype g and originating from dam and sire with genotype g_d and g_s, nb_m the number of selected males, $\omega_{[g_d g_s]gf(t)}$ the proportion of selected female ($a = f$) with genotype g and originating from dam and sire with genotype g_d and g_s, $NLFC_{g(t)}$ the number of weaned lambs per transferred female with genotype g and d the discount rate (the discounted term $\frac{1}{1+d}$ is only applied on the lambs slaughtered from the commercial population as they were performed one generation later).

Profit computing

$Profit = \sum_t (NS_t * 1 - Geno_t * GC) * \left(\frac{1}{1+d}\right)^t$ where NS_t denotes the number of lambs slaughtered at time t, $Geno_t$ the number of genotyping at time t, GC the cost of genotyping (expressed as a proportion of the marginal value per lamb slaughtered) and d the discount rate.

Supplementary material S3

Overlapping model

The overlapping generation model was developed according to the OVI-TEST breeding program design. In addition to candidate categories, parental categories, reported on supplementary figure 1, were defined according to the dam parity and the age and status of sires (progeny testing sire, proven AI sire and natural mating sire). At each time, nb_F =12 000 females were inseminated by AI males (males being progeny tested and proven males) according to the combination of matings assessed, $x_{[g_d g_s]}$, which was constant over time. Following estrus induction, non-pregnant females were mated to ++ natural mating males in multi-sire joining groups to avoid *mm* progeny. No selection differential was considered along the dams-to-breed female pathway, whereas young males were selected among animals born from AI proven sires and sire dams (20% of the best ewes of the nucleus). According to their genotype, young males were selected as described for the discrete model. A fraction of males were progeny tested, and half of these males were then selected as proven sires. Natural mating sires were selected among non-progeny tested males. Non-selected males were slaughtered. Non-selected females within the nucleus were either slaughtered or transferred to the commercial population as parents in accordance with OVI-TEST practice.

Category definition. Candidate categories defined for the discrete model were extended by including dam parity, $l = 1{,}5$. As shown in Supplementary figure S1, additional categories were defined for adults: 12 categories for females, 1 year old and parity 1-5 (6 ages, 2 genotypes); and 21 categories for males, one-year old males (three genotypes), AI sires (5 ages, 3 genotypes) and natural mating sires (3 ages, one genotype). The number of one-year old males ($nb_{ym} = 150$), males in progeny testing ($nb_{ptm} = 30$) and proven males ($nb_{pvm} = 15$) were input parameters (table 1). The number of l parity dams depended on the proportion of l parity dams $\alpha_l = f^l / \sum_{i=1}^{5} f^i$ where f, constant until the maximum age, denotes the proportion of females still present and fertile after each reproductive cycle. The number of weaned lambs in the nucleus was computed as the sum of the initial prolificacy of non-carriers, the GE and the average EBV of the category. The number of animals per candidate category was computed as $c_{[g_d g_s] g la(t)} = x_{[g_d g_s]} * \beta_{[g_d g_s]g} * NLFN_{g_d l(t-1)} * \alpha_l * f_{AI} * 0.5$, where $x_{[g_d g_s]}$ denotes the number of dams with genotype g_d mated to sires with genotype g_s, $\beta_{[g_d g_s]g}$ the proportion of progeny with genotype g originating from the mating $[g_d g_s]$, α_l, the proportion of l parity dams, f_{AI} the female fertility after AI and $NLFN_{g_d l(t-1)}$ the number of weaned lambs per nucleus female of parity l and genotype g_d at $t-1$.

Genetic model. To estimate EBVs over time, a classical gene flow approach, $\mathbf{ebv_t} = \boldsymbol{P} * \mathbf{ebv_{t-1}} + \boldsymbol{d_t}$, was used. To compute elements of the transition matrix $\boldsymbol{P}$ and the selection differential for the young and proven male categories, we assumed a within genotype single truncation selection (as described in the discrete model) based on approximate EBV reliabilities. For newborn male candidates, the reliability (${\rho^2}_{[g_d g_s]glm}$) was computed on parental information and depended on their dam's parity (l) and the average of their sires' reliabilities (as proven males reliabilities were close, we only distinguished the paternal genotype but not the age of candidate sire):

${\rho^2}_{[g_d g_s]glm} = 0.25 * \left(\frac{l*h^2}{1+l*rep} + \sum_j p_{[g_d g_s]glm,j} * \frac{nd_j}{nd_j + \frac{4}{h^2-1}} \right)$ where h^2 and rep denote the heritability and repeatability of NLF, $p_{[g_d g_s]glm,j}$ the genetic contribution of sire category j at $t-1$ to the candidate category $[g_d g_s]glm$ at t, nd_j the average number of daughters per sire category j. The number of selected young males per genotype was $nb_{g_s m} = \sum_{g_d} x_{[g_d g_s]} * \frac{nb_{ym}}{nb_f}$ where nb_{ym} denotes the total number of young males selected per time. The selected proportion was equal to $\frac{nb_{gm}}{0.1 * \sum_l \sum_{[g_d g_s]} c_{[g_d g_s]glm(t)}}$ where 0.1 indicates the proportion of available candidates for selection (90% of male lambs were culled for non-genetic purposes). The probability that a candidate of z being selected $\omega_{z(t)}$, the proportion of gm parents selected from z $q_{z(t)}$, and the selection differential $d_{z(t)}$ were computed as described in the discrete model. For proven males, the reliability was exclusively computed on progeny information: ${\rho^2}_{PT} = \frac{nd_{PT}}{nd_{PT} + \frac{4}{h^2-1}}$ where nd_{PT} denotes the average number of daughters per male in progeny testing. At each time, 50% of males in progeny testing were selected as proven males. For females ($a = f$), selected newborns were randomly chosen among AI candidates, thus the proportion of selected female for a given candidate category was $q_{[g_d g_s]glf(t)} = \frac{c_{[g_d g_s]glf(t)}}{\sum_{[g_d g_s]} \sum_l c_{[g_d g_s]glf(t)}}$. The transition matrix $\boldsymbol{P}$ was build according to category definition and the selection process implemented.

The element $p_{k,j}$ represented the genetic contribution of the category j at $t-1$ to the category k at t. The $k = 1, nc$ rows and $j = 1, nc$ columns corresponded to:

$k\ (j) = 1{,}90 : [g_d g_s]glf$ female candidate categories at $t\ (t-1)$

$k\ (j) = 103, 192$: $[g_d g_s] glf$ male candidate categories at $t\ (t-1)$

candidates Males $k\ (j)$	Females $k\ (j)$	$[g_d g_s]$	l	g
103	1	++/++	1	++
104	2	++/m+	|	|
105	3	++/m+	|	|
106	4	m+/++	|	|
107	5	m+/++	|	|
108	6	m+/mm	1	++
109	7	++/++	2	++
|	|	|	|	|
115	12	m+/mm	2	++
|	|	|	|	|
133	30	++/++	1	m+
|	|	|	|	|
139	36	m+/mm	1	m+
|	|	|	|	|
|	|	|	|	|
192	90	m+/mm	5	mm

$k\ (j) = 91{,}96$: ++ females from one year old to parity 5 at $t\ (t-1)$

$k\ (j) = 97{,}102$: m+ females from one year old to parity 5 at $t\ (t-1)$

$k\ (j)$	l	g	
91	0	++	one year old females
|	|	|	
96	5	++	females in parity 5
97	0	m+	one year old females
|	|	|	
102	5	m+	females in parity 5

$k\ (j) = 103{,}192$: $[g_d g_s] glm$ male candidate categories at $t\ (t-1)$

$k\ (j) = 193{,}195$: ++, m+, mm 1 year old male categories at $t\ (t-1)$

$k\ (j) = 193{,}198$: Natural mating male (aged from 2 to 4) categories at $t\ (t-1)$ only ++

$k\ (j) = 199{,}201$: male in progeny testing categories at $t\ (t-1)$

$k\ (j) = 202{,}204$: male awaiting progeny results categories at $t\ (t-1)$

$k\ (j) = 205{,}213$: proven male (age from 4 to 6) categories at $t\ (t-1)$

$k\ (j)$	$g_s{}'$	$\nu_{jfg_s{}'}$	$\nu_{jmg_s{}'}$	
193	++			
194	m+	0	0	one year old male
195	mm			
196	++	0	0	natural mating male, 1st use
|	|	0	0	natural mating male, 2nd use
198	++	0	0	natural mating male, 3rd use
199	++			
|	|	0.40	0	AI male in Progeny Testing
201	mm			
202	++			
|	|	0	0	AI male waiting for Progeny Results
204	mm			
205	++			
|	|	0.27	0.45	AI proven male, 1st use
207	mm			
208	++			
|	|	0.20	0.35	AI proven male, 2nd use
210	mm			
211	++			
|	|	0.13	0.2	AI proven male, 3rd use
213	mm			

$\nu_{ja g_s{}'}$, the contribution sire category j of genotype $g_d{}'$ to candidate progeny of sex a.

Contribution of female (j = 92-96, 98-102) categories at t-1 to female (k = 1,90) and male (k = 103,193) candidate categories at t

Given category definitions, only one dam category contributed to one candidate category. The parental contribution of female category gl (genotype g and parity l) at $t-1$ to candidate category $[g_d{}' g_s{}']g'l'a$ at t was equal to 0.5 for $l' = l$ and $g_d{}' = g$. For example, the contribution of the female category $j = 99$ $(g = g_d = m+, l = 2)$ to candidate category was equal to 0.5 for candidate categories $k = (10; 11; 40; 41; 42; 71{,}72)$.

Contribution of male categories at t-1 to female (k = 1,90) and male (k = 103,193) candidate categories at t

Several sire categories contributed to female (males in progeny testing and proven males) and male (proven males) candidate categories according to their relative use. Therefore, an additional input parameter, $\nu_{ja g_s{}'}$ was considered as the contribution of sire category j of genotype $g_s{}'$ to candidate progeny of sex a.

$p_{[g_d g_s]gla,j} = 0.5 * \nu_{ja g_s{}'}$ for $g_s{}' = g_s$

Contribution of female (male) categories at t-1 to female (male) categories at t.

One year old females (males) with genotype g at t came from candidate females (males) with genotype g at $t-1$

$P_{91,[g_d g_s]++la} = q_{[g_d g_s]++la(t)}$ and $P_{92,[g_d g_s]m+la} = q_{[g_d g_s]m+la(t)}$ for females ($a = f$)

$P_{193,[g_d g_s]++la} = q_{[g_d g_s]++la(t)}$, $P_{194,[g_d g_s]m+la} = q_{[g_d g_s]m+la(t)}$ and $P_{195,[g_d g_s]mmla} = q_{[g_d g_s]mmla(t)}$ for males ($a = m$)

For older female and male categories, the transition matrix reflected aging by having only one element of each row equal to 1. For example, females belonging to the female category with genotype ++ and parity 1 at t belonged to the one year old female category with genotype ++ at $t-1$, $p_{93,91} = 1$.

Profit function. The profit function was adapted to the real case of the OVI-TEST breeding program. Non-selected males were slaughtered as well as non-selected females born from matings [*m+/m+*] and [*m+/mm*] to avoid the transfer of *mm* females to the commercial population (no transferred females were genotyped). In accordance with the real proportions observed, half of females born from natural mating sires were transferred to the commercial population and half were slaughtered. Their number, nb_{NBg}, according their genotype g was computed as:

$$nb_{NBg} = (1 - f_{AI}) * f_R * \sum_l \alpha_l * \left(\sum_{g_d=++} x_{[g_d g_s]} * NLFE_{g_d l(t-1)} + 0.5 \sum_{g_d=m+} x_{[g_d g_s]} * NLFE_{g_d l(t-1)} \right)$$ for g =++ and

$$nb_{NBg} = (1 - f_{AI}) * f_R * \sum_l \alpha_l * 0.5 \sum_{g_d=m+} x_{[g_d g_s]} * NLFE_{g_d l(t-1)}$$ for g =m+

where f_{AI} denotes the fertility on induced estrus, α_l the proportion of females in parity l, $x_{[g_d g_s]}$ the number of dams with genotype g_d mated to a sire with genotype g_s, $NLFE_{g_d l(t-1)}$ the number of lambs per female of genotype g_d at $t-1$ on natural estrus and f_R the fertility on natural estrus of ewes that failed to become pregnant on induced estrus. We assumed a delta of -0.1 in prolificacy between natural and induced estrus. The prolificacy, or number of lambs per female, was computed as the sum of initial prolificacy, the genotype effect and half of dam's EBV (weighted by the parity contribution) and half of sire's EBV (weighted by the sire contribution to females $v_{jf g_s'}$). For non-selected females born from AI, 75% were transferred and 25% slaughtered. The genotyping needs and costs as well as the discounted revenues were computed using the methodology described for the discrete model. The profit function was computed from time horizon 2 to 21.

Supplementary material S4

Optimization: description of linear constraints

As the population was constant over time, the range of variation was from 0 to nb_f for each mating. Two types of constraint were defined. First, at any t, the sum of matings was equal to the number of females: $\frac{1}{nb_F} * \sum_{[g_d g_s]} x_{[g_d g_s]} = 1$. Second, for any genotype*sex ga, available candidates produced at $t-1$ had to be sufficient to select parents (or replacements) at t. The latter constraints were assumed to be linear as we did not consider the increase in prolificacy due to the genetic progress. This resulted in defining a linear inequality linked to each category with sex a and genotype g: the left term represents the number of parents to be selected and the right term to the number of candidates. For the discrete model, the inequalities were:

$$x_{[++/++]} + x_{[++/m+]} + x_{[++/mm]} \leq \sum_{[g_d g_s]} x_{[g_d g_s]} * \beta_{[g_d g_s]++} * NLF_{0,g=g_d} * 0.5 \quad \text{for } ♀_{++},$$

$$x_{[m+/++]} + x_{[m+/m+]} + x_{[m+/mm]} \leq \sum_{[g_d g_s]} x_{[g_d g_s]} * \beta_{[g_d g_s]m+} * NLF_{0,g=g_d} * 0.5 \quad \text{for } ♀_{m+},$$

$$x_{[++/++]} + x_{[m+/++]} \leq \sum_{[g_d g_s]} x_{[g_d g_s]} * \beta_{[g_d g_s]++} * NLF_{0,g=g_d} * 0.5 * \frac{nb_f}{nb_m} \quad \text{for } ♂_{++},$$

$$x_{[++/m+]} + x_{[m+/m+]} \leq \sum_{[g_d g_s]} x_{[g_d g_s]} * \beta_{[g_d g_s]m+} * NLF_{0,g=g_d} * 0.5 * \frac{nb_f}{nb_m} \quad \text{for } ♂_{m+},$$

$$x_{[++/mm]} + x_{[m+/mm]} \leq \sum_{[g_d g_s]} x_{[g_d g_s]} * \beta_{[g_d g_s]mm} * NLF_{0,g=g_d} * 0.5 * \frac{nb_f}{nb_m} \quad \text{for } ♂_{mm},$$

where $x_{[g_d g_s]}$ denotes the number of dams with genotype g_d mated to a sire with genotype g_s, $\beta_{[g_d g_s]g}$ the proportion of progeny with genotype g originating from the mating $[g_d g_s]$, $NLF_{0,g}$ the number of weaned lambs prior to selection given the genotype g and nb_f and nb_m the number of dams and sires.

For the overlapping model, we extended these inequalities by multiplying the left term by α_1, the proportion of female in first parity and by computing, in the right term, the number of candidates produced across dam parity categories.

Article 3 (court) : Optimal and practical strategies to manage an ovulation rate mutation located on the X chromosome in a French sheep breed

J. Raoul, L. Bodin, J.M. Elsen and A.A. Swan. 2017. Proc. Assoc. Advmt. Anim. Breed. Gen., Townsville, Australia, 22.

OPTIMAL AND PRACTICAL STRATEGIES TO MANAGE AN OVULATION RATE MUTATION LOCATED ON THE X CHROMOSOME IN A FRENCH SHEEP BREED

J. Raoul[1,2,3], L. Bodin[2], J.M. Elsen[2] and A.A. Swan[3]

[1] Institut de l'Elevage, Castanet-Tolosan, France
[2] GenPhySE, INRA, Castanet-Tolosan, France
[3] Animal Genetics and Breeding Unit, University of New England, Armidale, Australia

SUMMARY

In French sheep breeding programs, several mutations affecting ovulation rate have been discovered. For mutations located on the X chromosome, the optimal management of such genes is still a challenge because nucleus flocks are small compared to Australian or New Zealand ram breeding flocks. A deterministic model was developed, and using sequential quadratic programming methodology, the combination of mating types that maximized the profit across a range of genotype costs was determined. Results show that even if losses of genetic gain were quite high compared to the gain without the major gene, the optimal use of an ovulation rate mutation located on the X chromosome was beneficial. At the current costs, the optimal strategy that gave the maximal profit was based on four different mating types. A strategy based on only the use of carrier females mated to non-carrier males gave similar results to the optimal strategy in terms of profit and genetic gain. This strategy could be adopted by French breeding programs where this kind of mutation segregates.

INTRODUCTION

The number of lambs produced per female has a large impact on profitability in meat oriented sheep production. Several mutations affecting ovulation rate, and thus number of lambs, have been identified. For example, Booroola (Piper and Bindon 1982; Davis *et al.* 1982), BMP15-Inverdale (Davis *et al.* 1982) or BMP15-Grivette (Demars *et al.* 2013), and GDF9-Cambridge (Hanrahan *et al.* 2004). Most often, these polymorphisms have a positive effect on heterozygous carrier productivity. However, in homozygous ewes, these polymorphisms lead to sterility or excessive prolificacy and high rates of neonatal lamb mortality. Therefore homozygous females are undesirable for commercial production.

Several strategies can be implemented to manage these mutations, as outlined by Amer *et al.* (1998) for mutations carried by the X chromosome (i.e. Inverdale gene) and Raoul *et al.* (2017) for mutations carried by an autosomal chromosome: the proportion of each parental genotype is defined according to the sex and matings organised. These balance high frequency of heterozygous females with genetic gain. Increasing the frequency of heterozygotes leads to a change in the proportion of available candidates which affects the overall selection differential of parents and consequently genetic gain. Amer *et al.* (1998) assessed two strategies to manage the Inverdale gene and found that depending on the strategy implemented, the loss of genetic gain was either 24%, or less than 5% compared to the gain without major gene. In the case of an autosomal polymorphism, strategies that enhance either genetic gain or heterozygous female frequency gave equal profit (Raoul *et al.* 2017) and were affected by the genotyping cost per animal.

In the French meat sheep production context, the average number of ewes per nucleus flock is about 300. With such limited flock sizes implementing a strategy which comprises a small proportion of a given mating type (less than 10%) is difficult. It is not practical at a single flock level, but could be organized via specialization of several nucleus flocks in which different flocks focus on a specific mating. This is difficult to co-ordinate, so for practical reason, French breeders would much prefer strategies based on at most two mating types. Strategies outlined for autosomal mutation management have already been discussed for French breeding programs (Raoul *et al.*

2017). The aim of this study is, for the case of a mutation carried by the X chromosome, to determine the combination of mating types that provide the maximal profit (optimal strategy) according to various genotyping costs. This optimal strategy will be compared with more practical strategies in terms of profit and genetic gain.

MATERIALS AND METHODS

A nucleus population representative of a typical French breeding program based on natural mating was modelled. A maternal production trait expressed once per year during female's reproductive life was considered as the only selected trait (*e.g.* milk production estimated through lamb weight at 30 days). Each year, 8000 ewes were mated to 200 rams. Because homozygous carrier females were not used for reproduction, 2 genotypes, non-carriers and carriers were respectively considered for males ([+] and [m]) and females ([++] and [m+]) leading to 4 mating types: 1) ♀ [++] x ♂ [+], 2) ♀[++] x ♂ [m], 3) ♀ [m+] x ♂ [+] and 4) ♀ [m+] x ♂ [m]. As the flock management was assumed to be in a steady-state, the proportion of each mating type across time was constant. The newborn candidates were divided into categories according to their parental genotypes (*i.e.* 4 matings), their sex and their own genotype (2 genotypes for males and 3 genotypes for females). Generations were overlapping and the maximum reproductive life was 6 years for males and females, with a maximum parity of 5 (i.e., from 2 to 6 years of age), leading to a replacement proportion close to 24%.

At each generation, new parents were selected within sex*genotype categories by truncation selection on EBVs: 4 truncation thresholds (2 per parental genotype) were determined across the candidate EBV distributions. For example, [++] female replacement were selected from progeny of mating types 1 and 3. Considering dam parity, these female were selected across 10 EBVs distributions. Whatever their parental genotype or dam's age, we selected females whose EBV was above the unique truncation threshold. Given those thresholds, selection differential and genetic contribution to the next generation (*i.e.* probability of gene origin) were calculated for each candidate category. Evolution of genetic values of parents and their progeny across time for the maternal trait was derived using the gene flow methodology proposed by Hill (1974): a transition matrix representing the gene flow from categories at year t to categories at year t+1 was built from genetic contributions to newborns and accounting for ageing of parents.

Discounted revenues and costs were computed for each cycle (year). The revenues were proportional to the number of lambs sold per year which was equal to the number of live lambs produced minus the number selected for replacement, and the number of live lambs produced by ewes transferred to a commercial flocks. The costs included genotyping costs made at the nucleus level and proportional breeding costs per ewe (nucleus and transferred ewes). It was assumed that 50% of newborn females would still be available after parent selection, and these surplus females would be transferred to a commercial flock where they could be retained for up to 5 parities. These female were not genotyped and only females from mating types 1, 2 and 3 were transferred. It was assumed that independently of their genotype, the selected maternal trait was related to the cost per ewe, because the trait was determined based on milk production, with higher production levels reducing feed costs per lamb. The overall profit was computed as the sum of discounted revenues minus costs over a long-term time horizon (year 5 to year 30). This overall profit was assessed for the following sets of parameters: number of lambs produced = 1.5 for non-carrier females, and +0.5 additional lambs for heterozygous females. Given the fertility, the lamb viability (higher for lambs born from non-carrier), the number of lambs weaned per ewe joined for non-carrier and carrier ewes were 1.22 and 1.44 respectively. The income per lamb sold was assumed to be constant and the production cost per lamb depended on the dam's genetic value for the selected trait and genotype. Three genotyping costs were tested: no cost, 10 and 20 € per genotyped animal.

For a given genotyping cost, the relative proportion of mating types that gave the maximum profit (the optimal strategy) was determined using an algorithm based on sequential quadratic programming methodology. The gain in the absence of the major gene and two simplified strategies was also assessed based on 1 mating only, ♀ [m] x ♂ [++] (S1, corresponding to the "self-sustaining scheme" outlined by Amer (1998)) or 2 mating types, ♀ [++] x ♂ [m] and ♀ [m+] x ♂ [+], (named S2). The proportion of each mating types of these strategies is shown in Table 1.

Table 1: Proportion (%) of each mating type of alternative strategies assessed for the management of an ovulation rate mutation[1] located on the X chromosome.

Mating type	♀[++]x♂[+]	♀[++]x♂[m]	♀[m+]x♂[+]	♀[m+]x ♂[m]
Gain without major gene	100	0	0	0
S1	0	0	100	0
S2	0	60	40	0

[1] Biallelic locus (X chromosome) influencing the number of lambs per female (1.5 for [++] and 2.0 for [m+]).

RESULTS AND DISCUSSION

Table 2 gives the proportion of each mating type in the nucleus that maximizes profit according to the genotyping cost. Results show that when genotyping costs were not included (cost=0), the best strategy was to bred only carriers females and mate them to non-carrier males.

Table 2: Percentage of each mating type in the optimal strategy to manage an ovulation rate mutation[1] located on the X chromosome, according to three genotyping costs (€).

		Mating type			
	genotyping costs	♀[++]x♂[+]	♀[++]x♂[m]	♀[m+]x♂[+]	♀[m+]x ♂[m]
optimal strategy	0	0	0	100	0
	10	21	49	12	18
	20	39	57	0	4

[1] Biallelic locus (X chromosome) influencing the number of lambs per female (1.5 for [++] and 2.0 for [m+]).

For a genotyping cost equal to 10 €, the optimal strategies combined the 4 mating types. The main mating type was non-carrier females mated to carrier males (49% of all matings). In this strategy 30% of the nucleus females were carriers. For a genotyping cost equal to 20, the proportion of non-carrier females mated to carrier males reached 57%. The proportion of carrier females in the nucleus reduced to 4% which corresponded to the minimum requirement to replace carrier males and produced heterozygous females transferred to commercial flocks.

Table 3 shows the genetic gain achieved by the nucleus for all strategies assessed, the genotyping requirements, the frequencies of heterozygous females (nucleus and transferred) and the profit. Apart from the heterozygous frequencies, all results are expressed relative to values obtained for the optimal strategy when there was no genotype cost (=100 in the first row of Table 3).

Results show that when genotyping costs were not included, the optimal strategy maximized the heterozygous female frequency in the nucleus. In this case, a proportion of m+ females were selected for the nucleus, whereas all ++ females were available for transfer. This lead to a reduction in the heterozygous frequency of transferred females to 24%. When genotyping costs were included, the strategy maximized the heterozygous frequency of transferred females. In this case, mating type 2 (♀[++]x ♂[m]) which produces m+ females without genotyping was used, allowing production of heterozygous females to be transferred to a commercial flock. For a moderate genotyping cost (10€), the number of genotyping remained at a significant level and allowed implementation of a strategy providing a substantial genetic gain. For a high genotyping cost, the number of genotypes was very

low and limited to genotyping male progeny of the mating ♀[m+]x ♂[m] only, implemented to replace male carriers. Even if this mating produced homozygous carrier females which were culled, it allowed a higher genetic value of carrier males and a higher genetic gain compared to the use of the mating ♀[++]x ♂[m] This strategy maintained the high proportion of heterozygous females transferred to a commercial flock and limited losses in genetic gain.

Table 3: Genetic gain, genotyping needs, heterozygous female frequencies and profit of various strategies according to the genotyping costs (€).

	Geno. Costs	Genetic gain[1]	Genotyping requirements[2]	Het. freq (nucleus)	Het. freq (transferred)	Profit[3]
Optimal strategy	0	100.0	100	1.00	0.24	100.0
	10	100.4	27	0.29	1.00	79.1
	20	85.1	4	0.04	1.00	74.6
Gain without major gene	-	125.4	0	0.00	0.00	72.5
S1[4]	0	100.0	100	1.00	0.24	100.0
	10	100.0	100	1.00	0.24	77.1
	20	100.0	100	1.00	0.24	54.2
S2[5]	0	103.5	59	0.40	1.00	85.8
	10	103.5	59	0.40	1.00	72.4
	20	103.5	59	0.40	1.00	58.9

[1] 100=genetic gain obtained for the optimal strategy at null genotyping costs
[2] 100=number of genotype for the optimal strategy at null genotyping costs
[3] 100= profit obtained for the optimal strategy at null genotyping costs
[4] Simplified strategy based on one mating type ♀ [m] x ♂ [++]
[5] Simplified strategy based on two mating types ♀ [++] x ♂ [m] and ♀ [m+] x ♂ [+]

The genetic gains for the S1 and S2 strategies were similar to those obtained for optimal strategies, and losses of genetic gain ranged from 22 to 25%, compared to gain without the major gene, similar to the results obtained by Amer *et al.* (1998). Profit obtained for S1 was higher than S2 except at the high genotyping cost. In this case, simple management of the mutation gave lower profit than its eradication. Given the current genotyping cost, approximately 10 €, S1 is a strategy which could be considered for French breeding programs. This strategy has quite high genotyping requirements (two genotyped animals per selected replacement) but results in profitability similar to the optimal strategy and a high productivity in the nucleus flocks. The use of a tool combining parentage assignment and mutation genotyping, which is available in France, would decrease the genotyping cost and make application the S1 strategy more attractive.

REFERENCES

Amer P.R., McEwan J.C., Dodds K.G. and Davis G.H. (1998) *Proc. NZ Soc. Anim. Prod.*, **58**: 157

Piper L.R. and Bindon B.M. (1982) *In the Booroola Merino*, **9:**20. CSIRO, Melbourne

Davis G.H., Montgomery G.W., Allison A.J., Kelly R.W. and Bray A.R. (1982) *NZ J. Agric. Res.* **25**: 525–529

Demars J., Fabre S., Sarry J., Rossetti R., Gilbert H., Persani L., Tosser-Klopp G., Mulsant P., Nowak Z., Drobik W., Martyniuk E. and Bodin L. (2013) *PLoS Genetics* **9:**4

Hanrahan J.P., Gregan S.M., Mulsant P., Mullen M., Davis G.H., Powell R., Galloway S.M (2004) *Biol. Reprod.* **70**:900

Hill W.G. (1974). *Animal Production*, **18**:117

Raoul J., Palhière I., Astruc J.M., Swan A.A. and Elsen J.M. (2017) submitted to *Animal*

Chapitre 4 – Intérêt technique et économique d'un panel SNPs très basse densité pour la sélection génomique en ovins

Résumé des articles 4 et 5

L'intérêt de la sélection génomique repose dans la mise à disposition précoce de prédictions génomiques pour les candidats à la sélection, notamment pour les caractères exprimés uniquement par les femelles. Dans un programme de sélection classique avec testage, les mâles évalués sur descendance ont des prédictions précises de leur valeur génétique mais sont utilisés tardivement. Dans un programme de sélection génomique, les jeunes mâles ont généralement des prédictions génomiques dont la précision est intermédiaire entre précisions sur ascendance et descendances obtenues classiquement. Selon la précision des prédictions génomiques, le raccourcissement de l'intervalle de génération induit par l'utilisation précoce des mâles peut générer un progrès génétique supérieur à une situation ou les mâles sont évalués sur descendance.

Les prédictions génomiques reposent sur une population de référence, c'est-à-dire d'individus dont on dispose à la fois d'information phénotypique et du génotype. Pour de nombreuses populations ovines, l'établissement de cette population de référence reste une difficulté, notamment lorsque les effectifs du noyau de sélection sont faibles et que les phénotypes des mâles sont peu informatifs à cause d'un faible nombre de filles contrôlées par mâle. L'efficacité de la sélection génomique est également modulée par le nombre de candidats génotypés (*i.e.* l'intensité de sélection avec laquelle les reproducteurs sont sélectionnés). Pour des raisons de coûts, la plupart des programmes de sélection qui ont adopté un schéma

génomique utilisent des puces basses densités (de l'ordre de la dizaine de milliers de SNPs) pour génotyper les candidats.

Dans cette étude, une sélection génomique intégrant l'utilisation d'un panel SNP très basse densité a été évaluée. Afin d'augmenter la taille de la population de référence, les femelles ont été génotypées à l'aide du panel très basse densité. Ces femelles étaient filles et petites-filles de mâles génotypés en moyenne densité (50 000 SNPs). L'obtention des génotypes moyenne densité des filles par imputation était envisageable compte-tenu de la présence de leur ascendant paternel dans la population de référence pour l'imputation. La présence des pères limite effectivement l'effet défavorable du très faible nombre de marqueurs sur la précision de l'imputation. A l'aide de simulations stochastiques, un schéma classique basé sur la sélection sur descendance des mâles d'insémination a été comparé à des schémas génomiques variant par la composition de la population de référence (avec ou sans les mères) et le type de génotypes utilisés (moyenne densité ou moyenne densité imputée pour les candidats mâles et mères). Les objectifs de cette étude étaient donc, pour des noyaux de sélection ovins d'effectifs limités, de quantifier les conséquences sur le progrès génétique et la maîtrise de la consanguinité de i) l'accroissement de la taille de la population de référence avec des génotypes de femelles et ii) l'imputation des génotypes des candidats mâles et femelles de la très basse densité (≤ 1000 SNPs) vers la moyenne densité.

Les individus ont été simulés et leurs génomes assimilés à des haplotypes réels phasés issus de génotypes de moyennes densités (≈50 000 SNPs) de mâles d'insémination de la race Lacaune lait. Le sexe de chacun des individus était affecté aléatoirement et les générations suivantes obtenues par accouplements aléatoires entre 1200 femelles et 100 mâles par cycle. Le génome des individus générés à chaque cycle était issu de la réunion de deux haplotypes, maternel et paternel, obtenus après recombinaisons (dont le nombre moyen suivait une loi de

Poisson d'espérance 1 Morgan et les localisations suivaient une loi uniforme) et mutations (taux de $1x10^{-6}$ et localisation suivant une loi uniforme). Une population de fondateurs de 5000 femelles avec une structure pedigree était ainsi obtenue. Cette population était subdivisée en dix groupes (*i.e.* troupeaux), chaque lignée femelle étant affectée à un seul groupe. Afin de simuler les phénotypes des individus, un millier de marqueurs était aléatoirement choisi comme QTL de l'unique caractère en sélection, et leur effet simulé par tirage dans une loi Gamma. Pour chaque femelle, les phénotypes étaient simulés en fonction de leur valeur génétique vraie (somme des effets aux QTL) et de l'héritabilité (h^2=0.25) et répétabilité (rep=0.50) du caractère.

La mise en place d'une sélection génomique devant se faire pour des populations jusqu'alors sélectionnées classiquement, et dont le génome présentait une structure du déséquilibre de liaison de populations en sélection, dix années de sélection classique ont ensuite été appliquées. Les femelles sortaient aléatoirement de la population en fonction de paramètres démographique (probabilité de sortie) dépendant de leur âge. Un phénotype était simulé annuellement pour les femelles toujours présentes et une évaluation BLUP modèle animal pour le caractère sélectionné était réalisée pour l'ensemble de la population. L'évaluation était réalisée avec le logiciel Blupf90 développé par Misztal *et al.* (2009). Dans ce design les jeunes femelles de renouvellement n'étaient pas sélectionnées mais choisies préférentiellement parmi les pères d'insémination. Environ 600 jeunes mâles étaient présélectionnés par troncature sur les index parentaux et en pratique issus des accouplements raisonnés entre les meilleurs femelles et les meilleurs mâles. Chaque année, les dix meilleurs mâles sur ascendance étaient évalués sur descendance par insémination (mâles en testage). Après la première production de leurs filles, les cinq meilleurs mâles étaient sélectionnés sur index comme améliorateurs et utilisés pendant un maximum de quatre années. Les mâles de

monte naturelle étaient sélectionnés sur index parentaux parmi les candidats non retenus pour le testage et utilisés pendant un maximum de quatre années.

Cette période de sélection classique était poursuivie par quinze années de sélection classique ou sélection génomique. Le type de sélection n'avait pas d'influence sur le mode de choix des femelles de renouvellement. Dans le schéma génomique, les 600 jeunes mâles étaient présélectionnés sur index génomiques parentaux évalués par single step GBLUP avec le logiciel Blupf90. Environ 270 jeunes mâles génotypés étaient annuellement disponibles compte-tenu de la sélection des aptitudes fonctionnelles (50% d'élimination) et la mortalité (10%). Dix mâles étaient sélectionnés sur index génomique puis utilisés par insémination pendant un maximum de deux années. Les mâles de monte naturelle étaient sélectionnés sur index génomique parmi les candidats non retenus pour l'insémination et utilisés pendant un maximum de quatre années. Les scénarios de sélection génomique évalués variaient selon la quantité et qualité d'information génomique. Les pères étaient toujours génotypés en moyenne densité et les jeunes candidats mâles en moyenne densité ou en très basse densité puis leurs génotypes imputés vers la moyenne densité (« moyenne densité imputés »). Certains scénarios incluaient les génotypes des mères, moyenne densité ou bien « moyenne densité imputés ». Le panel très basse densité a été sélectionné parmi les marqueurs non retenus comme QTL et dont la fréquence de l'allèle mineur est supérieure à 0.25. Les phases d'imputation ont été réalisées avec le logiciel Fimpute développé par Sargolzaei *et al.* (2009). L'ensemble des scénarios a été évalué sur la base de 50 réplications.

L'application d'un schéma génomique a augmenté le progrès génétique annuel de 26% par rapport au schéma classique lorsque la population de référence était composée de génotypes moyenne densité des pères et les candidats mâles génotypés en moyenne densité. Lorsque les candidats mâles disposaient de génotypes « moyenne densité imputés », le surplus

de progrès génétique a été réduit mais restait conséquent avec +22%. La prise en compte du génotype des mères dans l'évaluation permettait d'accroître le progrès génétique annuel de 54% dans le cas de génotypes moyenne densité et de 42% dans le cas de génotypes « moyenne densité imputés » Ces résultats de scénarios avec imputation étaient basés sur l'utilisation d'un panel de 1000 SNPs et de phases d'imputation valorisant l'information généalogique. L'utilisation d'une densité inférieure ou bien l'absence de valorisation de l'information généalogique réduisait voire annulait le surplus de gain génétique observé.

Dans le second article, le modèle stochastique a été utilisé pour étudier la sensibilité des résultats à la taille de la population de référence initiale quand seuls les pères étaient génotypés en moyenne densité et les candidats en très basse densité. Cette population contenait les pères de l'ensemble des candidats présents lors de la mise en place de la sélection génomique (année 10) et les autres pères présents aux 2, 4, 6, 8 ou 10 cycles de reproduction précédents. Ensuite une comparaison des schémas génomique et classique à coût constant a été réalisée pour trois coûts de génotypage unitaire très basse densité. La combinaison taille de la population historique*nombre de candidats génotypés annuellement qui maximisait le progrès génétique annuel a été déterminé pour le schéma génomique puis comparé au progrès du schéma classique.

Malgré une légère érosion du progrès génétique annuel lorsque la taille de la population de référence diminuait, les résultats obtenus pour un horizon de 15 années étaient proches et montrent que le progrès était peu sensible à la taille initiale de la population de référence tant que les pères étaient présents. Le progrès génétique était également peu sensible à la gamme de coût unitaire du génotypage très basse densité considéré dans l'étude. A coût constant, le progrès génétique du schéma génomique était toujours supérieur (+15 à +18%) à celui du schéma classique avec testage sur descendance.

Article 4 : Using a very low density SNP panel for genomic selection in a breeding program for sheep

Jérôme Raoul, Andrew A. Swan and Jean-Michel Elsen. 2017. Genetics, Selection, Evolution. 49:76.

Raoul *et al. Genet Sel Evol (2017) 49:76*
DOI 10.1186/s12711-017-0351-0

GSE Genetics Selection Evolution

RESEARCH ARTICLE Open Access

Using a very low-density SNP panel for genomic selection in a breeding program for sheep

Jérôme Raoul[1,2*], Andrew A. Swan[3] and Jean-Michel Elsen[2]

Abstract

Background: Building an efficient reference population for genomic selection is an issue when the recorded population is small and phenotypes are poorly informed, which is often the case in sheep breeding programs. Using stochastic simulation, we evaluated a genomic design based on a reference population with medium-density genotypes [around 45 K single nucleotide polymorphisms (SNPs)] of dams that were imputed from very low-density genotypes ($\leq$ 1000 SNPs).

Methods: A population under selection for a maternal trait was simulated using real genotypes. Genetic gains realized from classical selection and genomic selection designs were compared. Genomic selection scenarios that differed in reference population structure (whether or not dams were included in the reference) and genotype quality (medium-density or imputed to medium-density from very low-density) were evaluated.

Results: The genomic design increased genetic gain by 26% when the reference population was based on sire medium-density genotypes and by 54% when the reference population included both sire and dam medium-density genotypes. When medium-density genotypes of male candidates and dams were replaced by imputed genotypes from very low-density SNP genotypes (1000 SNPs), the increase in gain was 22% for the sire reference population and 42% for the sire and dam reference population. The rate of increase in inbreeding was lower (from − 20 to − 34%) for the genomic design than for the classical design regardless of the genomic scenario.

Conclusions: We show that very low-density genotypes of male candidates and dams combined with an imputation process result in a substantial increase in genetic gain for small sheep breeding programs.

Background

Although technical and economic interests of genomic breeding programs in sheep have been positively assessed by various authors [1–5], only a few countries deliver genomic breeding values for meat and dairy sheep [6]. For many sheep breeds, the remaining main obstacle is to build an efficient reference population based on medium-density (MD ~ 50 K) genotypes and estimated breeding values (EBV) of sires to estimate the effects of single nucleotide polymorphisms (SNPs). Genomic prediction accuracy depends mainly on the number of animals included in the reference population, the accuracy of their EBV, their relationship with target animals (mostly candidates without phenotypes), and the effective size of the population [7]. In sheep, compared to dairy cattle, the effective population size is generally larger [8], and especially for maternal traits, EBV of both artificial insemination (AI) and naturally-mated sires are less accurate due to smaller numbers of progeny per sire. An increase in reference population size could counterbalance these unfavorable factors. However, this increase is limited when the population itself is small and the reference population is based on sires only. Including records and genotypes from lower tiers of the population is promising [9], whereas, so far, multi-breed approaches have not led to the expected increases in genomic prediction accuracy [10]. Another way to increase the reference

*Correspondence: Jerome.raoul@inra.fr
[2] GenPhySE, INRA, Castanet-Tolosan, France
Full list of author information is available at the end of the article

population would be to include females. Studies on dairy cattle breeding programs show that this strategy is efficient, especially when the reference population is based on a limited number of sires and/or records [11–15]. The impact of including females in the reference population on a sheep breeding program has never been assessed.

In sheep, the profitability at the nucleus level remains a critical factor for the design of breeding programs [16]. Implementation of genomic selection is mostly likely only if the cost is similar to that of the current design. In a genomic design, AI sires are no longer progeny-tested but both the animals in the reference population and the selection candidates must be genotyped. Since genotyping costs are quite high in sheep relative to the economic value of breeding animals, the number of genotypes has a large influence on the profitability of the design. Including dams in the reference population would increase the genotyping costs. Besides, the magnitude of the selection differential on genomic estimated breeding values (GEBV) of male candidates is a critical factor that determines the additional benefit of genomic selection breeding programs [17]. Increasing the selection differential requires additional candidate genotypes resulting in additional genotyping costs as well.

Very low-density (VLD) genotypes in association with imputation techniques would reduce the genotyping costs of a genomic design. VLD panels based on a few hundred SNPs are available at lower cost than low-density genotypes (~ 15 K) and are mainly used for parentage assignment [18–20]. Since candidate SNPs for parentage assignment are widely available across breeds (e.g. 9269 SNPs had a minor allele frequency higher than 0.30 in at least 20 French sheep populations [20]), new panels under development will probably be close to 1000 SNPs. Imputation techniques based on common SNPs that are present on both MD and VLD panels can be used to infer missing MD genotypes of male candidates and dams, thus to exploit MD genotypes of selected sires as a reference population. Population-based methods use the linkage disequilibrium (LD) between SNPs and haplotype frequencies only, whereas population- and family-based methods also include co-segregation information based on pedigree. The factors that affect imputation accuracy [21–35] and the relation between imputation accuracy and genomic prediction quality [21, 25–27, 30, 31, 34, 35] are well documented. In sheep, Moghaddar et al. [31] found a correlation close to 1 between GEBV computed from real versus imputed genotypes with an average imputation accuracy of 0.96. The imputation accuracy depends on: (1) the characteristics of the low-density panel with respect to the minor allele frequency, the number, spacing and localization of SNPs [21–27, 30, 33, 34]; (2) the linkage between adjacent SNPs [22, 24]; and (3) the characteristics of the reference population including size, single or multi-breed population and relationship with imputed animals [21, 24, 26, 27, 29–31, 34]. As observed with genomic prediction accuracy, imputation accuracy increases as close relatives are included in the reference population and pedigree information is used [21, 24, 26, 29, 30, 34].

Focusing on a breeding program applied to a small population of purebred sheep in which both AI and natural mating sires are used, the objectives of this study were to quantify the impact of (1) increasing the reference population size with female genotypes and (2) imputing genotypes of male candidates and females from very low- to medium-density SNP panels. The presence of all sires and grand-sires of dams and male candidates in the reference population for imputation is expected to limit the detrimental effect of using a VLD panel on imputation accuracy. Various designs were compared, including a classical selection scheme based on progeny testing of AI sires as a baseline scenario, and several genomic selection schemes. Five scenarios were assessed for genomic selection by varying the reference population component (with or without dams), and genotype information (MD or imputed MD genotypes for male candidates and dams).

Methods

To compare different designs of breeding programs, we developed a stochastic simulation model where individuals, including their genome, were simulated. Stochastic events were simulated using the NAG FORTRAN library [the Numerical Algorithms Group (NAG), Oxford, UK]. Parents were randomly mated and replacements randomly selected during 20 reproductive cycles to establish a founder population. Ten years of selection for a maternal trait were then simulated by applying a classical progeny test design. At this time (Time = 10), the average LD (r^2) between SNPs was equal to 0.13 at 50 kb, 0.07 at 200 kb and 0.05 at 1000 kb. These values of r^2 were within the range of estimated r^2 and Ne reported by Kijas et al. [36] for the Merino breed (r^2 ~ 0.10 at 50 kb, Ne = 900), the Suffolk breed (r^2 ~ 0.13, Ne = 569) and the Poll Dorset breed (r^2 ~ 0.19 at 50 kb, Ne = 318). The next 15 years were simulated by applying either a classical or a genomic design. The main steps of the simulation are described in Fig. 1.

Founder population

A total of 11,816 real genotypes of 54,241 bi-allelic SNPs (Illumina OvineSNP50 BeadChip-Illumina ©) from progeny-tested sires of the dairy Lacaune sheep breed were used to initialize population genomes (initial population). The genotypes were cleaned (call frequency ≥ 0.95,

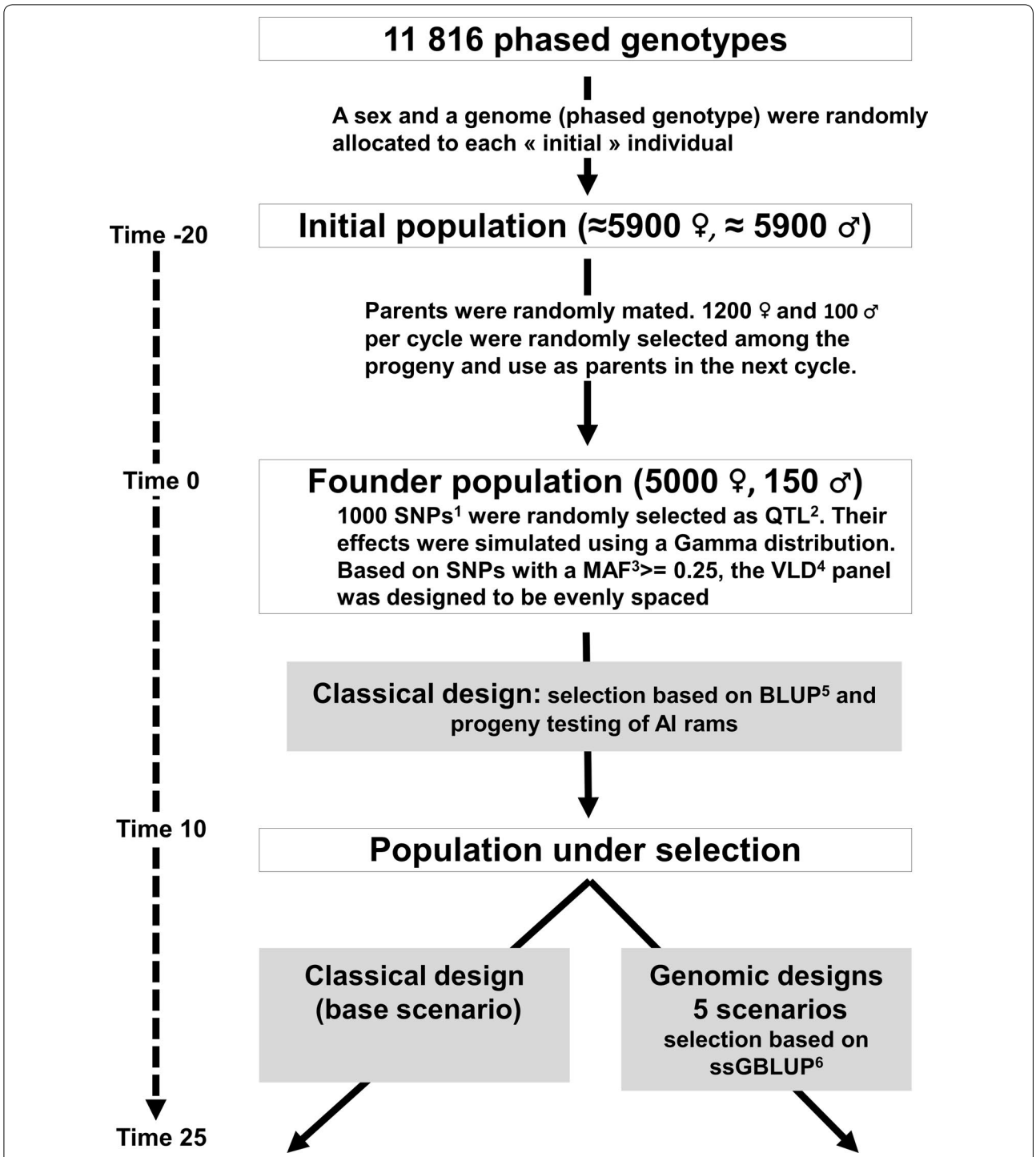


Fig. 1 General overview of the simulation steps. [1]SNP, single nucleotide polymorphism; [2]QTL: quantitative trait locus; [3]MAF: minor allele frequency; [4]VLD: very low density; [5]BLUP: best linear unbiased prediction; [6]ssGBLUP: single step genomic BLUP

Hardy–Weinberg equilibrium, Mendelian inheritance compatibilities) and only autosomal SNPs with a known position on the genome were retained: 47,706 SNPs were used for the genome simulations. Using genealogical information, phased chromosomes were obtained from Fimpute [37]. At Time -20, the phased genome of the 11,816 individuals and a randomly assigned sex were used as a starting point for each simulation run.

Accuracy of genomic predictions depends on the relationships between individuals of the reference population and candidate animals [31, 38]. To obtain the genomes of a small population of females mated to a limited number of males, random mating cycles were performed using numbers of males and females close to the replacement rates used in selection in the subsequent cycles. The number of random mating cycles has a limited effect on LD (differences in LD between various numbers of cycles were lower than 0.01 at 50 kb, 200 kb and 1000 kb within 10 to 30 cycles). To obtain the desired demographic structure, 20 random mating cycles were chosen. At each cycle, 1200 females were randomly selected from females that were born in the previous two cycles and randomly mated to 100 males that were selected among males born in the previous cycle. Given the number of progeny per dam allocated according to a prior distribution (single = 0.4, twin = 0.5, triplet = 0.1), selection from two cycles was required to generate a sufficient number of female replacements. The parental gametes were generated to produce new individual genomes: the number of crossovers was simulated following a Poisson distribution $P(\lambda)$ with λ the chromosome length expressed in Morgan. Positions of the crossovers were allocated following a uniform distribution along the chromosome and a parental strand was randomly chosen (i.e. paternal or maternal) to start the haplotype reconstruction of the gamete. Mutations were simulated with a mutation rate of 1×10^{-6} per SNP per meiosis. At Time 0, founders were randomly selected according to a demographic structure (i.e. percentage of individuals per age-category) among individuals that were created in the last seven cycles, aged from 1 to 7 years old at Time 0.

Quantitative trait loci and phenotype simulation

One thousand quantitative trait loci (QTL) underlying the maternal trait under selection were randomly selected among SNPs with a minor allele frequency higher or equal to 0.05. Following Hayes and Goddard [39], QTL effects were drawn from a Gamma distribution with a shape parameter of 0.4 and a scale parameter of 1.66. Assuming (1) no dominance effect, (2) Hardy–Weinberg equilibrium and (3) all the additive variance is explained by the QTL, the additive genetic variance was $VA = \sum_{i=1}^{nqtl} 2p_i(1-p_i)q_i^2$ [40], where p_i denotes the allele frequency of the first allele for QTL i and q_i its effect. The QTL effects were rescaled to make the additive variance VA equal to 1. The maternal permanent effects, pe, were drawn from a normal distribution with mean 0 and variance $\frac{rep-h^2}{h^2}$ where $rep = 0.5$ and $h^2 = 0.25$ denote the repeatability and the heritability of the trait. The residual effects were drawn from a normal distribution with mean 0 and variance $\frac{1-rep}{h^2}$. The kth phenotype y_{jk} of a female j was simulated as $y_{jk} = TBV_j + pe_j + e_{jk}$ with TBV_j the true breeding value depending on the effects of all QTL and the female genotype.

Estimation of breeding values

The estimation of breeding values was based on an animal best linear unbiased prediction (BLUP) model [41] for the classical design and genomic evaluation on an animal single-step genomic (G)BLUP [42] for the genomic design. Both types of evaluation were performed using the Blupf90 software [43] developed by Misztal et al. The genomic relationship matrix was built following Van Raden [44]. Marker inconsistencies between parents and progeny were due to imputation errors. Errors detected by Blupf90 resulted in the removal of the corresponding progeny genotype before the evaluation. Coefficients of inbreeding based on pedigree information were computed using the methodology developed by Aguilar and Misztal [45] and implemented in the Blupf90 software.

Very low-density and medium-density panels

SNPs with a minor allele frequency higher or equal to 0.25 and not selected as a QTL were considered as candidates for inclusion in the VLD panel. To select n_{vld} evenly spaced SNPs ($n_{vld} = 250$, 500 and 1000), the genome was divided into n_{vld} windows. Each window was subdivided into three parts and a SNP was randomly selected from candidate SNPs located in the central part. If no candidate was available, a SNP located within the window or in adjacent windows was selected. The MD panel (46,706 SNPs) included all SNPs except those selected as QTL.

Population and breeding designs

Around 5000 females divided into 10 flocks were recorded each year for the trait under selection. According to Hill [46], and neglecting the correlation between progeny sizes of sires to breed sires and sires to breed dams and between progeny sizes of dams to breed sires and dams to breed dams, the effective population size was around 180. At each reproductive cycle (length = 1 year), half of the breeding females were selected on EBV for mating by AI. Assuming a fertility rate of 55% on induced estrus, females that did not conceive to AI and females not selected for AI were then randomly mated to natural mating sires allocated to their flock assuming a fertility rate of 90%. The average number of females per naturally-mated sire was equal to 35. To prevent inbreeding, a male (AI or natural mating) could not be mated to a female belonging to its dam's flock. The number of progeny per dam depended on the reproduction mode (induced or natural estrus) and parity (first versus second and more). At each cycle, some dams were randomly culled with a proportion varying from 0.10 after parity 1 to 0.50 on

average after parity 6. The maximum parity was 7 and around 24% of dams were replaced per cycle. Replacement females, allocated to their dam's flock, were first randomly chosen among females that were born from an AI sire and then among females that were born from a natural mating sire. There is no difference between the classical and genomic designs, presented in Fig. 2, regarding how females were selected.

About 600 male candidates were preselected among newborn males (~ 3500) based on the parental mean EBV in the classical design and on the parental mean GEBV in the genomic design. In practice, all candidates were born from proven AI sires (classical design) and genomic AI sires (genomic design) given their genetic superiority. About 10% of male candidates died within their first year. At 1 year of age, 50% of the male candidates were culled for non-genetic factors. In the classical design, the ten candidates with the highest parental mean EBV (out of 270 candidates) were selected and mated across flocks by AI to 1100 females (110 per male on average) for progeny-testing. Two cycles later, the five AI sires with the highest EBV (including progeny records) were selected as proven AI sires and used at most for four cycles (from 4 to 7 years old). At each cycle, 1400 females were mated to proven AI sires. In the genomic design, MD or imputed MD genotypes of candidates were available. The ten

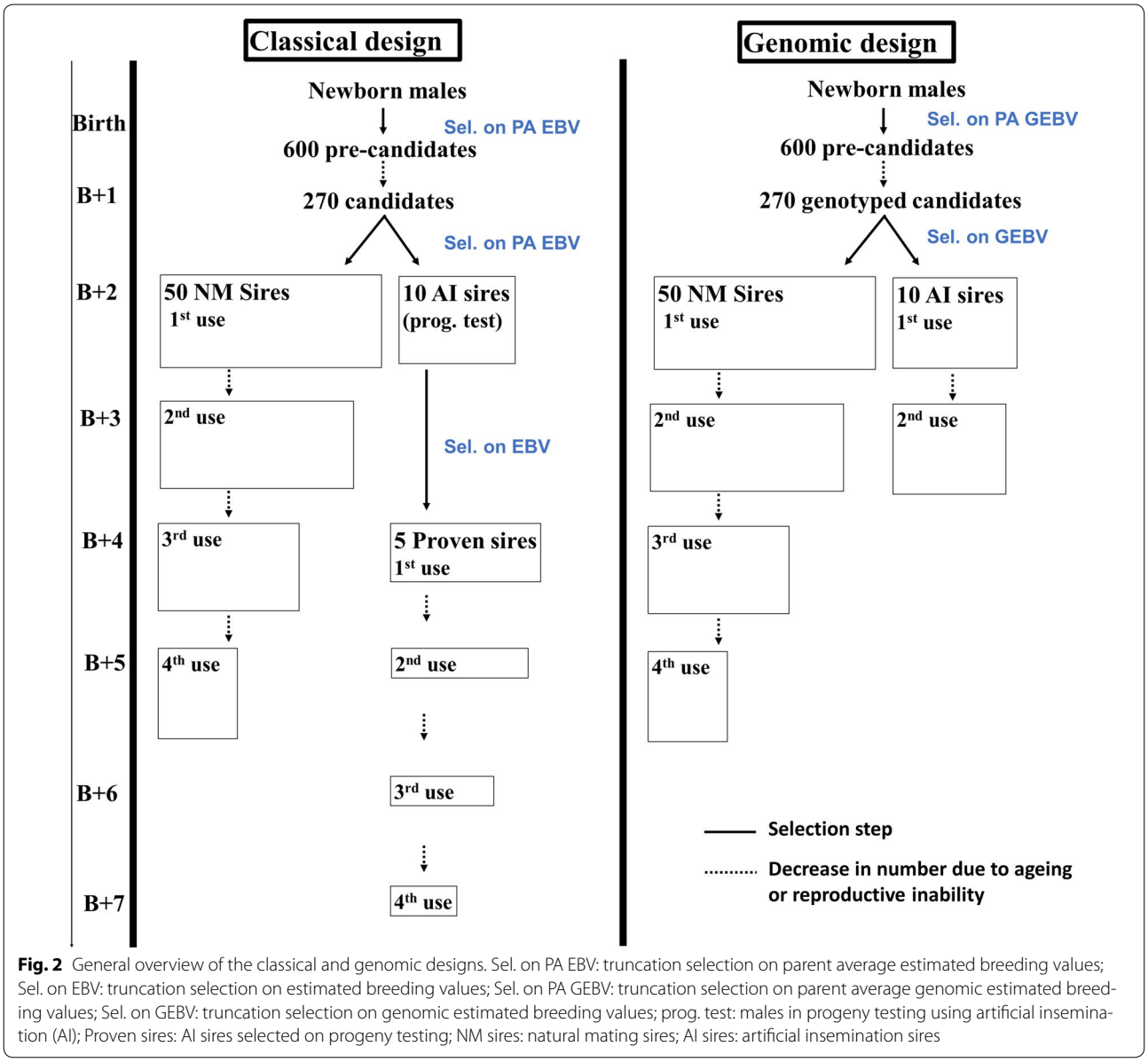


Fig. 2 General overview of the classical and genomic designs. Sel. on PA EBV: truncation selection on parent average estimated breeding values; Sel. on EBV: truncation selection on estimated breeding values; Sel. on PA GEBV: truncation selection on parent average genomic estimated breeding values; Sel. on GEBV: truncation selection on genomic estimated breeding values; prog. test: males in progeny testing using artificial insemination (AI); Proven sires: AI sires selected on progeny testing; NM sires: natural mating sires; AI sires: artificial insemination sires

candidates with the highest GEBV were selected as AI males and mated across flocks to 1100 females. Genomic AI sires were used at most for another cycle and mated to 1400 females. To mimic the variation in the AI dose production per male, the inseminated dams were randomly allocated to one AI sire with a maximum number of AI per male equal to 140 for 2-year old sires and to 250 for older sires. About 50 naturally-mated male replacements were selected among 260 candidates not selected as AI sires, based on their parental mean EBV in the classical design or on their GEBV in the genomic design. These naturally-mated sires were used at most for four cycles (from 2 to 5 years old).

Five genomic scenarios (Table 1) were assessed. Some scenarios included an imputation step to reconstruct the MD genotypes of individuals genotyped with the VLD panel. The imputation was performed using the software Fimpute developed by Sargolzaei et al. [37]. The reference population for imputation was based first on the historical population of AI and natural mating sires at Time 10: about 50 progeny tested sires (each with ~ 30 daughters), 50 proven sires (~ 125 daughters) and 400 sires (~ 9 daughters). From Time 10 to 25, new selected sires were added, i.e. about 10 AI and 50 natural mating sires per year. The scenarios differed in the choice of genotyped animals (including dams or not), the SNP panel used for the genotyping of male candidates and dams (MD or VLD followed by imputation to MD) and the information used for imputation (including pedigree or not). In the GSsc scenario, sires and male candidates are genotyped using the MD panel. In the GSs_Ic scenario, sires are genotyped using the MD panel and male candidates using the VLD panel. In the GSscd scenario, sires, male candidates and dams are genotyped using the MD panel. In the GSs_Icd scenario, sires are genotyped using the MD panel and male candidates and dams using the VLD panel. To assess the importance of pedigree information in the genomic design, the last scenario, GSs_Icd_pop, was similar to GSs_Icd but an imputation method that ignored pedigree information (population based) was used.

Simulation outputs

At each cycle and for each scenario, the average TBV and inbreeding coefficient of females in their first parity were computed. The annual genetic gain and annual rate of inbreeding were estimated as the regression slopes of the average TBV and average inbreeding coefficient of first parity females on time between cycles 10 and 25. The means and standard deviations presented are based on 50 replicates. At time 25, we computed the average Pearson correlation between the TBV and the (G)EBV for dams and male candidates as well as the average concordance rate of imputed genotypes (the number of correctly imputed alleles divided by the number of imputed alleles) of dam and male imputed genotypes.

Results

Table 2 includes the annual genetic gain, the annual increase in inbreeding, (G)EBV accuracies, and concordance rates of imputed dams and male candidates for six designs. Computed at time 25, we assumed that GEBV accuracies and concordance rates were close to the upper bounds of values obtained over the simulation time. We used a VLD panel of 1000 SNPs for designs that included an imputation step to the MD panel. Table 3 provides the same information for designs based on female imputed genotypes using panels of 1000, 500 or 250 SNPs.

Table 1 Information taken into account for (genomic) breeding value estimation and imputation step according to scenarios

Scenarios[a]	CS[b]	GSsc[c]	GSs_Ic[c]	GSscd[c]	GSs_Icd[c]	GSs_Icd_pop[c]
Genotypes (G[d]) or imputed genotype (IG[e])						
Sires		G	G	G	G	G
Male candidates		G	IG	G	IG	IG
Dams				G	IG	IG
Imputation methodology: population (P)/familial (F)			P + F		P + F	P

[a] Phenotypes and pedigree were included in all scenarios

[b] CS = classical selection design

[c] GS = genomic selection design; GSsc, sires and candidates had medium-density genotypes; GSs_Ic, sires had medium-density genotypes and candidates had medium-density genotypes imputed from very low-density genotypes; GSscd, sires, candidates and dams had medium-density genotypes; GSs_Icd, sires had medium-density genotypes and candidates and dams had medium-density genotypes imputed from very low-density genotypes; GSs_Icd_pop, sires had medium-density genotypes and candidates and dams had medium-density genotypes imputed from very low-density genotypes without using the pedigree information

[d] Medium-density genotypes (46 K)

[e] Medium-density genotypes (46 K) imputed from very low-density genotypes ($\leq$ 1000)

Table 2 Genetic gain, inbreeding rate, (G)EBV accuracies, and imputation concordance rates for six scenarios[a] with VLD panels of 1000 SNPs (standard deviations for 50 replicates shown in brackets)

Scenarios[a]	CS[b]	GSsc[c]	GSs_Ic[c]	GSscd[c]	GSs_Icd[c]	GSs_Icd_pop[c]
Genetic gain[d] (σ_a/year)	0.162 (0.015)	0.205 (0.019)	0.197 (0.021)	0.249 (0.016)	0.230 (0.014)	0.179 (0.020)
Inbreeding rate/year[e]	0.0043 (0.0011)	0.0034 (0.0006)	0.0033 (0.0006)	0.0028 (0.0005)	0.0031 (0.0006)	0.0040 (0.0008)
(G)EBV accuracy[f]						
Dams[g]	0.71 (0.02)	0.76 (0.02)	0.75 (0.02)	0.87 (0.01)	0.83 (0.02)	0.74 (0.03)
Male candidates	0.36 (0.07)	0.53 (0.06)	0.51 (0.06)	0.71 (0.05)	0.63 (0.05)	0.43 (0.07)
Imputation concordance rate[h]						
Dams[g]					96.1 (0.1)	91.1 (0.2)
Old females[i]					93.3 (0.1)	88.3 (0.2)
Male candidates			94.7 (0.2)		96.5 (0.1)	92.3 (0.2)

[a] Scenarios based on imputation were performed with a very low-density 1000-SNP panel

[b] CS = classical selection design

[c] GS = genomic selection design; GSsc, sires and candidates had medium-density genotypes; GSs_Ic, sires had medium-density genotypes and candidates had medium-density genotypes imputed from very low-density genotypes; GSscd, sires, candidates and dams had medium-density genotypes; GSs_Icd, sires had medium-density genotypes and candidates and dams had medium-density genotypes imputed from very low-density genotypes; GSs_Icd_pop, sires had medium-density genotypes and candidates and dams had medium-density genotypes imputed from very low-density genotypes without using the pedigree information

[d] Computed as the slope of the average true breeding value of females in first parity between time10 and time25

[e] Computed as the slope of the average inbreeding coefficient of females in first parity between time10 and time25

[f] Computed as the average Pearson correlation between the true breeding value and (genomic) estimated breeding values of animals at time25

[g] Dams mated at time 25

[h] Computed as the average of number of correctly imputed SNP divided by the number of imputed SNP obtained for the imputation realized at time 25

[i] Dams not present anymore at time 25

Annual genetic gain

Annual genetic gains indicated in Tables 2 and 3 correspond to the slope of the average TBV of females in parity 1 from year 10 to 25 for each design. Results show that the genetic gain for a genomic design based on a sire reference population (GSsc: $0.205\sigma_a$) was 27% higher than for the classical design (CS: $0.162\sigma_a$). When male candidates were imputed from the VLD, the genetic gain slightly decreased but remained significantly different from that of the classical design (GSs_Ic: $0.197\sigma_a$). Including dams in the reference population doubled the additional genetic gain of the genomic design (+ 54%) compared to the classical design (GSscd: $0.249\sigma_a$). When candidates and dams were imputed, based on both family and population information, the increase in gain was lower (GSs_Icd: $0.230\sigma_a$) but remained higher than that of a genomic design based on a sire reference population (GSs_Ic: $0.197\sigma_a$). When imputation was based on population information, the genetic gain (GSs_Icd_pop: $0.179\sigma_a$) was close to the gain achieved in the classical design.

Inbreeding

The rates of inbreeding shown in Tables 2 and 3 correspond to the slope of the average inbreeding coefficient of females in parity 1 from year 10 to 25 for each design. Apart from when the imputation method did not take pedigree information into account, the increase in inbreeding was significantly slower for the genomic scenarios (from 0.0028 to 0.0034 per year) than for the classical design (0.0043 per year). Within genomic scenarios, the increase in inbreeding per year was lower as the number and quality of genotypes increased.

GEBV accuracies and imputation concordance rate

The EBV accuracies shown in Tables 2 and 3 were calculated as the average Pearson correlation between TBV and (G)EBV of females in parity 1–7 (dams) and young male candidates at time 25. Accuracies for dams were slightly higher in the genomic design based on a sire reference population (GSsc: 0.76) than in the classical design (CS: 0.71). Including dams in the reference population resulted in a substantial increase in accuracy either when they were genotyped with MD (GSscd: 0.87) or genotyped with VLD and imputed to MD using population and family information (GSs_Icd: 0.87). When pedigree information was not used for imputation, the GEBV accuracy for dams was lower (GSs_Icd_pop: 0.74) than in the genomic design without dams in the reference population. Accuracy for male candidates was lower in the classical design (CS: 0.36), since they were selected on mid-parent EBV, compared to the genomic designs where their own genomic information was included. For these males, accuracies increased from 0.43 to 0.71 as

Table 3 Genetic gain, inbreeding increase, (G)EBV accuracies, and imputation concordance rates for the GSs_Icd genomic design using VLD densities of 250, 500, and 1000 SNPs (standard deviations for 50 replicates shown in brackets)

Scenarios	GSs Icd[a]	GSs Icd[a]	GSs_Icd[a]
Number of SNPs	1000	500	250
Genetic gain[b] (σ_a/year)	0.230 (0.014)	0.183 (0.016)	0.175 (0.015)
Inbreeding[c]	0.0031 (0.0006)	0.0037 (0.0007)	0.0040 (0.0010)
GEBV accuracy[d]			
Dams[e]	0.83 (0.02)	0.74 (0.02)	0.73 (0.03)
Male candidates	0.63 (0.05)	0.45 (0.07)	0.38 (0.02)
Imputation concordance rate[f]			
Dams[e]	96.1 (0.1)	91.8 (0.3)	87.3 (0.4)
Old females[g]	93.3 (0.1)	88.7 (0.2)	84.5 (0.3)
Male candidates	96.5 (0.1)	92.4 (0.3)	88.0 (0.4)

[a] GSs_Icd = genomic selection design, sires had medium-density genotypes and candidates and dams had medium-density genotypes imputed from very low-density genotypes

[b] Computed as the slope of the average true breeding value of females in first parity between time10 and time25

[c] Computed as the slope of the average inbreeding coefficient of females in first parity between time10 and time25

[d] Computed as the average Pearson correlation between the true breeding value and (genomic) estimated breeding values of animals at time 25

[e] Dams mated at time 25

[f] Computed as the average of number of correctly imputed SNP divided by the number of imputed SNP obtained for the imputation realized at time 25

[g] Dams not present anymore at time 25

the quality of their own genomic information (imputed genotypes based either on population or population and family information, MD genotypes) increased and by including dams in the reference population.

The concordance rates of imputation reported in Tables 2 and 3 correspond to the number of correctly imputed alleles divided by the number of imputed alleles at Time 25. Results show that dams and young male candidates are imputed with higher accuracy than old females (+ 3%). Adding pedigree information (GSs_Icd) resulted in a higher imputation accuracy of + 4.2% for male candidates and + 5.0% for females compared to ignoring the pedigree in the imputation process (GSs_Icd_pop). Adding female VLD genotypes resulted in a gain of + 1.8% for male candidates.

Effect of the number of SNPs in the VLD panel on the GSs_Icd design

Table 3 shows that the genetic gain was much affected by the number of VLD SNPs used for imputation. The additional gain obtained for 1000 SNPs ($0.230\sigma_a$) fell to $0.183\sigma_a$ for 500 SNPs, while the gain for 250 SNPs was close to, but still significantly higher ($0.175\sigma_a$) than, the genetic gain obtained for the classical design ($0.162\sigma_a$). Compared to the use of 1000 SNPs, the decrease in accuracy was approximately 0.1 for dams irrespective of the number of SNPs, 500 or 250. These accuracies were close to the accuracy obtained when dams were not included in the reference population (GSsc, GSs_Ic). The same comparison for male candidates shows that the decrease was larger when 250 SNPs (− 0.25) were used than when 500 SNPs (− 0.18) were used. Across all numbers of SNPs, the magnitude of the decrease in concordance rate was similar regardless of the category considered: from − 4.1 to − 4.6 for 500 SNPs and from − 8.5 to − 8.8 for 250 SNPs compared to 1000 SNPs.

Discussion

In this study, our reference scenario the classical design produced a rate of genetic gain of $0.162\sigma_a$. This result is close to $0.173\sigma_a$, which is the annual genetic gain from 1986 to 1999 estimated from real data for the Red Faced Manech breed [47], a breeding program similar to the simulation design (single trait selection until 2003). Compared to the classical design, genomic designs generated more annual genetic gain and limited the increase in inbreeding. Sensitivity to a lower heritability ($h^2 = 0.10$ instead of 0.25) and a higher efficiency of the classical design (proportion selected of 0.33 instead of 0.5 to select proven AI sires) were also assessed. Based on ten replicates (results not shown), differences in genetic gain were slightly smaller for a low heritability or a higher efficiency of the classical design, but were of a similar magnitude to the results obtained with the classical design, GSscd and GSs_Icd in this study.

The rate of increase in inbreeding was lower in genomic scenarios for three main reasons: first, the average contribution of natural mating sires was identical across designs whereas the average contribution of AI sires was smaller in the genomic design (genomic AI sires were systematically culled after they had been used for two cycles); second, the selection intensity applied on parent average genetic value to select male candidates in the genomic design was lower than that to select AI and natural mating sires in the classical design; and third, genomic information was expected to reduce the relative importance of pedigree information in breeding value estimation through improved estimation of the Mendelian sampling term [17]. The latter might explain why within genomic scenarios, the increase in the rate of inbreeding was lower as the quality and quantity of the genomic information improved.

Performance of a genomic selection design based on a reference population of related sires

Compared to the classical design, a genomic design based on a reference population of related sires resulted

in an increase in genetic gain of 26.3% per year when candidates were genotyped with the MD panel (GSsc). This increase is inferior to the increase in gain of 50% calculated by Shumbusho et al. [1] for the Red Faced Manech breed using a deterministic model. Although the designs were similar, Shumbusho et al. [1] computed the accuracy of GEBV based on a reference population of 2000 animals with a constant number of records using the formula of Daetwyler et al. [48], whereas in our study, the cumulated number of genotyped sires (included in the single-step GBLUP) increased from around 500 (Time 10) to 1250 (Time 25), and the number of daughters per sire was highly variable due to both use of AI and natural mating. For the GSsc scenario, the increase in accuracy of GEBV was moderate for dams (7%) but considerable for young males (47%). When candidate male genotypes were imputed from genotypes based on 1000 SNPs (GSs_Ic), the increase in gain (+ 21.6%) compared to that obtained with the classical design was lower than with GSsc. The average concordance rate, equal to 94.7%, is of the same order as results obtained by Wang et al. [25] and corroborates that an acceptable imputation quality based on a VLD panel could be achieved provided that sires are included in the training population. The GEBV accuracies for male candidates were slightly lower for GSs_Ic than for GSsc ($P = 0.068$) in agreement with results obtained in sheep by Moghaddar et al. [31] for a similar range of imputation accuracies.

Sires of male candidates were selected among young males in the genomic design and progeny-tested males in the classical design. Although the accuracy of GEBV of young males was lower than that of progeny-tested males (0.71—results not shown), higher selection intensities and a shorter average generation interval resulted in a higher genetic gain. Regardless of the design, male candidates were preselected based on parent average genetic values. The genomic AI sires were selected on their own genomic value before 2 years of age with a proportion selected of 1/27 versus 1/2 at 4 years old after progeny-testing for AI sires in the classical design. Since genomic AI sires were used for a maximum of 2 years, the average generation interval was reduced (3 years instead of 4). The natural sires were used and selected at the same age with a proportion selected close to 1/5 but on their own genomic value in the genomic design or on parent average genetic value in the classical design.

Value of female VLD genotypes

The comparison between scenarios with (GSscd) or without females (GSsc) included in the reference population shows that adding dam genotypes, along with their phenotypes, resulted in a doubling of the increase in genetic gain of the genomic design. Genotyping dams led to an increase in the average accuracy of GEBV for both dams (14%) and male candidates (34%). The effect of including cow genotypes in the reference population on accuracy and bias of genomic prediction has been widely reported in dairy cattle [11–15, 49–55]. Genetic gain varied depending on the population and reference population structure. For example, Koivula [14] reported a small additional gain (increase in accuracy of + 2 to 4%) when the reference population contained about 4400 sires, whereas McNugh [11] reported that including a large female population in the reference population increased the annual genetic gain by a factor of 3. In our study, the increase in gain was also large. Buch et al. [12] and Gonzalez-Recio et al. [13] showed that the increase in genomic prediction accuracy due to female information was most important when phenotypes and sizes of sire progeny groups were limited. Thus, this result was expected given the structure of our reference population: the number of sires was limited and most sire progeny groups were generated by natural mating, resulting in small numbers of daughters and thus inaccurate phenotypes.

When male candidates and dams were genotyped with a 1000-SNP panel and imputed genotypes were included for genomic evaluation (GSs_Icd), the additional genetic gain was reduced but it remained higher (+ 41%) than that obtained with the classical design. With a concordance rate of imputed animals of about 96%, the accuracy of GEBV of male candidates and dams decreased compared to that in GSscd (respectively − 11.3 and − 4.6%) but remained higher than accuracies obtained with a reference population based on sires (respectively + 15.9% and + 8.4% pts). Compared to the classical design, the increase in accuracy with a 1000-SNP panel was + 41% but this decreased to + 8% with a 250-SNP panel. Lower imputation accuracies obtained with lower density SNP panels and larger numbers of genotypes discarded from the genomic evaluation due to parent-progeny mismatches (results not shown), resulted in no increase in GEBV accuracy of dams when they were genotyped with the 250- or 500-SNP panel: both panels gave the same accuracies as the GSsc scenario in which dams were not included in the reference population. For male candidates, the accuracy of GEBV was lower for 500 (− 40.0%) and 250 (− 65.8%) SNPs than the accuracy obtained with 1000 SNPs but the increases in gain for genomic scenarios based on 250 and 500 SNPs, compared to the classical design, were highly significant ($P < 0.001$). Comparing GSs_Icd based on 250 SNPs with the classical design, the differences in accuracy of GEBV for dams and male candidates were small (+ 0.02) but highly significant

($P < 0.001$) for dams and moderate for male candidates ($P = 0.0565$).

Importance of using pedigree information in the imputation of MD genotypes from VLD genotypes

In GSs_Icd_pop, we removed pedigree information in the imputation of male candidate and dam genotypes. The comparison of GSs_Icd_pop with GSs_Icd shows that removing pedigree information in the imputation step decreased the concordance rate by 5 points and 4.2 points, respectively for dams and male candidates. A high proportion of genotypes (around 80%) were removed in GSs_Icd_pop, due to mismatches between parents and progeny (results not shown). In GSs_Icd_pop, the annual genetic gain remained higher (+ 9.5%) than in the classical design but was substantially lower than in GSs_Icd (+ 45%). The contribution to the analysis of co-segregation between VLD and MD SNPs has already been reported in previous studies which highlight the beneficial effect of including relatives in the reference population for imputation [21, 24, 26, 27, 29–31, 34], especially sires and grandsires [24]. The effect of including close relatives in the reference population on genomic prediction accuracy was also pointed out by previous studies on simulated and real data [38, 56]. Habier et al. [57] showed that co-segregation, as well as LD and additive genetic relationships, all contribute to the capture of QTL effects, whereas Sun et al. [58] showed that explicitly modelling the co-segregation results in higher accuracy of genomic prediction when the recent effective population is small. Such results suggest that a small reference population that contains all dams and sires of candidates can result in substantial additional genetic gain based on a better genomic prediction accuracy.

Feasibility of a genomic design including imputed genotypes on dams in the reference population

Higher gain was obtained with a genomic design when including dam medium-density genotypes in the reference population. The economic value of such a scenario should be assessed over the whole population, including both the nucleus and commercial levels. However, as discussed for Australian sheep and beef cattle industries by van der Werf and Banks [16], for an individual breeder or nucleus, the main objective is to achieve the maximum economic return within the company. Regarding the value of breeding stock, a lower genetic gain but cheaper design can be relevant for sheep breeding programs. For a French small ruminant breeding program, the cost of AI males including their progeny testing is assumed to be at least 400 euros per male per year [59]. In a genomic selection design, due to AI sires being used earlier and over a shorter period, the corresponding cost would be reduced, but genotyping costs have to be considered for building the reference population and identifying candidates. Implementing a GSscd design would require a major investment compared to the current classical design since all dams have to be genotyped. The investment for implementing a GSs_Icd design would be the same as that for the classical design, for example if VLD genotypes of dams were supported by breeders for parentage assignment purposes. However, the current price of the very low-density panel available for sheep (around 20 euros) limits its use for parentage assignment by breeders. A combined use of a VLD panel for both parentage and genomic selection might be cost-effective.

Plieschke et al. [15] noted that including female genotypes led to increased accuracy of GEBV without bias as long as the females chosen for genotyping were randomly sampled. In our study, all dams were genotyped. The main purpose was to assess the usefulness for genomic selection of VLD genotypes used for parentage assignment. From a practical point of view, one can imagine that only a certain proportion of flocks will use parentage assignment based on SNPs. In that case, there could be variation across flocks, and potentially within a flock, of the average dam GEBV accuracy according to the VLD genotyping status of dams. This could result in a bias since male candidates are selected on parent mean GEBV.

Conclusions

Using a stochastic model, we compared classical and genomic selection designs. Five genomic scenarios were assessed by varying the structure of the reference population (with or without genotyping females) and the genotyping panels that were used (medium density, or imputed to medium from very low density). Compared to the classical design with progeny testing, genomic scenarios generated more genetic gain and limited the increase in inbreeding. This superiority was based on higher selection differentials that are applied to male candidates to select sires for both AI and natural mating. The combined use of very low-density genotypes for male candidates and dams together with imputation resulted in lower genetic gain than scenarios based on medium-density genotypes. However, the increase in gain was substantial compared to that in the classical design. Using very low-density SNP panels might be more profitable at the nucleus level given its lower cost relative to medium- or low-density SNP panels and multipurpose use (parentage assignment and genomic selection).

Authors' contributions

JR and JME designed the study. JME established the methodology to create the founder population and genome simulation. JR wrote the simulation program, performed analysis and drafted the manuscript. JR, JME and AS contributed in interpreting the results and revising the manuscript. All authors read and approved the final manuscript.

Author details
[1] Institut de l'Elevage, Castanet-Tolosan, France. [2] GenPhySE, INRA, Castanet-Tolosan, France. [3] Animal Genetics and Breeding Unit, University of New England, Armidale, Australia.

Acknowledgements
We gratefully acknowledge Misztal et al. and Sargolzaei et al. for respectively making the Blupf90 software and the Fimpute software available.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
The datasets generated during and/or analysed during this study are available from the corresponding author on reasonable request.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Funding
This study was funded by Institut de l'Elevage, INRA (Department of Animal Genetic and metaprogram SELGEN), Apis-gène and AGBU (Animal Genetic and Breeding Unit). AGBU is a joint unit between the University of New England and the Australian Department of Primary Industry. We gratefully acknowledge the funding.

Publisher's Note

Received: 19 May 2017 Accepted: 17 October 2017
Published online: 24 October 2017

References

1. Shumbusho F, Raoul J, Astruc JM, Palhière I, Elsen JM. Potential benefits of genomic selection on genetic gain of small ruminant breeding programs. J Anim Sci. 2013;91:3644–57.
2. van der Werf JHJ, Banks RG, Clark SA, Lee SJ, Daetwyler HD, Hayes BJ, et al. Genomic selection in sheep breeding programs. In: Proceedings of the 10th world congress on genetics applied to livestock production, 17–22 August 2014; Vancouver. 2014. http://agbu.une.edu.au/publications/10wcgalp/351_paper_10381_manuscript_1648_0.pdf. Accessed 06 April 2017.
3. Casellas J, Piedrafita J. Accuracy and expected genetic gain under genetic or genomic evaluation in sheep flocks with different amounts of pedigree, genomic and phenotypic data. Livest Sci. 2015;182:58–63.
4. Granleese T, Clark SA, Swan AA, van der Werf JHJ. Increased genetic gains in sheep, beef and dairy breeding programs from using female reproductive technologies combined with optimal contribution selection and genomic breeding values. Genet Sel Evol. 2015;47:70.
5. Shumbusho F, Raoul J, Astruc JM, Palhiere I, Lemarié S, Fugeray-Scarbel A, et al. Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. Animal. 2016;10:1033–41.
6. Rupp R, Mucha S, Larroque H, McEwan J, Conington J. Genomic application in sheep and goat breeding. Anim Frontiers. 2016;6:39–44.
7. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci. 2009;92:433–43.
8. Hall SJG. Effective population sizes in cattle, sheep, horses, pigs and goats estimated from census and herdbook data. Animal. 2016;10:1778–85.
9. Santos BFS, van der Werf JHJ, Gibson JP, Byrne TJ, Amer PR. Genetic and economic benefits of selection based on performance recording and genotyping in lower tiers of multi-tiered sheep breeding schemes. Genet Sel Evol. 2017;49:10.
10. Moghaddar N, Swan AA, van der Werf JH. Comparing genomic prediction accuracy from purebred, crossbred and combined purebred and crossbred reference populations in sheep. Genet Sel Evol. 2014;46:58.
11. McHugh N, Meuwissen THE, Cromie AR, Sonesson AK. Use of female information in dairy cattle genomic breeding programs. J Dairy Sci. 2011;94:4109–18.
12. Buch LH, Kargo M, Berg P, Lassen J, Sørensen AC. The value of cows in reference populations for genomic selection of new functional traits. Animal. 2012;6:880–6.
13. Gonzalez-Recio O, Coffey MP, Pryce JE. On the value of the phenotypes in the genomic era. J Dairy Sci. 2014;97:7905–15.
14. Koivula M, Strandén I, Aamand GP, Mäntysaari EA. Effect of cow reference group on validation reliability of genomic evaluation. Animal. 2016;10:1061–6.
15. Plieschke L, Edel C, Pimentel ECG, Emmerling R, Bennewitz J, Götz KU. Systematic genotyping of groups of cows to improve genomic estimated breeding values of selection candidates. Genet Sel Evol. 2016;48:73.
16. van der Werf J, Banks R. Effective breeding programmes down under. J Anim Breed Genet. 2017;134:1–2.
17. Pryce JE, Daetwyler HD. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim Prod Sci. 2012;52:107.
18. Clarke SM, Henry HM, Dodds KG, Jowett TWD, Manley TR, Anderson RM, et al. A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. PLoS One. 2014;9:e93392.
19. Heaton MP, Leymaster KA, Kalbfleisch TS, Kijas JW, Clarke SM, McEwan J, et al. SNPs for parentage testing and traceability in globally diverse breeds of sheep. PLoS One. 2014;9:e94851.
20. Tortereau F, Moreno CR, Tosser-Klopp G, Servin B, Raoul J. Development of a SNP panel dedicated to parentage assignment in French sheep populations. BMC Genet. 2017;18:50.
21. Habier D, Fernando RL, Dekkers JCM. Genomic selection using low-density marker panels. Genetics. 2009;182:343–53.
22. Zhang Z, Druet T. Marker imputation with low-density marker panels in Dutch Holstein cattle. J Dairy Sci. 2010;93:5487–94.
23. Hayes BJ, Bowman PJ, Daetwyler HD, Kijas JW, van der Werf JHJ. Accuracy of genotype imputation in sheep breeds. Anim Genet. 2012;43:72–80.
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. Wang C, Habier D, Peiris BL, Wolc A, Kranis A, Watson KA, et al. Accuracy of genomic prediction using an evenly spaced, low-density single nucleotide polymorphism panel in broiler chickens. Poult Sci. 2013;92:1712–23.
26. Cleveland MA, Hickey JM. Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. J Anim Sci. 2013;91:3583–92.
27. Wellmann R, Preuß S, Tholen E, Heinkel J, Wimmers K, Bennewitz J. Genomic selection using low density marker panels with application to a sire line in pigs. Genet Sel Evol. 2013;45:28.
28. Hozé C, Fouilloux MN, Venot E, Guillaume F, Dassonneville R, Fritz S, et al. High-density marker imputation accuracy in sixteen French cattle breeds. Genet Sel Evol. 2013;45:33.
29. He S, Wang S, Fu W, Ding X, Zhang Q. Imputation of missing genotypes from low- to high-density SNP panel in different population designs. Anim Genet. 2015;46:1–7.
30. Bolormaa S, Gore K, van der Werf JHJ, Hayes BJ, Daetwyler HD. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. Anim Genet. 2015;46:544–56.
31. Moghaddar N, Gore KP, Daetwyler HD, Hayes BJ, van der Werf JHJ. Accuracy of genotype imputation based on random and selected reference sets in purebred and crossbred sheep populations and its effect on accuracy of genomic prediction. Genet Sel Evol. 2015;47:97.
32. Xiang T, Ma P, Ostersen T, Legarra A, Christensen OF. Imputation of genotypes in Danish purebred and two-way crossbred pigs using low-density panels. Genet Sel Evol. 2015;47:54.
33. Ventura RV, Lee MA, Miller SP, Clarke SM, McEwan JC. Assessing imputation accuracy using a 15 K low density panel in a multi-breed New Zealand sheep population. Proc Assoc Advmt Anim Breed Genet. 2015;21:302–5.

34. Toghiani S, Aggrey SE, Rekaya R. Multi-generational imputation of single nucleotide polymorphism marker genotypes and accuracy of genomic selection. Animal. 2016;10:1077–85.
35. Badke YM, Bates RO, Ernst CW, Fix J, Steibel JP. Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. G3 (Bethesda). 2014;4:623–31.
36. Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, et al. Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Anim Genet. 2014;45:754–7.
37. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014;15:478.
38. Clark SA, Hickey JM, Daetwyler HD, van der Werf JH. 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. Genet Sel Evol. 2012;44:4.
39. Hayes BJ, Goddard ME. The distribution of the effects of genes affecting quantitative traits in livestock. Genet Select Evol. 2001;33:209–29.
40. Falconer DS. Introduction to quantitative genetics. London: Oliver and Boyd; 1960.
41. Henderson CR. Best linear unbiased estimation and prediction under a selection model. Biometrics. 1975;31:423–47.
42. 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 score1. J Dairy Sci. 2010;93:743–52.
43. Misztal I. Complex models, more data: simpler programming. Interbull Bull. 1999;20:33–42.
44. Van Raden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91:4414–23.
45. Aguilar I, Misztal I. Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. J Dairy Sci. 2008;91:1669–72.
46. Hill WG. A note on effective population size with overlapping generations. Genetics. 1979;92:317–22.
47. Astruc JM, Barillet F, Barbat A, Clement V, Boichard D. Genetic evaluation of dairy sheep in France. In Proceedings of the 7th world congress on genetics applied to livestock production, 19–23 August 2002; Montpellier. 2002.
48. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One. 2008;3:e3395.
49. Thomasen JR, Sørensen AC, Lund MS, Guldbrandtsen B. Adding cows to the reference population makes a small dairy population competitive. J Dairy Sci. 2014;97:5822–32.
50. Su G, Ma P, Nielsen US, Aamand GP, Wiggans G, Guldbrandtsen B, et al. Sharing reference data and including cows in the reference population improve genomic predictions in Danish Jersey. Animal. 2016;10:1067–75.
51. Gao H, Madsen P, Nielsen US, Aamand GP, Su G, Byskov K, et al. Including different groups of genotyped females for genomic prediction in a Nordic Jersey population. J Dairy Sci. 2015;98:9051–9.
52. Pryce JE, Hayes BJ, Goddard ME. Genotyping dairy females can improve the reliability of genomic selection for young bulls and heifers and provide farmers with new management tools. In: Proceedings of the international committee for animal recording 38th annual meeting, 28 May–1 June 2012; Cork; 2012.
53. Calus MPL, de Haas Y, Veerkamp RF. Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. J Dairy Sci. 2013;96:6703–15.
54. Dassonneville R, Baur A, Fritz S, Boichard D, Ducrocq V. Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. Genet Sel Evol. 2012;44:40.
55. Edel C, Pimentel ECG, Plieschke L, Emmerling R, Götz KU. Short communication: the effect of genotyping cows to improve the reliability of genomic predictions for selection candidates. J Dairy Sci. 2016;99:1999–2004.
56. Lee SH, Weerasinghe WMSP, Wray NR, Goddard ME, van der Werf JHJ. Using information of relatives in genomic prediction to apply effective stratified medicine. Sci Rep. 2017;7:42091.
57. Habier D, Fernando RL, Garrick DJ. Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics. 2013;194:597–607.
58. Sun X, Fernando R, Dekkers J. Contributions of linkage disequilibrium and co-segregation information to the accuracy of genomic prediction. Genet Sel Evol. 2016;48:77.
59. Buisson D. Sélection génomique des races ovines laitières françaises: analyse des schémas actuels, première modélisation de scénarios génomiques et bilan technico- économique. 2012. https://dumas.ccsd.cnrs.fr/dumas-00801194/document. Accessed 25 Apr 2017.

Article 5 (court): Assessment of a genomic design for French meat sheep breeding program

J. Raoul, A.A. Swan & J.M. Elsen. 2017. World Congress on Genetics Applied to Livestock Production. Auckland, New-Zealand. **Accepted** to be included in the programme (oral presentation or poster) – under Review

Assessment of a genomic design for French meat sheep breeding program

J. Raoul[1,2], A.A. Swan[3] & J.M. Elsen[2]

[1] Institut de l'Elevage, BP 42118, 31321 Castanet-Tolosan, France
[2] GenPhySE, INRA, 31326 Castanet-Tolosan, France
[3] Animal Genetics and Breeding Unit, University of New England, 2350 Armidale, Australia

Summary

The majority of French meat sheep populations use both insemination (AI) and natural mating sires. Usually, AI sires are progeny tested and then the best are used as proven sires to produce male candidates. A breeding program based on genomic selection would be an alternative. Using a stochastic model where both individuals and their genome were simulated, we assessed a genomic design. The reference population was based on sires genotyped with a medium density panel (MD = 50K SNPs), including two to ten generations of sires born before the implementation of the genomic scheme and all sires born after. For sire replacement, newborn progeny were first preselected on parent average genomic estimated breeding values (GEBV) and then genotyped with a very low density panel (VLD = 1K SNPs). MD genotypes of candidates were imputed using the software Fimpute and GEBV computed with an animal model single step Genomic BLUP using the software Blupf90. Males selected for replacement were then genotyped with the MD panel to update the reference population. We assessed the sensitivity of the genetic gain to various sizes of the initial reference population, and compared the genetic gain of genomic and classical designs at a fixed cost for three different prices of a VLD genotyping. Within the range of values assessed for the initial reference population size and VLD genotyping cost, no significant difference across genomic schemes was observed. At a fixed cost, the annual genetic gain was higher for genomic designs (+18%) than for the classical design.

Keywords: breeding program, stochastic simulation, genomic selection, imputation, sheep

Introduction

Implementing a genomic breeding program is still a challenge for small meat sheep populations where estimated breeding values (EBVs) of both artificial insemination (AI) and naturally mated sires are less accurate due to low progeny numbers per sire, compared to dairy cattle for example. Reaching a high genomic prediction accuracy such as predicted by Daetwyler *et al.* (2008) would require establishment of a reference population based on a large number of animals, which is infeasible for small sheep breeding nuclei. Including records and genotypes from lower tiers of the population can solve this issue (Santos et al., 2017) though this might be difficult to achieve due to the lack of recording in commercial flocks. However, when the genotyped reference population included all sires and grandsires of candidates, Raoul *et. al* (2017) found an additional genetic gain of + 21.6% compared to a classical scheme when a full genomic scheme was adopted, including steps to impute 50K SNPs "Medium Density" (MD) genotypes from 1K SNPs "Very Low Density" (VLD) genotypes of male candidates.

French meat sheep breeding programs are based on collective management of males coordinated by breeding societies. Male candidates enter into collective stations where AI and natural mating sires are selected. AI males are progeny tested and at least two years later, the best AI sires based on progeny records are used as proven sires to produce male candidates. In a genomic design, male candidates might first be selected on parent average Genomic EBV (GEBV) and genotyped at MD after an imputation from a VLD panel. Then, replacements might be selected among candidates on their own GEBV. Male candidate genotypes would be purchased by the breeding society. To implement such a design, the cost needs to be similar to the current design cost as in sheep, the profitability at the nucleus level remains a critical factor (van der Werf and Banks, 2017). The current investment dedicated to the progeny testing would be allocated to the VLD genotypes of male candidates and the MD genotypes of sires (initial reference population and newly selected sires).

Focusing on a breeding program applied to a small population of purebred sheep, we used stochastic simulations to assess (1) the sensitivity of the additional gain to lower sizes of the initial reference population, and (2) the genetic gain of genomic and classical designs at a fixed cost for three different prices of VLD genotyping.

Materials and methods

Using a stochastic model, we simulated individuals and their genome. 50 K real genotypes were allocated to individuals as their genome. To obtain a founder population of 5000 females with a pedigree structure we first realized random reproductive cycles. Then we applied a classical design based on the progeny testing of AI sires over 10 years. Only one maternal selected trait was considered. The next fifteen years we applied either a classical design or a genomic design. The model included the establishment of the founder population, the QTL and phenotype simulation, the VLD and MD SNP panel simulation, the imputation of MD genotypes using FImpute software developed by Sargolzaei et al. (2014) and the estimation of breeding values by single step GBLUP using the Blupf90 software developed by Misztal et al. (1999). The key design information relevant to this study is described below, and the model is fully described in Raoul et al. (2017).

Around 5000 females divided into 10 flocks were recorded (one record per year) for the selected trait. Each year, half of the breeding females were selected on EBV and mated to an AI sire. Those which did not conceive to AI were mated to a natural mating sire, along with females specifically selected for natural mating. The number of progeny per dam depended on mode of reproduction and parity. Some dams were randomly culled after a reproductive cycle and the maximum parity was seven. Around 24% of dams were replaced per year by females preferentially chosen among newborn progeny from AI matings. There was no difference between the classical and genomic designs regarding female selection.

Male candidates were preselected among newborn progeny based on the parent average EBV in the classical design and on the parent average GEBV in the genomic design. In practice, all candidates were born from proven AI sires (classical design) and genomic AI sires (genomic design) given their genetic superiority. In the classical design, the ten candidates with the highest parent average EBV were selected and mated across flocks by AI to be progeny tested. Two years later, the five AI sires with the highest EBV (including progeny records) were selected as proven AI sires and used at most for four years. In the genomic design, male candidates were genotyped with the VLD panel and their MD genotypes imputed. The ten candidates with the highest GEBV were selected as genomic AI males and used at most for two years. Naturally mated male replacements were selected, based either on their parent average EBV in the classical design or GEBV in the genomic design, among candidates not selected for AI. Naturally mated males were used at most for four years, no further selection was applied.

In the genomic scenario, all selected sires were then genotyped with the MD panel to update the reference population.

The annual genetic gain was estimated as the regression slopes of the average true breeding value (TBV) of first parity females on time between years 10 and 15 and 10 and 25. The means and standard deviations presented are based on 50 replicates.

To assess the sensitivity of the genetic gain to the initial reference population size, we built different initial reference populations: all sires used when the design shifted to genomic selection (time 10) were included. In addition other sires used in the 2, 4, 6, 8 or 10 years prior to the start of genomic selection were included in the reference population. Whatever the size of the initial population, all sires selected in subsequent years were included in the reference population and the number of genotyped candidates per year was constant (270). To compare classical and genomic designs at a fixed cost we first determined the variable costs in the genomic and classical scheme: MD and VLD genotyping and the costs of maintaining AI sires. The money saved by using AI sires for a shorter time period in the genomic design was invested in MD genotypes (75€) of the initial reference population, in selected sires of subsequent years, and in VLD genotypes of male candidates. First we determined for three VLD genotyping costs, the optimal investment of either an increase the initial reference population size, or an increase in the number of candidates per year over a fifteen year time period. Then, the genetic gain of the classical and optimized genomic designs (time 10-25) were compared across VLD genotyping costs.

Results and Discussion

Table 1 gives the annual genetic gain for the classical design and genomic designs varying in the size of their initial reference population. We observed a slight sensitivity of the genetic gain Time 10-15 to the size of the initial reference population but, computed on time 10-25, the gain did not differ significantly. These results supported that the presence of candidate sires in the ongoing reference population was more important than the size of the initial reference population for the design we assessed. The average accuracy of candidate GEBVs in genomic designs was 0.12 points higher than the average EBV accuracy (0.39) in the classical scheme (results not shown).

Table 1. Annual Genetic gain for a classical and genomic designs varying in the size of their initial reference population (50 replicates)

Design		CS[1]	GS[2]				
Initial reference population size[3]			175	210	300	400	500
Years prior to start of GS[4]			2	4	6	8	10
Gain (genetic STD/year)							
Time 10-15	mean	0.173	0.173	0.175	0.179	0.185	0.188
	STD	0.021	0.021	0.023	0.020	0.020	0.022
Time 10-25	mean	0.167	0.193	0.192	0.196	0.197	0.197
	STD	0.014	0.018	0.019	0.019	0.017	0.018
Gain, Time 10-25 (%, classical =100)		100.0	115.4	114.8	117.2	118.0	118.4

[1]Classical design, no genomic information was used. Male replacements were selected on progeny records.
[2]Genomic design, male replacements were selected on their GEBV among 270 genotyped candidates.
[3]The initial reference population included sires used at time 10 and other sires used from previous years (from the last two years to the last ten years)
[4]Number of years prior to the start of genomic selection for inclusion of sires in the reference population

In Raoul *et al.* (2017), a reference population only including sires was sufficient to impute MD genotypes from VLD genotypes with a concordance rate close to 0.95, although a scenario where the imputation was performed without the pedigree information gave substantial losses of genetic gain (-22.2%). Previous studies (e.g. Clark et al. 2012) also highlighted the importance of close relatives in the reference population for the accuracy of genomic prediction, while Misztal et al. (2013) reported that using single step GBLUP increased accuracy compared to multi-step GBLUP. Given the heritability of the trait (0.25), it is highly likely that the moderate GEBV accuracy might be enhanced by a substantial larger reference population. However even if the increase was moderate, the genomic design was superior to the classical design.

Table 2 gives the maximal Annual Genetic gain for a genomic design at fixed costs according to three VLD genotyping costs and a MD genotyping cost of 75 €. The optimized combination of the size of the historical reference population and number of genotyped candidates per year is also given.

Table 2. Maximal annual genetic gain for a genomic design at fixed costs according to three VLD genotyping costs.

	VLD genotyping cost (€/unit)		
	7.5	15	22.5
Gain expressed in genetic standard deviations	0.197	0.194	0.193
Initial reference population size: years prior to start of GS[1]	Y=10	Y=8	Y=4
Number of VLD genotypes per year[2]	270	200	215

[1]Number of years prior to the start of genomic selection for inclusion of sires in the reference population
[2]VLD: very low density SNPs panel on male candidates

Irrespective of the VLD genotyping cost, the genomic design was superior to the classical design (0.167). Surprisingly, no significant differences in terms of genetic gain were observed according to the three VLD genotyping costs. The decrease in the historical reference population size as well as the decrease in the number of candidates had little impact on the genetic gain. In this study, for practical reasons, we considered that a maximum of 300 preselected males were genotyped and sires selected among 270 genotyped candidates (ten percent were culled or died for other reasons). The selection intensity of genomic sires is little affected by reducing the number from 270 (intensity = 2.186) to 200 (intensity = 2.063). This slight decrease in intensity and the low sensitivity to the initial reference population size might explain the stability of results across the range of genotyping costs considered.

We conclude that for small sheep populations, a genomic design based on a reference population including all sires and the VLD genotyping of male candidates is an alternative to current breeding programs based on progeny testing.

List of References

Clark SA, JM Hickey, HD Daetwyler, JH 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. GSE. 44:4.

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:E3395.

Misztal I. 1990. Complex models, more data: simpler programming. Interbull Bull. 20:33-42.

Misztal I, SE Aggrey, WM Muir. 2013. Experiences with a single-step genome evaluation. Poultry Sci. 92:2530–4.

Raoul J, AA Swan, JM Elsen. 2017. Using a very low density SNP panel for genomic selection in a breeding program for sheep. GSE. 49(1):76

Santos BFS, JHJ van der Werf, JP Gibson, TJ Byrne, PR Amer. 2017. Genetic and economic benefits of selection based on performance recording and genotyping in lower tiers of multi-tiered sheep breeding schemes. GSE.;49:10.

Sargolzaei M, JP Chesnais, FS Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478.

van der Werf J, R. Banks. 2017. Effective breeding programmes down under. JABG. 134:1–2.

Chapitre 5 – Discussion générale

Résultats obtenus

Les objectifs de ces travaux de thèse étaient d'apprécier, pour les populations ovines et caprines, l'intérêt technique et la rentabilité économique de l'utilisation d'un panel SNPs très basse densité.

Dans une première étude (Raoul *et al.*, 2016), nous avons évalué à l'aide d'un modèle déterministe, le gain génétique annuel additionnel et la rentabilité économique d'un accroissement du taux de filiation paternelle des femelles via l'assignation de parenté. En effet cette assignation peut être obtenue avec une très bonne précision (>95%) en utilisant un panel de 250 SNPs. Nous avons montré que l'accroissement du taux de filiation paternelle des femelles via l'assignation de parenté permettait un gain génétique additionnel mais n'était pas toujours économiquement rentable. Le type de programme de sélection (avec ou sans épreuve sur descendance pour les mâles d'insémination, monte naturelle exclusive) et la taille de la population commerciale devaient donc être considérés pour apprécier la rentabilité de l'assignation de parenté. Pour les programmes de sélection basés exclusivement sur la monte naturelle et un caractère peu héritable (0.10), +16.9% de gain génétique asymptotique additionnel a été obtenu pour un taux de filiations de filiations paternelles connues passant de 5 à 100%. Ce gain additionnel était restreint à +10% pour les programmes avec insémination mais sans testage et +5% lorsque les mâles d'insémination étaient évalués sur descendance. La

rentabilité économique était fortement dépendante du ratio gain/coût (gain monétaire dû à un écart-type génétique additionnel du caractère sélectionné divisé par le coût d'une assignation). Un ratio de cinq était nécessaire pour rentabiliser l'assignation au niveau du seul noyau de sélection tandis qu'un ratio de deux était suffisant lorsque le noyau représentait un tiers de la population totale (les revenus de l'assignation étant basés dans ce dernier cas sur l'augmentation de productivité de cette population totale et non du seul noyau).

Dans une seconde étude nous avons comparé à l'aide d'un modèle déterministe le gain génétique pour un caractère maternel et la rentabilité économique de différents modes de gestion des accouplements en présence d'un gène majeur d'ovulation, favorable à l'état hétérozygote, présent sur un autosome (Raoul *et al.*, 2017a) ou le chromosome X (Raoul *et al.*, 2017b). Le génotypage pour des gènes majeurs identifiés peut en effet être obtenu en incluant les SNPs causaux ou en très fort déséquilibre de liaison sur une puce très basse densité. Les génotypes étant connus, il s'agit de définir la stratégie de gestion c'est-à-dire déterminer la fréquence de chaque génotype chez les mâles et femelles et organiser les accouplements en fonction des génotypes (proportion de chaque type d'accouplements). Un modèle en générations discrètes montrait la sensibilité des solutions optimales à l'effet du gène et au coût unitaire du génotypage dans le cadre d'un locus autosomal. Des solutions basées sur un nombre limité de types d'accouplements (deux ou trois) étaient proches de la solution optimale en termes de rentabilité économique mais pouvaient se traduire par des valeurs très différentes des critères techniques (gain génétique, taux d'hétérozygotes dans le noyau et la population commerciale, nombre de génotypages requis). Un modèle à générations chevauchantes adapté au cas réel du programme de sélection Lacaune viande Ovi-Test a montré que la stratégie basée sur deux types d'accouplements (mâles hétérozygotes x femelles

non porteurs, mâles non porteurs x femelles hétérozygotes) était proche de la solution optimale, entre 93 et 100% de la rentabilité maximale, quel que soit le coût unitaire du génotypage. Les résultats obtenus pour un gène majeur présent sur le chromosome X montre que sa gestion était rentable par rapport à son éradication mais se traduisant par une perte de progrès génétique plus importante que dans le cas d'un gène majeur autosomal.

Dans une troisième étude (Raoul *et al.*, 2017c) nous avons estimé à l'aide d'un modèle stochastique le gain génétique pour un caractère maternel dans le cas d'un programme de sélection classique ou génomique. Dans le design classique, les pères à mâles étaient sélectionnés après l'épreuve de la descendance. Dans le design génomique, les pères à mâles étaient sélectionnés sur index génomique. Plusieurs scénarios génomiques ont été évalués, variant par la quantité d'information génomique disponible (mères génotypées ou non) et la qualité des génotypes des candidats et mères (génotypés en moyenne densité ou en très basse densité puis imputés vers la moyenne densité). Nous avons montré la supériorité d'un schéma génomique, +22%, au schéma classique lorsque l'ensemble des pères étaient génotypés en moyenne densité et les candidats en très basse densité (1000 SNPs) puis imputés vers la moyenne densité. Lorsque les mères étaient également génotypées (basse densité puis imputation), le gain génétique annuel doublait et était largement supérieur (+42%) à celui du schéma classique. Une étude complémentaire basée sur le même modèle (Raoul *et al.*, 2017d) a montré que, pour un schéma génomique évalué sur 15 années, ce gain était très peu sensible à la taille de la population de référence initiale malgré un gain génétique plus faible sur les premières années de sélection en cas de taille très limitée. Comparés à coût fixe, les schémas génomiques génèrent un gain génétique annuel supérieur au schéma classique. La solution optimale entre taille de la population de référence initiale et nombre de candidats mâles

génotypés par cycle a donné un gain génétique additionnel stable pour la gamme de coûts évalués (génotypage en très basse densité compris entre 7.5 et 22.5 euros).

Développement de l'assignation dans les populations ovines et caprines françaises

L'intérêt technique et économique de l'assignation de parenté a fait l'objet de l'étude présentée au chapitre deux. Compte tenu du coût actuel de la technique, son utilisation visant à accroître la quantité de généalogie n'est économiquement pas rentable même dans le cas où un gain génétique additionnel substantiel a été calculé. L'utilisation du modèle pour le traitement de quelques cas d'étude montre que le coût d'opportunité de l'assignation soit son coût maximum visant à un équilibre entre recettes et coûts à 20 ans se situe entre 2 Euros par assignation pour les caprins (résultats non présenté dans l'article) à 7.1 Euros en race ovine allaitante Romane. Ce calcul de coût d'opportunité intègre l'expression du gain génétique additionnel au niveau de la population commerciale c'est-à-dire le surplus de recettes générées dans les élevages commerciaux. Si le surplus de recettes était comptabilisé au niveau du seul noyau, le coût d'opportunité diminuerait et s'écarterait encore davantage du coût actuel de l'assignation. Compte tenu de ces éléments, l'utilisation de panels SNPs très basse densité à des fins uniques d'assignation de paternité pour accroître la quantité de filiations paternelles des femelles n'est pas opportune aux coûts actuels. Alternativement, l'utilisation des panels très basse densité pour d'autres valorisations (*e.g.* gestions de gènes majeurs) pourrait rendre cette assignation économiquement attractive.

Dans l'étude présentée au chapitre deux, les conséquences de l'assignation de parenté sur la planification des accouplements raisonnées n'ont pas été considérées. En méthode classique, les femelles sont accouplées à un seul mâle tandis qu'en méthode alternative elles sont accouplées à un groupe de mâles parmi lesquels le père de chaque nouveau-né est assigné sur marqueurs moléculaires. Si les femelles possédant les plus hautes valeurs génétiques ne peuvent être luttées par les mâles possédant les plus hautes valeurs génétiques (accouplement non dirigés), l'assignation de parenté pourrait entraîner une moindre efficacité du programme de sélection. Ce risque est faible dans les troupeaux de taille conséquente mais existe dans les troupeaux de petite taille et pour lesquels le niveau génétique des mâles est hétérogène.

Au-delà de la simplification du travail, un autre intérêt du panel très basse densité est sa capacité à détecter des erreurs de filiation. Les conséquences techniques et économiques des erreurs de filiations ont été appréciées, principalement dans les programmes de sélection mettant en œuvre un testage sur descendance des mâles d'insémination bovins (Gerldermann *et al.*, 1986 ; Ron *et al.*, 1996 ; Israel et Weller, 2000 ; Roughsedge *et al.*, 2001 ; Senneke *et al.*, 2004 ; Sanders *et al.*, 2006 ; Parlato et van Vleck, 2012 ; Winkelman, 2013), en caprin (Garritsen *et al.*, 2015) et en porcins (Long *et al.*, 1990). Ces erreurs affectent la qualité des évaluations génétiques, le classement des reproducteurs et en conséquence le progrès génétique de la population en sélection. Un gain génétique additionnel peut donc être espéré par la réduction des erreurs de filiation, notamment en élevage allaitant pour lequel les taux d'erreur observés sont généralement supérieurs. Ce gain devrait être pris en compte pour évaluer l'intérêt de l'assignation de parenté comme alternative aux méthodes classiques d'obtention des paternités (insémination et monte naturelle en paternité).

Gestion des gènes majeurs d'ovulation dans les populations ovines

La gestion des mutations entrainant une hyper-prolificité des femelles porteuses (gènes majeurs d'ovulation) est un enjeu prépondérant dans les populations ovines allaitantes françaises (Bodin *et al.*, 2011). Les cas étudiés dans Raoul *et al.*, 2017a et 2017b, correspondent à des situations pour lesquelles l'effet de la mutation causale est délétère à l'état homozygote, situation usuelle dans le cas des gènes majeurs d'ovulation.

L'étude de la relation entre génotype et prolificité mais aussi entre génotype et autres caractères (notamment maternels) dans chacun des fonds polygéniques est un préalable à l'établissement de règles de gestion. La quantité d'informations disponibles pour réaliser ces études est très variable selon les populations. Le démarrage du génotypage systématique des agnelles a démarré en 2010 en race Lacaune viande et plus tardivement pour les races Noire du Velay (2015), Grivette (2016) et Mouton vendéen (2016). Compte-tenu de la sensibilité de la solution optimale à l'effet du gène et au coût unitaire du génotypage, il est difficile de fournir, excepté pour la race Lacaune viande (Raoul *et al.*, 2017a), une recommandation aux gestionnaires des populations ovines françaises avant d'avoir analysé les performances de leurs femelles génotypées.

La gestion optimale a été déterminée avec un modèle statique (Raoul *et al.*, 2017a, 2017b) : pour chacune des solutions, la fréquence de l'allèle favorable et la proportion de chaque type d'accouplement étaient fixes au cours du temps. La phase de transition entre l'état actuel et l'état optimal (ou proche de) n'a pas été étudiée. Pour maximiser un critère technique (gain génétique) évalué sur un temps suffisamment large, on peut admettre que la phase de

transition soit négligeable et que les solutions optimales obtenues restent valides. Pour un critère économique, l'horizon évalué ne peut être indéfini et la phase de transition entre état initial et état final devrait être prise en compte. Nous avons assumé que les solutions optimales obtenues restaient également valides à condition que la différence des fréquences alléliques entre états initial et final ne soit pas trop importante.

Un modèle dynamique permettrait de simuler une évolution des fréquences alléliques et des proportions de type d'accouplements au cours du temps et d'intégrer la phase de transition entre état actuel et état « optimal ». Le nombre de variables à optimiser serait alors plus important, soit une combinaison de 6 variables (types d'accouplements) dans le modèle statique et de 6t variables, t étant le nombre de cycles de reproduction, dans le modèle dynamique. En plus du rallongement du temps de calcul, cet accroissement du nombre de variables complexifie la formulation du problème, notamment des contraintes et peut entraîner des problèmes de convergence de l'algorithme utilisé pour rechercher la solution optimale. A noter que selon la longueur de l'horizon, les fréquences alléliques peuvent ne pas être stabilisées en fin d'horizon.

Dans Raoul *et al.* (2017a), la prolificité était le seul caractère sélectionné tandis que dans Raoul *et al.* (2017b), seul un caractère maternel, apparenté à la valeur laitière était sélectionné. Un accroissement infini de la prolificité moyenne de la population n'est pas souhaitable car conduirait à des problèmes d'hyper-prolificité. La prolificité moyenne de la population pourrait être plafonnée à une valeur maximale (et optimale). Cet objectif pourrait être atteint plus ou moins vite en fonction du niveau actuel de la population, par l'amélioration de la valeur polygénique des femelles et/ou l'accroissement de la fréquence des femelles hétérozygotes. Cette notion de valeur cible n'a pas été prise en compte dans Raoul *et al.* (2017a).

Au-delà de la valeur cible, la canalisation de la variance de la taille de portée (*i.e.* réduction de la variabilité des tailles de portées) n'a pas été abordée.

Le gain phénotypique du caractère sélectionné pourrait être en compétition avec les gains polygéniques espérés pour d'autres caractères inclus dans l'objectif de sélection. Traiter cette situation nécessiterait l'utilisation d'un modèle dynamique dans lequel les pondérations d'un objectif multi-caractères, susceptibles d'évoluer au cours du temps, seraient incluses dans les variables à optimiser.

L'optimisation des accouplements se limite aux situations pour lesquelles un seul gène majeur est connu. Dans certaines populations ovines telles qu'en races Lacaune viande et Noire du Velay, plusieurs mutations, identifiées ou suspectées, ségrégent (Bodin *et al.*, 2011). En général, ces autres mutations sont présentes à une fréquence assez faible. Il serait envisageable de modéliser une telle situation à condition de connaître les effets conjugués des allèles favorables.

Pour des mutations portées sur des autosomes, le modèle reviendrait à traiter la situation présentée dans Raoul *et al.* (2017a), à condition que la prolificité des femelles porteuses hétérozygotes des deux mutations soit similaire à celle des porteuses hétérozygotes d'une seule mutation. En cas d'hyper-prolificité trop importante des femelles porteuses hétérozygotes des deux mutations, le modèle peut-être extrapolé à trois génotypes pour les femelles (non porteuse, porteuse hétérozygote locus 1, porteuse hétérozygote locus 2) et neuf génotypes pour les mâles soit 27 types d'accouplements possibles. Si d'un point de vue modélisation, ce problème peut être considéré, on peut admettre que le choix d'une seule mutation à gérer, l'autre étant éradiquée, soit une solution plus pratique.

L'utilisation d'un panel très basse densité pour la sélection génomique

Grâce à une précision élevée des prédictions génomiques permise par la qualité de la population de référence et une réduction conséquente de l'intervalle de génération, les gains espérés par le développement de la sélection génomique sont très conséquents en bovin lait (Schaeffer, 2006). En Lacaune lait, ce gain est *a priori* moindre du fait d'une précision inférieure des prédictions génomiques et d'un intervalle de génération initial plus faible (Baloche *et al.*, 2014). Dans le cadre de nos simulations, le gain additionnel permis par le schéma génomique, sans incorporation des femelles, apparaît également moindre par rapport aux bovins laitiers.

Deux raisons similaires à celles énoncées pour la race Lacaune lait peuvent être évoquées : i) la taille de notre population de référence était très faible, de l'ordre de quelques centaines de pères à un peu plus d'un millier en fin de simulation et comprenait en majorité des mâles de monte naturelle (80%) avec chacun peu de filles, conduisant à une précision moindre par rapport aux bovins lait des prédictions génomiques calculées pour les candidats, ii) la réduction de l'intervalle de génération liée à la suppression du testage sur descendance est moindre car l'intervalle initial est plus faible et seulement une partie du renouvellement (100% des pères et environ un tiers des femelles) est issue des mâles améliorateurs dans le schéma classique.

La mise au place de la sélection génomique dans les principales races bovines laitière comme alternative au testage sur descendance des taureaux a été proposée sur la base de population de référence contenant plus d'un millier de taureaux connus sur descendance (Fritz *et al.*, 2010). En petits ruminants, la précision moindre des phénotypes (définis ici comme

l'ensemble de l'information connue sur l'individu et ses apparentés) et des tailles efficaces généralement plus importante rendaient nécessaire d'accroitre le nombre d'individus de la population de référence pour obtenir, sur la base des formulations déterministes (Daetwyller *et al.*, 2008 ; Goddard, 2009), des précisions similaires à celle des bovins laitiers. Il apparaissait ainsi difficile de proposer ce type de schéma aux populations ovines et caprines compte-tenu de leur situation, hormis pour quelques races disposant d'effectifs en testage conséquents. Des études ultérieures ont montré le gain en précision des prédictions génomiques apporté par la présence dans la population de référence d'individus apparentés (Clark *et al*., 2012) et par une méthode d'évaluation génomique par single step plutôt que multi-step (Misztal *et al.*, 2013). En accord avec ces études, les résultats que nous avons obtenus montrent, à coût équivalent, une supériorité du schéma génomique par rapport au schéma classique malgré un effectif limité de la population de référence.

Le gain apporté par les génotypages femelle semble limité pour la population Holstein (Koivula *et al.*, 2016) mais d'autant plus conséquent que la population de mâles génotypés est faible et leur phénotype imprécis pour les autres populations bovines laitières (McNugh *et al*., 2011 ; Buch *et al*., 2012 ; Pryce *et al*., 2012; Calus *et al*., 2013 ; Gonzalez-Recio *et al*., 2014 ; Thomasen *et al*., 2014 ; Gao *et al*., 2015 ; Plieschke *et al*., 2016 ; Su *et al*., 2016). Nos résultats montrent effectivement que le gain d'un schéma génomique par rapport à un schéma classique peut être doublé en incorporant les génotypages des femelles dans l'évaluation génomique. L'utilisation de l'assignation de parenté par les éleveurs comme alternative aux lots de paternité contribuerait ainsi à accroître le progrès génétique dans le cadre d'un schéma génomique. De même, la mise en place d'une sélection génomique dans un schéma basé sur l'insémination et génotypant systématiquement les agnelles pour la gestion d'un gène majeur (*e.g.* programme Ovi-test) devrait être très bénéfique en termes d'efficacité de la sélection.

Dans notre design, la précision de l'imputation pour les candidats males (et femelles) repose sur la capacité des panels très basse densité à caractériser la co-ségrégation entre marqueurs au sein des lignées paternelles. Quand les liens de parenté sont ignorés, l'imputation est réalisée à partir du déséquilibre de liaison entre marqueurs. La précision de l'imputation est alors beaucoup plus faible compte-tenu du faible déséquilibre de liaison existant entre la plupart des marqueurs moyenne densité présents et absents sur la puce très basse densité (Habier *et al.*, 2009 ; Huang *et al.*, 2009 ; Wellmann *et al.*, 2013 ; Toghiani *et al.*, 2016). La précision des prédictions génomiques des candidats est également favorisée par la présence des apparentés proches dans la population de référence. Les résultats obtenus dans Raoul *et al.*, (2017c, 2017d) impliquaient une organisation spécifique reposant sur une organisation très collective de la sélection et en particulier de la maîtrise de la voie mâle. Ce type d'organisation dans laquelle tous les pères, grands-pères des candidats sont présents dans la population de référence n'est pas facilement envisageable pour des populations avec un fonctionnement moins collectif. On peut en effet craindre que certains détenteurs fassent le choix de ne pas génotyper les pères ce qui pourrait affecter la capacité à imputer le génotypage de leur fils et petits-fils.

Les résultats obtenus impliquaient également que le critère de sélection soit contrôlé dans l'ensemble des élevages du noyau. Si le critère était mesuré uniquement sur une partie de la population, la précision moyenne des prédictions génomiques (moins d'information disponible) serait probablement affectée et en particulier la précision des candidats issus de pères dont la descendance serait partiellement voire non contrôlée. Les résultats obtenus ne peuvent donc pas être extrapolés à des caractères contrôlés sur une partie du noyau seulement (*e.g.* difficile à contrôler).

Dans Raoul *et al.*, (2017c, 2017d), les scenarios génomiques étaient comparés à un schéma classique mettant en œuvre un testage sur descendance des mâles d'insémination. Les ressources financières actuellement dédiées au testage pourraient être affectées aux besoins en génotypages engendrés par la sélection génomique. De ce fait les gains génétiques annuels produits par les différents schémas pouvaient être comparés à coût identique (Raoul *et al.*, 2017d).

L'intérêt technique de la sélection génomique pour les populations sans testage mais avec insémination et les populations en monte naturelle exclusive n'a pas été évalué. La mise en place de la sélection génomique dans ces types de programme représenterait une surcharge financière qui ne serait *a priori* pas compensable par une diminution des coûts liée à la modification de leur organisation initiale. Les effectifs de mâle de monte naturelle sont contraints par les besoins en reproduction et les effectifs de mâles d'insémination déjà adaptés aux besoins de production de doses dans le contexte de l'IA en semences fraîches et d'une rotation rapide des mâles. Il n'y a donc pas de réduction espérée de l'intervalle de génération, d'ores et déjà très court pour ces programmes tel que c'est le cas par exemple pour les programmes de sélection porcins (Tribout *et al.*, 2012). Le surplus éventuel de progrès génétique annuel proviendrait d'un gain de précision des prédictions des valeurs génétiques des parents des individus présélectionnés et de la possibilité de réaliser une pression de sélection sur index génomiques propres parmi les individus présélectionnés (candidats génotypés).

La comparaison entre schéma classique et génomique ne pourrait être réalisée à coût constant. L'intérêt économique de la sélection génomique serait apprécié par une approche coûts/recettes comparant les coûts générés par les besoins en génotypages et le surplus de

recettes généré dans la population. Une supériorité technique du schéma génomique par rapport au schéma classique basé sur l'IA mais sans évaluation sur descendance est attendu. Dans ce cas, le design génomique est identique à celui modélisé dans Raoul *et al*. (2017c) tandis que le schéma classique sans testage est moins performant que le schéma classique avec testage.

Pour la comparaison des schémas génomique et classique basés sur l'utilisation exclusive de la monte naturelle, le résultat serait plus incertain. La prise en compte d'information génomique et la sélection sur index propres pourrait se traduire par un gain additionnel mais celui-ci est difficilement quantifiable *a priori*.

Le cas des populations ovines allaitantes sélectionnant les aptitudes maternelles et bouchères mais réalisant un testage de leurs mâles d'insémination sur aptitudes bouchères uniquement est un cas différent qui n'a pas été évalué. Dans ces programmes, les mâles d'insémination sont sélectionnés après leur première utilisation. L'intérêt de mobiliser les ressources actuellement dédiées au testage des aptitudes bouchères (coût élevé du phénotypage hors noyau et en abattoir) pour l'adoption d'un schéma génomique pourrait être évalué. L'intérêt dépendrait dans ce cas des évolutions relatives des gains génétiques sur les aptitudes bouchères et maternelles.

Limites des modèles utilisés

Caractère sélectionné

Les objectifs de sélection et l'organisation des programmes de sélection ovins et caprins présentent une diversité importante entre pays. C'est le cas par exemple pour les ovins allaitants en termes de dispositifs génétiques (i.e. acteurs et organisations des structures impliquées dans l'amélioration génétique), de caractères sélectionnés, de méthodes d'évaluations génétiques et de modes de diffusion (de Rancourt et Raoul, 2013). Les programmes de sélection modélisés dans le cadre de cette thèse ont visé à traiter le cas de populations françaises. Dans l'ensemble des modèles utilisés, la sélection portait sur un seul caractère exprimé par les femelles. Il s'agissait d'une simplification par rapport à la réalité.

Pour les populations laitières, certains caractères mâles sont évalués (*e.g.* aptitudes à la production de semence) mais les caractères sélectionnés sont essentiellement exprimés par les femelles (Palhière *et al.*, 2015 ; Barillet *et al.*, 2016). La sélection sur un seul caractère peut être assimilée à une sélection sur index multi-caractères et le progrès génétique annuel par caractère déduits du progrès génétique annuel sur l'index.

Pour les populations allaitantes, l'objectif de sélection inclut également les caractères bouchers (Cheype *et al.*, 2013). Les évaluations sont uni-caractère et la sélection des aptitudes maternelles et bouchères est réalisée sur seuils indépendants. Mise à part la croissance post-sevrage, évaluée en ferme pour un nombre limité d'individus, les caractères bouchers sont généralement évalués et sélectionnés en station de contrôle individuel sur performances propres et parfois sur performances de la descendance pour les mâles d'insémination par la procréation d'agneaux hors du noyau de sélection (Ménissier et Bouix, 1992).

L'accroissement du taux de filiations paternelles ne peut pas affecter la précision des évaluations des caractères bouchers car l'information maternelle n'est pas considérée dans les modèles d'évaluation génétique en station et sur descendance (Bouix, communication personnelle). La sélection bouchère, ou bien encore la sélection des aptitudes à la production de semences, affectent le nombre de candidats. Ce phénomène a été pris en compte dans le calcul des différentielles de sélection pour le caractère maternel. Par contre il n'y a pas de conséquences attendues sur le progrès génétique annuel boucher donc celui-ci n'a pas été quantifié dans le cadre de cette étude.

Pour les mêmes raisons, les aptitudes bouchères étant considérées indépendantes des aptitudes maternelles et du génotype au locus du gène majeur d'ovulation, le progrès génétique annuel des aptitudes bouchères n'a pas été quantifié pour les populations ovines allaitantes. Cette indépendance entre aptitudes maternelles et bouchères pour les races ovines françaises n'est pas étayée dans la littérature mais est de fait institutionnalisée par l'utilisation de modèles uni-caractères pour l'évaluation génétique officielle (Tiphine *et al.*, 2013).

Concernant la comparaison de programmes génomiques, le choix de ne pas considérer les aptitudes bouchères peut conduire à sous-estimer l'intérêt du schéma génomique pour les populations allaitantes. Les animaux contrôlés pour ce caractère sont ceux génotypés en très basse densité et, s'ils sont sélectionnés, en moyenne densité. Cela ouvre donc la possibilité d'un gain supplémentaire pour ce caractère lié à un gain en précision des prédictions de la valeur génétique des mâles voire à la disponibilité d'une prédiction pour les femelles dans le cas d'une évaluation ssGBLUP du caractère. Le gain est cependant probablement assez faible car les héritabilités des aptitudes bouchères sont relativement élevées, entre 0.25 et 0.5 selon les caractères, et la sélection réalisée à un âge précoce.

L'analyse de la rentabilité économique des plans de sélection

Le choix de l'horizon est partiellement arbitraire. Nous avons considéré un horizon large de 60 années (soit entre 12 et 15 générations en fonction des programmes de sélection) dans l'étude sur l'assignation, différents horizons (court terme : 2-6 générations ; long terme : 16-20 générations et large horizon : 2-20 générations) dans l'étude sur la gestion de gènes d'ovulation et un horizon court, 15 années soit 3-4 générations, dans l'étude sur l'utilisation de puces très basse densité pour la sélection génomique. Les modèles de flux de gènes dans le cas de modélisation déterministe ou les modèles stochastiques permettent de caractériser le niveau génétique des groupes d'individus ou individus au cours du temps et donc de caractériser le décalage entre coûts et recettes. Compte-tenu de ce décalage, les coûts engagés en fin d'horizon induiront des recettes qui seront partiellement (ou pas) comptabilisées. Le choix d'un horizon large permet de minimiser cet effet de « bordure ». Lorsque l'étude revient à mesurer l'effet d'un investissement (*e.g.* intérêt de l'assignation), le délai de récupération du capital investi est un critère intéressant car indépendant du choix d'un horizon. Lorsque l'étude porte sur la comparaison de différentes stratégies de gestion, chacune variant par leurs besoins en investissements (*e.g.* besoins en génotypages des différentes stratégies de gestion de gènes majeurs d'ovulation), les choix d'un horizon et d'une approche coût-recettes s'imposent.

Les temps de calcul, et donc les types de modèle, influencent également le choix de l'horizon. Pour l'étude de l'assignation, les temps de calcul pour un jeu de paramètres d'entrée étaient de quelques secondes. Pour l'étude sur la gestion des gènes majeurs, incluant la recherche des valeurs optimales d'un vecteur de décision de 6 éléments (les types d'accouplement), les temps de calcul étaient de quelques minutes. Dans le cadre du modèle stochastique développé pour l'étude des schémas génomiques, les temps de calcul pour un

scénario (15 années) variaient de quelques heures à une dizaine d'heures en fonction du nombre d'individus génotypés et de la réalisation ou non d'étapes d'imputation.

Les approches coûts-recettes intègrent généralement les recettes réalisées au sein de la population commerciale en s'appuyant sur une modélisation des flux de gènes du noyau vers les élevages commerciaux. Si d'un point de vue théorique il convient, lors de l'évaluation d'une stratégie, de prendre en compte les recettes générées dans l'ensemble de la population, la concentration des coûts au niveau du seul noyau de sélection peut poser un problème pratique : le surplus de recettes générées au niveau de la population commerciale ne se traduit pas forcément par un flux financier des élevages commerciaux vers les élevages du noyau. Une stratégie peut donc être rentable au niveau de la population et déficitaire au niveau du noyau de sélection (Raoul *et al.*, 2016). La comparaison à coûts constants lorsqu'elle est possible évite ce problème. La comparaison peut se faire alors uniquement sur les seuls critères techniques, gain génétique annuel et évolution de la consanguinité.

La comparaison à coûts constants semble également plus robuste que l'approche couts-recettes. Les coûts variables entre stratégies sont en général facilement identifiables et assez aisés à estimer. Les recettes dépendent du gain économique par unité du caractère et de l'évolution du niveau génétique dans les différentes strates de la population (âge des individus, élevages commerciaux ou du noyau, …) qui sont multipliés par le nombre d'expressions du caractère. En général le nombre d'expressions du caractère est très important : pour un caractère femelle exprimé annuellement dans un noyau de 5000 femelles et une population commerciale de 10 000 femelles ayant en moyenne trois productions, le nombre (non actualisé) d'expressions sur un horizon économique de dix ans est de l'ordre de la centaine de milliers. Une imprécision, même faible, de l'évolution génétique dans la

population et du gain économique par unité du caractère est donc démultipliée par le nombre d'expressions du caractère. L'hypothèse de la stabilité du gain économique par unité du caractère quel que soit le mode de production de l'élevage commercial, est une source d'imprécision. De même la connaissance souvent partielle des flux d'animaux entre noyau et élevages commerciaux affecte l'estimation du niveau génétique dans les élevages commerciaux. Il conviendrait donc dans une approche coûts-recettes d'analyser la sensibilité des résultats au gain économique par unité du caractère et à la qualité du transfert du progrès génétique vers la population commerciale.

L'analyse à coûts fixes serait une option préférable mais la capacité à rendre compte de difficultés d'ordre organisationnel ou sanitaire de certaines stratégies existe quel que soit le type d'analyses économiques. Dans Raoul *et al.* (2017a), la stratégie basée sur la spécialisation de quelques élevages dans la production de mâles porteurs homozygotes (quelques pourcents des accouplements du noyau) ont donné des résultats proches en termes de profit qu'une stratégie basée sur deux types d'accouplement en proportions équilibrées, réalisable par tous les élevages du noyau. Pour un même profit, la seconde stratégie serait plus pratique et moins exposé au risque sanitaire. De telles considérations, dont l'impact économique serait complexe à estimer, sont difficiles à intégrer lors de la formalisation de la fonction objective.

Conclusion

A l'aide de simulations déterministes et stochastiques, l'intérêt technique et économique de l'utilisation d'un panel de marqueurs moléculaires très basse densité a été évalué dans les populations en sélection ovines et caprines et permis d'obtenir les résultats suivants : i) utiliser un tel panel pour accroître, quand elle est limitée, la quantité de filiations paternelles n'est pas toujours rentable, ii) la stratégie de gestion des gènes d'ovulation qui maximise la rentabilité économique du plan de sélection a été déterminée par optimisation et des stratégies simples à implémenter, qui donnent des rentabilités proches de la rentabilité maximale, ont été proposées, iii) un programme de sélection génomique basé sur un panel très basse densité, permet à coût constant une efficacité supérieure aux programmes basés actuellement sur le testage sur descendance des mâles.

Références bibliographiques

Aguilar I, Misztal I. 2008. Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. Journal of Dairy Sciences. 91:1669–1672.

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. Journal of Dairy Sciences. 93:743–52.

Amer PR, McEwan JC, Dodds KG and Davis GH. 1998. Cost benefit analysis of commercial use of the Inverdale prolificacy gene. In Proceeding of the New Zealand Society of Animal Production, 58, 157–160.

Analla M, A Sanchez-Palma, A Munõz-Serrano, and JM Serradilla. 1995. Simulation analysis with BLUP methodology of different data structures in goat selection schemes in Spain. Small Ruminants Research. 17:51–55.

Auvray B, JC McEwan, SA Newman, M Lee, and KG Dodds. 2014. Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. Journal of Animal Sciences. 92:4375–4389.

Astruc JM, F Barillet, A Barbat, V Clement, and D Boichard. 2002. Genetic evaluation of dairy sheep in France. In: 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France. p. 231–234.

Astruc JM, G Baloche, D Buisson, J Labatut, G Lagriffoul, H Larroque, C Robert-Granié, A Legarra, F Barillet. 2016. La sélection génomique des ovins laitiers en France. INRA Productions Animales, INRA Editions. 29 (1) 41-56.

Badke YM, Bates RO, Ernst CW, Fix J, Steibel JP. 2014. Accuracy of Estimation of Genomic Breeding Values in Pigs Using Low-Density Genotypes and Imputation. Genes, Genomes, Genetics. 4:623–31.

Baloche G, A Legarra, G Sallé, H Larroque, JM Astruc, C Robert-Granié, and F Barillet. 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. Journal of Dairy Sciences. 97:1107–1116.

Barillet F, G Lagriffoul, PG Marnet, H Larroque, R Rupp, D Portes, F Bocquier et JM Astruc. 2016. Objectifs de sélection et stratégie raisonnée de mise en œuvre à l'échelle des populations de brebis laitières françaises. INRA Productions Animales, INRA Editions. 29 (1) 19-40.

Bell AM, JM Henshall, S Gill, K Gore, JW Kijas. 2013. Success rates of commercial SNP based parentage assignment in sheep. In: Proc. Assoc. Adv. Anim. Breed. Genet. 20: p. 278–281.

Besbes B, 1990. Evaluation du contrôle de la paternité chez les ovins : comparaison des réponses à la sélection obtenues en présence et en absence du contrôle de la paternité. PhD Diss. CIHEAM, Zaragoza, Spain.

Bichard M, 1971. Dissemination of genetic improvement through a livestock industry. Animal Production. 13:401–411.

Bijma P, and JA Woolliams. 1999. Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics 151:1197–1210.

Bijma P, and JA Woolliams. 2000. On the relation between gene flow theory and genetic gain. Genetics Selection Evolution. 32:99–104.

Bodin L, E Di Pasquale, S Fabre, M Bontoux, P Monget, L Persani and P Mulsant. 2007. A Novel Mutation in the Bone Morphogenetic Protein 15 Gene Causing Defective Protein Secretion Is Associated with Both Increased Ovulation Rate and Sterility in Lacaune Sheep. Endocrinology. 148(1):393-40.

Bodin L, Raoul J, Demars J, Drouilhet L, Mulsant P, Sarry J, Tabet C, Tosser-Klopp G, Fabre S, Boscher MY, Tiphine L, Bertrand C, Bouquet PM, Maton C, Teyssier J, Jouannaux C, Hallauer J, Cathalan D, Gueux J and Pocachard M. 2011. Etat des lieux et gestion pratique des gènes d'ovulation détectés dans les races ovines françaises. 18èmes Rencontres Recherches Ruminants, Paris, France, 393–400.

Boichard D, F Guillaume, A Baur, P Croiseau, MN Rossignol, MY Boscher, T Druet, L Genestout, JJ Colleau, L Journaux, V Ducrocq, and S Fritz. 2012. Genomic selection in French dairy cattle. Animal Production Science. 52:115.

Bolormaa S, K Gore, JHJ van der Werf, BJ Hayes, and HD Daetwyler. 2015. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. Animal Genetics. 46:544–556.

Boulanger A, F Grosclaude and MF Mahé. 1984. Polymorphisme des caséines alpha s1 et alpha s2 de la chèvre(Capra hircus). Genetics Selection Evolution. 16 (2), 157-176.

Brard S, A Ricard. 2015. Is the use of formulae a reliable way to predict the accuracy of genomic selection? Journal of Animal Breeding and Genetics. Vol. 132, 3: 207–217.

Browning SR, and BL Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics. 81:1084–1097.

Buch LH, Kargo M, Berg P, Lassen J, Sørensen AC. 2012. The value of cows in reference populations for genomic selection of new functional traits. Animal. 6:880–6.

Buisson D. 2012. Sélection génomique des races ovines laitières françaises : analyse des schémas actuels, première modélisation de scénarios génomiques et bilan technico- économique. In: Sciences agricoles.

Calus MPL, de Haas Y, Veerkamp RF. 2013. Combining cow and bull reference populations to increase accuracy of genomic prediction and genome-wide association studies. Journal of Dairy Sciences. 96:6703–15.

Capgènes. Le schéma de sélection –Site web : http://www.capgenes.com/spip.php?article78. Accédé le 04/09/2017.

Carillier C, H Larroque, and C Robert-Granié. 2014. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genetics Selection Evolution. 46:67.

Carillier-Jacquin C, Larroque H and Robert-Granié C. 2016. Including α s1 casein gene information in genomic evaluations of French dairy goats. Genetics Selection Evolution, 48:54.

Carillier-Jacquin C, Larroque H and Robert-Granié C. 2017. Vers une sélection génomique chez les caprins laitiers. INRA Productions Animales, INRA Editions, 30 (1) 19-30.

Casellas J, Piedrafita J. 2015. Accuracy and expected genetic gain under genetic or genomic evaluation in sheep flocks with different amounts of pedigree, genomic and phenotypic data. Livestock Sciences. 182:58–63.

Cheype A, J Guerrier, F Tortereau, D François, JP Poivey, K Chile and J Raoul. 2013. Economical weighting of breeding objectives and definition of total merit indexes in BMC sheep breed. In: EAAP annual meeting. Nantes, France. Book of abstracts No. 19, p. 319.

Cheype A. 2016. Evaluation sur descendances des aptitudes bouchères des races ovines allaitantes. Résultats des séries 2014 et mise en testage 2015. Compte rendu 00 16 203 026. Département Génétique et Phénotypes. Institut de l'élevage.

Clark SA, JM Hickey, HD Daetwyler, and JH 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.

Clarke SM, HM Henry, KG Dodds, TWD Jowett, TR Manley, RM Anderson, and JC McEwan. 2014. A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. PLoS ONE 9:e93392.

Cleveland MA, and JM Hickey. 2013. Practical implementation of cost-effective genomic selection in commercial pig breeding using imputation. Journal of animal science. 91:3583–3592.

Clop A, F Marcq, H Takeda, D Pirottin, X Tordoir, B Bibé, J Bouix, F Caiment, JM Elsen, F Eychenne, C Larzul, E Laville, F Meish, D Milenkovic, J Tobin, C Charlier and M Georges. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. 2006. Nature Genetics; New York 38.7 : 813-8.

Costard AD, ZG Vitezica, CR Moreno, and JM Elsen. 2008. A dynamic deterministic model to optimize a multiple-trait selection scheme. Journal of Animal Science. 87:885–894.

Costard AD, and JM Elsen. 2011. Optimization of Gene-Assisted Selection in Small-Sized Populations: Comparison of Deterministic and Stochastic Approaches. Frontiers in Genetics. 2.

Daetwyler HD, B Villanueva, and JA Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:E3395.

Daetwyler HD, S Bolormaa, DJ Brown, JHJ Van der Werf, BJ Hayes. 2013. A genomic prediction cross-validation approach combining ewe repeated phenotypes and ram daughter trait deviations. In: Proc. Assoc. Adv. Anim. Breed. Genet. 20: p. 360–363.

Dassonneville R, Baur A, Fritz S, Boichard D, Ducrocq V. 2012. Inclusion of cow records in genomic evaluations and impact on bias due to preferential treatment. Genetics Selection Evolution. 44:40.

Davis GH, Montgomery GW, Allison AJ, Kelly RW and Bray AR. 1982. Segregation of a major gene influencing fecundity in progeny of Booroola sheep. New Zealand Journal of Agriculture Research 25, 525–529.

Dekkers JCM and van Arendonk JAM. 1998. Optimizing selection for quantitative traits with information on an identified locus in outbred populations. Genetical Research, 71, 257-275.

Demars J, Fabre S, Sarry J, Rossetti R, Gilbert H, Persani L, Tosser-Klopp G, Mulsant P, Nowak Z, Drobik W, Martyniuk E and Bodin L. 2013. Genome-wide association studies identify two novel BMP15 mutations responsible for an atypical hyperprolificacy phenotype in sheep. PLoS Genetics 9:4

Dodds KG, ML Tate, and JA Sise. 2005. Genetic evaluation using parentage information from genetic markers. Journal of Animal Sciences. 83:2271–2279.

Douguet M, JM Astruc, and G Thomas. 2015. Résultats de contrôle laitier - France 2014. http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/resultats-de-controle-laitier-france-2013.html (Accessed 15 November 2015).

Douguet M, JM Astruc, G Thomas. Résultats de Contrôle Laitier France 2016. Compte rendu 0017 201 002. Département Génétique et Phénotypes. Institut de l'élevage, 2017.

Dréo J, A Pétrowski, P Siarry, E Taillard. Métaheuristiques pour l'optimisation difficile. Ed. Eyrolles, Paris, 2003.

Drouilhet L, Lecerf F, Bodin L, Fabre S and Mulsant P. 2009. Fine mapping of the FecL locus influencing prolificacy in Lacaune sheep. Animal Genetics, 40, 804–812.

Ducrocq V, and RL Quaas. 1988. Prediction of genetic response to truncation selection across generations. Journal of Dairy Sciences. 71:2543–2553.

Fritz S, F Guillaume, P Croiseau, A Baur, C Hoze, R Dassonneville, MY Boscher, L Journaux, D Boichard, and V Ducrocq. 2010. Mise en place de la sélection génomique dans les trois principales races françaises de bovins laitiers. Renc Rech Ruminants. 17:455–458.

Edel C, Pimentel ECG, Plieschke L, Emmerling R, Götz K-U. 2016. Short communication: The effect of genotyping cows to improve the reliability of genomic predictions for selection candidates. Journal of Dairy Sciences. 99:1999–2004.

Elsen JM, and J Mocquot. 1974. Méthode de prévision de l'évolution du niveau génétique d'une population soumise à une opération de sélection et dont les générations se chevauchent. INRA Bull. tech. Dépt. Génét. Anim 17:30–54.

Elsen JM. 1977. Pondérations économiques des indices de sélection dans le cas des populations de race mixte. Ann. Génét. Sél. Anim 9(I):73–85.

Elsen JM. 1992. De l'optimisation au progrès génétique réalisé dans les schémas de sélection. INRA Prod. Anim. 237–242.

Elsen JM. 1993. Prediction of annual genetic gain and improvement lag between populations. Genetics Selection Evolution. 25:75–82.

Falconer DS. Introduction to Quantitative Genetics. Oliver & Boyd, London; 1960.

Fatet A, B Leboeuf, S Freret, X Druart, L Bodin, H Caillat, I David, I Palhière, P Boué, G Lagriffoul. 2008. L'insémination dans les filières ovines et caprines. Renc. Rech. Ruminants, 15.

Fernando R, and M Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genetics Selection Evolution. 21:467–477.

Fontanesi L, G Schiavo, E Scotti, G Galimberti, DG Calò, AB Samorè, M Gallo, V Russo and L Buttazzoni. 2015. A retrospective analysis of allele frequency changes of major genes during 20 years of selection in the Italian Large White pig breed. Journal of Animal Breeding and Genetics, 132, 239–246.

Gao H, P Madsen, US Nielsen, GP Aamand, G Su, K Byskov, and J Jensen. 2015. Including different groups of genotyped females for genomic prediction in a Nordic Jersey population. Journal of Dairy Sciences. 98:9051–9059.

Garritsen C, E Van Marle-Köster, MA Snyman, and C Visser. 2015. The impact of DNA parentage verification on breeding value estimation and sire ranking in South African Angora goats. Small Ruminants Research. 124:30–37.

Geldermann H, U Pieper, and WE Weber. 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. Journal of animal science. 63:1759–1768.

Gill E, W Murray, and MH Wright. Practical optimization. Academic Press Inc, London, 1981.

Goddard M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica. 136:245–257.

Gomez-Raya L and G Klemetsdal. 1999. Two-stage selection strategies utilizing marker-quantitative trait locus information and individual performance. Journal of Animal Science, 77, 2008.

Gonzalez-Recio O, MP Coffey, JE Pryce. 2014 On the value of the phenotypes in the genomic era. Journal of Dairy Sciences. 97:7905–15.

Granleese T, SA Clark, AA Swan, JHJ van der Werf. 2015. Increased genetic gains in sheep, beef and dairy breeding programs from using female reproductive technologies combined with optimal contribution selection and genomic breeding values. Genetics Selection Evolution. 47:70.

Grasset D, J Bouix, B Bibé, H Leveziel, M Georges, and E Laville. 2009. Le gène d'hypertrophie musculaire du «Texel Belge»: identification, impact, introgression. Renc. Rech. Ruminants, 16.

Gengler N, S Abras, C Verkenne, S Vanderick, M Szydlowski, and R Renaville. 2008. Accuracy of Prediction of Gene Content in Large Animal Populations and its Use for Candidate Gene Detection and Genetic Evaluation. Journal of Dairy Sciences. 91:1652–1659.

Habier D, Fernando RL, Dekkers JCM. 2009. Genomic Selection Using Low-Density Marker Panels. Genetics. 182:343–53.

Habier D, RL Fernando, and DJ Garrick. 2013. Genomic BLUP Decoded: A Look into the Black Box of Genomic Prediction. Genetics. 194:597–607.

Hall SJG. 2016. Effective population sizes in cattle, sheep, horses, pigs and goats estimated from census and herdbook data. Animal. 10:1778–85.

Hamelin C. 2014. L'assignation de parenté chez les petits ruminants. Rapport de stage Licence professionnelle Génétique et développement de l'élevage, Université de Limoges, Institut de l'élevage, INRA.

Hanrahan JP, SM Gregan, P Mulsant, M Mullen, GH Davis, R Powell, SM Galloway. 2004. Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belclare sheep (Ovis aries). Biology of Reproduction. 70:900

Harder B, J Bennewitz, N Reinsch, M Mayer, and E Kalm. 2005. Effect of missing sire information on genetic evaluation. Arch. Tierz 48:219–232.

Hayes BEN, ME Goddard. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genetics Selection Evolution. 33:1.

Hayes BJ, PJ Bowman, AJ Chamberlain, ME Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Sciences. 92:433–43.

Hayes BJ, PJ Bowman, HD Daetwyler, JW Kijas, JHJ van der Werf. 2012. Accuracy of genotype imputation in sheep breeds. Animal Genetics. 43:72–80.

Hazel LN. 1943. The Genetic Basis for Constructing Selection Indexes. Genetics. 28(6): 476–490.

He S, S Wang, W Fu, X Ding, and Q Zhang. 2015. Imputation of missing genotypes from low- to high-density SNP panel in different population designs. Animal Genetics. 46:1–7.

Heaton MP, KA Leymaster, TS Kalbfleisch, JW Kijas, SM Clarke, J McEwan, JF Maddox, V Basnayake, DT Petrik, B Simpson, TPL Smith, CG Chitko-McKown, and the International Sheep Genomics Consortium. 2014. SNPs for parentage testing and traceability in globally diverse breeds of sheep. PLoS ONE 9:e94851.

Henderson CR. 1975. Best Linear Unbiased Estimation and Prediction under a Selection Model. Biometrics. 31:423-447.

Henryon M, T Ostersen, B Ask, AC Sørensen, and P Berg. 2015. Most of the long-term genetic gain from optimum-contribution selection can be realised with restrictions imposed during optimisation. Genetics Selection Evolution. 47.

Hill WG. 1971. Investment appraisal for national breeding programmes. Animal Production 13:37–50.

Hill WG. 1974. Prediction and evaluation of response to selection with overlapping generations. Animal Production 18:117–139.

Hozé C, MN Fouilloux, E Venot, F Guillaume, R Dassonneville, S Fritz, V Ducrocq, F Phocas, D Boichard and P Croiseau. 2013. High-density marker imputation accuracy in sixteen French cattle breeds. Genetics Selection Evolution. 45:33.

Huang Y, JM Hickey, MA Cleveland, and C Maltecca. 2012. Assessment of alternative genotyping strategies to maximize imputation accuracy at minimal cost. Genetics Selection Evolution. 44:25.

Hunter N, JD Foster, W Goldman, MJ Stear, J Hope, C Bostock. 1996. Natural scrapie in a closed flock of Cheviot sheep occurs only in specific PrP genotypes. Archives of Virology. 141 809-824.

Israel C, and JI Weller. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. Journal of Dairy Sciences. 83:181–187.

Kennedy BW, M Quinton, and JA Van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. Journal of Animal Science. 70:2000–2012.

Khang JVT. 1983. Méthodes d'analyse des données démographiques et généalogiques dans les populations d'animaux domestiques. Genetics Selection Evolution. 15:263–298.

Kijas JW, L Porto-Neto, S Dominik, A Reverter, R Bunch, R McCulloch, BJ Hayes, R Brauning, J McEwan and the International Sheep Genomics Consortium. 2014. Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal Genetics. 45:754–7.

Koivula M, I Strandén, GP Aamand, EA Mäntysaari. 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal. 10:1–6.

Lande R, and R Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics. 124:743–756.

Larzul C, E Manfredi and JM Elsen. 1997. Potential gain from including major gene information in breeding value estimation. Genetics Selection Evolution, 29, 161–184.

Lee SH, WMSP Weerasinghe, NR Wray, ME Goddard, JHJ van der Werf. 2017. Using information of relatives in genomic prediction to apply effective stratified medicine. Sciences Reports. 7:42091.

Legarra A, I Aguilar, and I Misztal. 2009. A relationship matrix including full pedigree and genomic information. Journal of Dairy Sciences. 92:4656–4663.

Legarra A, and ZG Vitezica. 2015. Genetic evaluation with major genes and polygenic inheritance when some animals are not genotyped using gene content multiple-trait BLUP. Genetics Selection Evolution. 47.

Le Roy P. 1989. Méthodes de détection de gènes majeurs. Applications aux animaux domestiques. Manuscrit de thèse, Université de Paris-sud, centre d'Orsay.

Le Roy P, Elsen JM. 2000. Principes de l'utilisation des marqueurs génétiques pour la détection des gènes influençant les caractères quantitatifs. INRA Productions Animales, INRA Editions, hors-série 211-215.

Long TE, RK Johnson, and JW Keele. 1990. Effects of errors in pedigree on three methods of estimating breeding value for litter size, backfat and average daily gain in swine. Journal of Animal Science. 68:4069–4078.

Loywyck V, G Lagriffoul. 2016. Compte rendu annuel sur l'insémination artificielle ovine – campagne 2015. Compte rendu 00 16 200 016. Département Génétique et Phénotypes. Institut de l'élevage.
Lynch M and B Walsh. 1998. Genetics and Analysis of Quantitative Traits. Sunderland, MA: Sinauer Associates, Inc.
Manfredi E, M Barbieri, F Fournet and JM Elsen. 1998. A dynamic deterministic model to evaluate breeding strategies under mixed inheritance. Genetics Selection Evolution 30, 127–148.
Martin P, J Raoul, and L Bodin. 2014. Effects of the FecL major gene in the Lacaune meat sheep population. Genetics Selection Evolution. 46:48.
McEwan JC. 1995. Industry use of the Inverdale gene (Fec X). In Proceeding of the New Zealand Society of Animal Production, 55, 298-300.
McHugh N, THE Meuwissen, AR Cromie, AK Sonesson. 2011. Use of female information in dairy cattle genomic breeding programs. Journal of Dairy Sciences. 94:4109–18.
Meuwissen TH. 1997. Maximizing the response of selection with a predefined rate of inbreeding. Journal of animal science. 75:934–940.
Meuwissen THE, BJ Hayes, and ME Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819–1829.
Meuwissen T, B Hayes, and M Goddard. 2013. Accelerating improvement of livestock with genomic selection. Annual Review of Animal Biosciences. 1:221–237.
Misztal I. 1999. Complex models, more data: simpler programming. Interbull Bulletin. 20:33-42.
Misztal I, SE Aggrey, WM Muir. 2013. Experiences with a single-step genome evaluation. Poultry Sciences. 92:2530–4.
Moghaddar N, AA Swan, JHJ Van der Werf. 2013. Accuracy of genomic prediction from multi-breed sheep reference population. In: Proc. Assoc. Adv. Anim. Breed. Genet. 20: p. 220–223.
Moghaddar N, AA Swan, and JH van der Werf. 2014. Comparing genomic prediction accuracy from purebred, crossbred and combined purebred and crossbred reference populations in sheep. Genetics Selection Evolution. 46:58.
Moghaddar N, KP Gore, HD Daetwyler, BJ Hayes, JHJ van der Werf. 2015. Accuracy of genotype imputation based on random and selected reference sets in purebred and crossbred sheep populations and its effect on accuracy of genomic prediction. Genetics Selection Evolution. 47:97.
Nicolas C. 2014. L'assignation de parenté chez les petits ruminants. Rapport de stage Licence professionnelle Productions animales, Université du Maine Institut de l'élevage, INRA.
Ollivier L. Eléments de génétique quantitative. 2ème édition revue et augmentée. INRA, 2002.
OVI-TEST, 2016. Official website. Retrieved on 23 September 2016 from http://www.lacaune-ovitest.com/fr/lacaune-viande-ovi-test.php.
Palhière I, JM Elsen, JM Astruc, F Barillet, B Bed'Hom, B Bibé, B Bouix, M Brochard, O Catrou, F Dion, D François, L Griffon, E Jullien, M Orlianges, G Perret and P Tribon. 2003. Breeding for scrapie resistance in France. In Proceeding of the International Workshop on Major Genes and QTL in Sheep and Goat, Toulouse, France. CD-ROM communication no. 3-03.
Palhière I, M Brochard, JM Astruc, F Barillet, B Bed'Hom, B Bibé, J Bouix, M Brochard, O Catrou, JM Elsen, D François, L Griffon, E Jullien, M Orlianges, G Perret, P Tribon. 2004. Breeding for scrapie resistance in France. In: EAAP annual meeting. Bled, Slovenia.
Palhière I, V Clément, I Croué, D Pinard. 2015. Un objectif de sélection qui augmente le profit des éleveurs caprins. Renc. Rech. Ruminants, 22.
Parlato E, and LD Van Vleck. 2012. Effect of parentage misidentification on estimates of genetic parameters for milk yield in the Mediterranean Italian buffalo population. Journal of Dairy Sciences. 95:4059–4064.
Pickering NK, KG Dodds, B Auvray, JC McEwan. 2013. The impact of genomic selection on genetic gain in the New Zealand sheep dual purpose selection index. In: Proc. Assoc. Adv. Anim. Breed. Genet. 20: p. 175–178.
Piper LR and BM Bindon. 1982. The Booroola Merino and the performance of medium non-Peppin crosses at Armidale. In the Booroola Merino, 9-20. CSIRO, Melbourne.

Plieschke L, C Edel, ECG Pimentel, R Emmerling, J Bennewitz, and KU Götz. 2016. Systematic genotyping of groups of cows to improve genomic estimated breeding values of selection candidates. Genetics Selection Evolution. 48.

Poutous M, B Vissac, M Calomiti, and M Chambeyron. 1962. Recherche théorique des conditions de rentabilité maximum de l'épreuve de descendance des taureaux d'insémination artificielle. In: Annales de zootechnie. Vol. 11. p. 233–256.

Pryce JE, HD Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Animal Production Sciences. 52:107.

Pryce JE, BJ Hayes, ME Goddard. 2012. Genotyping dairy females can improve the reliability of genomic selection for young bulls and heifers and provide farmers with new management tools. Proc. of the 38th ICAR Session ; 28.

de Rancourt M, and J Raoul. 2013. An international comparison of the main meat sheep genetic schemes. In: Proc. 19th International Farm Management Congress, SGGW, Warsaw, Poland.

Raoul J. 2011. Contrôles de filiations ovins allaitants 2011. Bulletin Geovial 11. Institut de l'élevage.

Raoul J, C Chantry-Darmon, L Barbotte, JLM Babilliot, MY Bosher, L Bodin. 2012. Parentage assignment with molecular markers in sheep: first results of an experiment and prospects. In: Proc. 63th Annual Meeting of the EAAP, Bratislava, Poland.

Raoul J, Palhière I, Astruc J. and JM Elsen. 2016. Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs. Journal of Animal Sciences. 94(9):3663-3683.

Raoul J, I Palhière, J Astruc, AA Swan, and JM Elsen. 2017a. Optimal mating strategies to manage a heterozygous advantage major gene in sheep. Animal, 1-10.

Raoul J, L Bodin, JM Elsen, and AA Swan. 2017b. Optimal and practical strategies to manage an ovulation rate mutation located on the X chromosome in a French sheep breed. In: Proc. Assoc. Adv. Anim. Breed. Genet. 22.

Raoul J, AA Swan, and JM Elsen. 2017c. Using a very low density SNP panel for genomic selection in a breeding program for sheep. Genetics Selection Evolution. 49:76.

Raoul J, AA Swan, and JM Elsen. 2017d. Assessment of a genomic design for French meat sheep breeding program. WCGALP 2018, Auckland, New-zealand.

Rendel JM, and A Robertson. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. Journal of Genetics. 50:1–8.

Robert-Granié C, A Legarra, and V Ducrocq. 2011. Principes de base de la sélection génomique. Productions Animales. 24:331.

Ron M, Y Blanc, M Band, E Ezra, and JI Weller. 1996. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. Journal of Dairy Sciences. 79:676–681.

Roughsedge T, S Brotherstone, and PM. Visscher. 2001. Bias and power in the estimation of a maternal family variance component in the presence of incomplete and incorrect pedigree information. Journal of Dairy Sciences. 84:944–950.

Rupp R, S Mucha, H Larroque, J McEwan, J Conington. 2016. Genomic application in sheep and goat breeding. Animal Frontiers. Vol. 6, No. 1.

Sánchez A, H Ilahi, E Manfredi and JM Serradilla. 2005. Potential benefit from using the αs1-casein genotype information in a selection scheme for dairy goats. Journal of Animal Breeding and Genetics, 122, 21–29.

Sanders K, J Bennewitz, and E Kalm. 2006. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. Journal of Dairy Sciences. 89:315–321.

Santos BFS, JHJ van der Werf, JP Gibson, TJ Byrne, PR Amer. 2017. Genetic and economic benefits of selection based on performance recording and genotyping in lower tiers of multi-tiered sheep breeding schemes. Genetics Selection Evolution. 49:10.

Sargolzaei M, JP Chesnais, FS Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478.

Schaeffer LR. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics. 123:218–223.

Schenkel FS, and LR Schaeffer. 2000. Effects of nonrandom parental selection on estimation of variance components. Journal of Animal Breeding and Genetics. 117:225–239.

Schenkel FS, LR Schaeffer, and PJ Boettcher. 2002. Comparison between estimation of breeding values and fixed effects using Bayesian and empirical BLUP estimation under selection on parents and missing pedigree information. Genetics Selection Evolution. 34:41–59.

Senneke, SL, MD MacNeil, and LD Van Vleck. 2004. Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. Journal of Animal Sciences. 82:2307–2312.

Shumbusho F, J Raoul, JM Astruc, I Palhière, and JM Elsen. 2013. Designing, technical evaluation and profitability estimation of breeding strategies based on molecular information for small ruminant species. Journal of Animal Sciences. 91:3644–3657.

Shumbusho F, J Raoul, JM Astruc, I Palhiere, S Lemarié, A Fugeray-Scarbel, and JM Elsen. 2016. Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. Animal.10:1033–41.

Smith C, and AJ Webb. 1981. Effects of major genes on animal breeding strategies1. Zeitschrift für Tierzüchtung und Züchtungsbiologie. 98:161–169.

Sodhi SS, DK Jeong, N Sharma, JH Lee, JH Kim, SH Kim, SW Kim and SJ Oh. 2013. Marker Assisted Selection-Applications and Evaluation for Commercial Poultry Breeding. Korean Journal of Poultry Science, 40, 223–234.

Solberg TR, B Heringstad, M Svendsen, H Grove, and TH Meuwissen. 2011. Genomic predictions for production-and functional traits in Norwegian red from BLUP analyses of imputed 54K and 777K SNP data. Interbull Bulletin.

Soller M, Genizi A, 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics, 34, 7-55.

Su G, P Ma, US Nielsen, GP Aamand, G Wiggans, B Guldbrandtsen, and MS Lund. 2016. Sharing reference data and including cows in the reference population improve genomic predictions in Danish Jersey. Animal. 10:1067–1075.

Sullivan PG. 1995. Alternatives for genetic evaluation with uncertain parentage. Canadian Journal of Animal Sciences. 75:31–36.

Sun X, R Fernando, J Dekkers.2016. Contributions of linkage disequilibrium and co-segregation information to the accuracy of genomic prediction. Genetics Selection Evolution. 48:77.

Swan AA, DJ Brown, HD Daetwyler, BJ Hayes, M Kelly, N Moghaddar, and JHJ. Van der Werf. 2014. Genomic evaluations in the Australian sheep industry. In: Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada.

Talenti A, I Palhière, F Tortereau, G Pagnacco, A Stella, EL Nicolazzi, P Crepaldi, G Tosser-Lopp and ADAPTmap Consortium. 2017. Functional SNP panel for parentage assessment and assignment in worldwide goat breeds. Submitted to Genetics Selection Evolution.

Thomasen JR, AC Sørensen, MS Lund, B Guldbrandtsen. 2014. Adding cows to the reference population makes a small dairy population competitive. Journal of Dairy Sciences. 97:5822–32.

Tiphine L, I David, J Raoul, J Guerrier, JP Praud, L Bodin, D François, E Jullien JP Poivey. 2013. Estimation of breeding values for meat sheep in France. In: EAAP annual meeting. Stavanger, Norway. Book of abstracts No. 17, p. 344.

Tiphine L. 2017. Bilan du contrôle de performances ovins allaitants. Campagne 2016. Compte rendu 0017 201 004. Département Génétique et Phénotypes. Institut de l'élevage.

Toghiani S, SE Aggrey, and R Rekaya. 2016. Multi-generational imputation of single nucleotide polymorphism marker genotypes and accuracy of genomic selection. Animal. 10:1077–85.

Tortereau F, C Moreno, G Tosser-Klopp, L Barbotte, L Genestout, and J Raoul. 2015. Development of a SNP parentage assignment panel for French sheep breeds. In: EAAP annual meeting. Warsaw, Poland. Book of abstracts No. 21, p. 510.

Tortereau F, C Moreno, G Tosser-Klopp, B Servin, J Raoul. 2017. Development of a SNP panel dedicated to parentage assignment in French sheep populations. BMC Genetics. 18(1):50.

Tribout T, C Larzul, and F Phocas. 2012. Efficiency of genomic selection in a purebred pig male line1. Journal of Animal Sciences. 90:4164-4176.

Tribout T. 2013a. Intérêt de la sélection génomique dans les programmes de sélection porcins : cas d'une lignée mâle de grande taille. Manuscrit de thèse. Institut des Sciences et Industries du Vivant et de l'Environnement (AgroParisTech).

Tribout T, C Larzul, and F Phocas. 2013b. Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. Genetics Selection Evolution. 45:40.

Van Eenennaam AL, RL Weaber, DJ Drake, MCT Penedo, RL Quaas, DJ Garrick, and EJ Pollak. 2007. DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting. Journal of Animal Sciences. 85:3159–3169.

Van Raden PM. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Sciences. 91:4414–4423.

VanRaden, PM, ME. Tooker, JR O'Connell, JB Cole, and DM. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. Genetics Selection Evolution. 49:32.

Van der Werf JHJ. 2014. Genomic selection in sheep breeding programs. In: Proc. 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada.

Van der Werf J, R Banks. 2017. Effective breeding programmes down under. Journal of Animal Breeding and Genetics. 134:1–2.

Ventura R, M Lee, S Miller, S Clarke, JC McEwan. 2015. Assessing imputation accuracy using a 15K low density panel in a multi-breed New Zealand sheep population. Proc Assoc Advmt Anim Breed Genet. 21: 302-5.

Villanueva B, R Pong-Wong, B Grundy and JA Woolliams. 1999. Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genetics Selection Evolution, 31, 115–133.

de Vries AG, A Sosnicki, JP Garnier and GS Plastow. 1998. The role of major genes and DNA technology in selection for meat quality in pigs. Meat Science, 49, 245–255.

Wang C, D Habier, BL Peiris, A Wolc, A Kranis, KA Watson, S Avendano, DJ Garrick, RL Fernando, SJ Lamont, and JCM Dekkers. 2013. Accuracy of genomic prediction using an evenly spaced, low-density single nucleotide polymorphism panel in broiler chickens. Poultry Science. 92:1712–1723.

Weller JI. 1994. Economic aspects of animal breeding. Published by Chapman & Hall, London, UK.

Wellmann, R., S Preuß, E Tholen, J Heinkel, K Wimmers, and J Bennewitz. 2013. Genomic selection using low density marker panels with application to a sire line in pigs. Genetics Selection Evolution. 45:28.

Winkelman AM. 2013. Effect of daughter misidentification on dairy sire evaluation. In: Proceedings of the Twentieth Conference of the Association for the Advancement of Animal Breeding and Genetics, Translating Science into Action, Napier, New Zealand, 20th-23rd October 2013. In: Proc. Assoc. Adv. Anim. Breed. Genet. p. 25–28.

Woolliams JA, P Bijma, and B Villanueva. 1999. Expected genetic contributions and their impact on gene flow and genetic gain. Genetics 153:1009–1020.

Xiang T, P Ma, T Ostersen, A Legarra, OF Christensen. 2015. Imputation of genotypes in Danish purebred and two-way crossbred pigs using low-density panels. Genetics Selection Evolution. 47:54.

Zhang Z, and T Druet. 2010. Marker imputation with low-density marker panels in Dutch Holstein cattle. Journal of Dairy Sciences. 93:5487–5494.

Liste des publications

Publications principales

Raoul J, Palhière I, Astruc J. and JM Elsen. 2016. Genetic and economic effects of the increase in female paternal filiations by parentage assignment in sheep and goat breeding programs. Journal of Animal Sciences. 94(9):3663-3683.

Raoul J, I Palhière, J Astruc, AA Swan, and JM Elsen. 2017a. Optimal mating strategies to manage a heterozygous advantage major gene in sheep. Animal, 1-10.

Raoul J, AA Swan, and JM Elsen. 2017c. Using a very low density SNP panel for genomic selection in a breeding program for sheep. Genetics Selection Evolution. 49:76.

Publications courtes

Raoul J, L Bodin, JM Elsen, and AA Swan. 2017b. Optimal and practical strategies to manage an ovulation rate mutation located on the X chromosome in a French sheep breed. In: Proc. Assoc. Adv. Anim. Breed. Genet. 22.

Raoul J, AA Swan, and JM Elsen. 2017d. Assessment of a genomic design for French meat sheep breeding program. WCGALP 2018, Auckland, New-Zealand. Accepted for oral presentation or poster.

Communications orales

En congrès international

Raoul, J., Palhiere, I., Astruc, J.-M., Elsen, J. M. (2015). Is annual genetic gain influenced by the rate of registered paternity in sheep and goat breeding programs ? Presented at 66. Annual Meeting of the European Federation of Animal Science (EAAP), Varsovie, Poland (2015-08-31 - 2015-09-04).

Raoul, J., Bodin, L., Elsen, J. M., Swan, A. A. (2017). Optimal and practical strategies to manage an ovulation rate mutation located on the X chromosome in a French sheep breed. Presented at 22. Conference Association Advancement Animal Breeding Genetics (AAABG), Townsville, Australia (2017-07-02 - 2017-07-05).

En séminaire

Raoul, J., Palhiere, I., Astruc, J.-M., Elsen, J. M. (2016). Genetic and economic interests of a Very Low Density SNPs chip in sheep and goat breeding programs. Présenté lors du Séminaire des thésards du département de génétique animale, Toulouse, France (2016-03-16 – 2016-03-17) et du séminaire des thésards du métaprogramme INRA SELGEN, Paris, France (2016-05-26).

Raoul, J., Elsen, J. M. (2016). Genetic and economic interests of a Very Low Density SNPs chip in sheep and goat breeding programs. Présenté lors du Séminaire R2D2, action du métaprogramme INRA SELGEN, Autrans, France (2016-05-18 – 2016-05-20).

Raoul J, I Palhière, J Astruc, AA Swan, and JM Elsen. (2016) Optimal mating plans to manage a major gene for ovulation rate in the Lacaune meat sheep breed. Presented at the Animal and Genetics Breeding Unit Seminar, Armidale, Australia (2016-10-05).

Raoul J, AA Swan, and JM Elsen. (2017) Using a very low density SNP panel for genomic selection in a breeding program for sheep. Presented at the Animal and Genetics Breeding Unit Seminar, Armidale, Australia (2017-06-19).