



HAL
open science

Pedigree-based methods to partition genetic gain and inbreeding load in dairy sheep

Simona Antonios

► **To cite this version:**

Simona Antonios. Pedigree-based methods to partition genetic gain and inbreeding load in dairy sheep. Animal biology. Université de Toulouse, 2024. English. NNT : 2024TLSEP099 . tel-04801619

HAL Id: tel-04801619

<https://theses.hal.science/tel-04801619v1>

Submitted on 25 Nov 2024

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

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

Doctorat de l'Université de Toulouse

préparé à Toulouse INP

Méthodes basées sur la généalogie pour partitionner le gain
génétique et le fardeau génétique chez les ovins laitiers

Thèse présentée et soutenue, le 15 novembre 2024 par

Simona ANTONIOS

École doctorale

SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries

Spécialité

Infectiologie, Physiopathologie, Toxicologie, Génétique et Nutrition

Unité de recherche

GenPhySE - Unité Génétique, Physiologie et Systèmes d'Elevage

Thèse dirigée par

Zulma VITEZICA et Silvia RODRIGUEZ RAMILO

Composition du jury

M. Joaquim CASELLAS, Président et Rapporteur, Universitat Autònoma de Barcelona

M. Ino ČURIK, Rapporteur, University of Zagreb

Mme Pauline MARTIN, Examinatrice, INRAE Île-de-France - Jouy-en-Josas - Antony

M. Jérôme RAOUL, Examineur, INRAE Occitanie-Toulouse

Mme Zulma VITEZICA, Directrice de thèse, Toulouse INP

Mme Silvia RODRÍGUEZ-RAMILO, Co-directrice de thèse, INRAE Occitanie-Toulouse

This work was funded by project ARDI (grant agreement EFA 208/16), and project ARDI2 (grant agreement EFA 032/01) from POCTEFA funds. This work also received funding from the Department of Animal Genetics of INRAE and Toulouse INP, France.



Interreg
POCTEFA



Cofinanciado por
la UNIÓN EUROPEA
Cofinancé par
l'UNION EUROPÉENNE



This work is dedicated to my parents,
with heartfelt gratitude for their boundless love,
unwavering support, and constant inspiration.

هذا العمل مُهدى إلى والدي،
تقديرًا لحبهم اللامحدود
ودعمهم المستمر وإلهامهم الدائم.

Acknowledgements

The completion of this thesis marks the culmination of three transformative years, during which I have been fortunate to receive support, guidance, and friendship from numerous individuals. Each of you, in your own unique way, has contributed to this journey, and for that, I am deeply grateful. Though it is impossible to individually acknowledge every person within the confines of this page, please know that each of you holds a special place in my heart, forever etched into the narrative of this accomplishment. I would like to thank a couple of people who have been involved closely in my daily work and in finalizing this thesis.

First and foremost, I wish to express my profound gratitude to my supervisor, Zulma. Your unwavering support, patience, and insightful guidance have been pivotal throughout my Ph.D. journey. You not only provided me with the opportunity to build upon the foundations laid during my master's thesis but also enriched my experience with your kindness and encouragement. The moments we shared, both professional and personal, including the delightful meals and conversations, have left an indelible mark on me. Your exemplary dedication to your work and your compassion for those around you continue to inspire me deeply. For all this and more, I am eternally grateful.

Silvia, as my co-supervisor and the first Ph.D. student under your mentorship, I am honoured to have been guided by you. Your advice, both in work and in life, your sense of humour, and your unwavering support have been invaluable. The time spent with you, Alba, and Oscar has added immeasurable joy to this experience. Thank you for making this journey memorable.

I also wish to extend my heartfelt thanks to Andres Legarra for your practical wisdom, advice, and collaboration. Your support has greatly enriched my Ph.D. work.

During my research, I was privileged to spend five months at the Roslin Institute in Edinburgh, UK. This experience was made unforgettable by the warmth and friendship of many people. Special thanks go to Gregor Gorjanc and Ivan Pocrnic for having received me; you have helped me (re)discover the fun in doing science. Gregor, your enthusiasm for science rekindled my own and your encouragement and advice have been invaluable. Ivan, thank you

for opening doors to new opportunities through the British Council, which led to further growth and the promise of future collaborations.

My journey was also enriched by so many great experiences with my colleagues and fellow students. To all those with whom I have shared discussions, papers, meals, office space, conferences, and countless memories, I extend my deepest thanks. You made my time at INRAE both enjoyable and fulfilling. A special mention to Marine, for sharing both the office and life's adventures, and to Coralie, for making me feel at home and for our unforgettable moments together.

I would like to acknowledge Professor Rabih Roumouz at the Lebanese University, whose encouragement and support in the student exchange program helped me discover and pursue my true passion.

To those who were not directly involved in my academic work but whose love and support were my bedrock: My parents, your belief in me and your constant encouragement have shaped the person I am today. Mother, your words, "fly high and follow your dreams, no matter how distant or difficult they seem," have guided me through every challenge. Your endless love, inspiration, and unwavering support gave me the strength to persevere. I could not have achieved this without you. I love you dearly. To my dearest Martin, thank you for your unceasing support, for listening to my worries and thoughts, especially during the final stages of my thesis. Your visits, love, and encouragement have been a source of immense comfort and strength.

I extend my sincere thanks to the members of my thesis committee—A. Legarra, J.M. Astruc, L. Varona, L. Sanchez-Rodriguez, and G. Restoux—for their invaluable contributions to my work. I am also deeply grateful to the thesis jury members—I. Curik, J. Casellas, P. Martin, and J. Raoul—for dedicating their time to review my research and provide insightful feedback.

Lastly, this thesis would not have been possible without the financial support from several institutions. I gratefully acknowledge the projects ARDI (grant agreement EFA 208/16), and ARDI2 (grant agreement EFA 032/01) from POCTEFA funds. I also extend my thanks to the Department of Animal Genetics of INRAE (Toulouse, France) and Toulouse INP, France, for funding my five-months stay in Edinburgh. The Genotoul Bioinformatics platform

Toulouse Occitanie (Bioinfo Genotoul) for providing the essential computing and storage resources.

To each and every one of you, thank you for being a part of this journey. Your contributions, in ways both large and small, have made this achievement possible.

Résumé

Cette thèse explore des méthodes basées sur la généalogie pour partitionner le gain génétique et le fardeau génétique (FG) dans les races ovines laitières françaises : Lacaune (LAC), Basco-Béarnaise (BB), Manech Tête Noire (MTN) et Manech Tête Rousse (MTR).

Le Chapitre 2 a utilisé une analyse rétrospective pour affiner la partition de la tendance génétique dans les échantillonnages mendéliens par catégorie d'animaux définies par le sexe et par la voie de sélection, ainsi que pour caractériser les contributions génétiques à long terme. Nous avons analysé le gain génétique pour la production laitière dans quatre races : LAC, BB, MTN et MTR. Les mères à béliers (MAB) et les mâles d'insémination artificielle (IA) ont été les sources les plus importantes de progrès génétique, comme l'a montré la décomposition des tendances de l'échantillonnage mendélien. Les contributions annuelles étaient plus variables pour les mâles d'IA que pour les MAB, étant donné que ces contributions ont été calculées en moyenne sur un plus petit nombre d'individus. En termes d'échantillonnage mendélien, les femelles ont contribué davantage que les mâles au gain génétique total, et nous interprétons cela comme étant dû au fait que les femelles constituent un plus grand réservoir de diversité génétique. En outre, nous avons calculé les contributions à long terme de chaque individu aux pseudo-génération suivantes. L'échantillonnage mendélien était plus important que la moyenne des parents pour déterminer la sélection des individus et leurs contributions à long terme. Ces contributions étaient plus significatives pour les mâles d'IA (dont la descendance est plus importante que celle des femelles) et en BB qu'en LAC (étant une race de taille plus importante).

Au Chapitre 3, la théorie qui montre la nature additive du FG est présentée. L'effet du FG et l'effet génétique additif (dans une population non consanguine) ont une corrélation négative dépendant de la fréquence des allèles, de la consanguinité et de la dominance. Nous avons calculé et décrit les coefficients de consanguinité partielle dans trois races : BB, MTN et MTR. Ensuite, nous avons inclus ces coefficients dans un modèle mixte en tant que covariables de régression aléatoire pour estimer la variance et les valeurs génétiques du FG pour la production laitière. Il existe une variance génétique pour le FG dans les races MTN et MTR, mais elle n'était pas différente de zéro pour BB. Comme attendu, nous avons estimé des corrélations génétiques négatives entre le FG et les valeurs génétiques estimées ; cependant,

elles étaient proches de zéro dans les trois races. La faible magnitude du FG ne justifie pas une sélection fondée sur ce critère.

Dans le Chapitre 4, nous avons évalué l'efficacité de l'intégration du FG dans les stratégies de sélection chez les ovins laitiers. Nous avons simulé 10 générations de sélection. Six scénarios qui diffèrent par les critères de sélection (uniquement les valeurs génétiques additive estimées du caractère, uniquement les valeurs génétiques estimées du FG, ou à la fois les deux) et les stratégies d'accouplement (minimiser le FG ou la consanguinité attendue dans la descendance) ont été évalués. Les scénarios ont été comparés en termes de gain génétique, coefficients et taux de consanguinité, taille efficace et précision de la sélection. Il est possible d'utiliser les prédictions des effets du FG pour sélectionner les animaux directement ou dans le cadre de stratégies d'accouplement. Cependant, la sélection basée sur le FG (en raison de sa variation et de sa magnitude) ne présente pas d'intérêt pratique. À la lumière de nos résultats, l'inclusion d'animaux génotypés pourrait améliorer la précision de la prédiction des FG individuelles. D'autres recherches sont nécessaires.

Summary

This thesis explores pedigree-based methods to partition genetic gain and inbreeding load in French dairy sheep breeds: Lacaune (LAC), Basco-Béarnaise (BB), Manech Tête Noire (MTN) and Manech Tête Rousse (MTR).

The Chapter 2 used a retrospective analysis to fine partitioning genetic trend in Mendelian samplings by categories of animals defined by sex and by selection pathways, and to similarly characterize long-term genetic contributions. We analysed genetic gain for milk yield in four dairy sheep breeds: LAC, BB, MTN and MTR. Dams of males and Artificial Insemination (AI) males were the most important sources of genetic progress as observed in the decomposition in Mendelian sampling trends. The yearly contributions were more erratic for AI males than for dams of males as they are averaged across a smaller number of individuals. Overall, in terms of Mendelian sampling, females contributed more than males to the total genetic gain, and we interpret that this is because females constitute a larger pool of genetic diversity. In addition, we computed long-term contributions from each individual to the following pseudo-generations. Mendelian sampling was more important than Parent Average to determine the selection of individuals and their long-term contributions. Long-term contributions were larger for AI males (with larger progeny sizes than females) and in BB than in LAC (with the latter being a larger population).

In Chapter 3, we presented theory that show the additive nature of the inbreeding load. The inbreeding load effect and the regular (in non-inbred population) additive genetic effect have a negative correlation depending on allele frequencies, inbreeding and dominance. We calculated and described the partial inbreeding coefficients in three French dairy sheep populations: BB, MTN and MTR. Then, we included these coefficients in a mixed model as random regression covariates, to predict genetic variance and breeding values of the inbreeding load for milk yield in the same breeds. There is genetic variance for inbreeding load in MTN and MTR breeds, but it was not different from zero for BB. As expected, we estimated negative genetic correlations between inbreeding load and breeding values; however, estimates were close to zero in the three sheep breeds. The small magnitude of inbreeding load does not warrant selection based on this criterion.

In Chapter 4, we evaluated the effectiveness of involving inbreeding load in selection strategies in a dairy sheep breeding scheme. We did this by simulation of 10 generations of

evaluations and selection. Six scenarios that differ in the criteria of selection (only breeding values, only breeding values of inbreeding load, or both genetic and inbreeding load breeding values) and mate allocation strategies (minimising inbreeding load or minimising expected future inbreeding) were evaluated. Scenarios were compared in terms of genetic gain, inbreeding coefficients, rate of inbreeding, effective population size, and accuracy of selection. The use of predictions of inbreeding load effects to select animals directly or in mating strategies is feasible. However, selection based on inbreeding load (due to its variation and magnitude) is not of practical interest. In light of our results, the inclusion of genotype animals could improve the accuracy of predicting individual inbreeding loads. Further research is needed.

List of figures

Figure 1.1. Pedigree of 7 animals.	4
Figure 1.2. Pedigree of animal 7	10
Figure 1.3. Partial inbreeding coefficient of animal 7 from founder 1	10
Figure 1.4. Partial inbreeding coefficient of animal 7 from founder 2	11
Figure 1.5. Partial inbreeding coefficient of animal 7 from animal 4.....	11
Figure 1.6. Distribution of the 5 French dairy sheep breeds in France (Races de France, 2011).	13
Figure 1.7. Evolution of the sheep milk production in France (Lagriffoul, personal communication).	13
Figure 1.8. The selection scheme of French dairy sheep breeds (Barillet, 1997).....	14

List of tables

Table 1.1. Pedigree of 7 animals.	4
Table 1.2. Number of individuals in the selection scheme in 2014 before genomic selection (Astruc <i>et al.</i> , 2016).	15
Table 1.3. Genomic breeding schemes in French dairy sheep in 2021 (Astruc <i>et al.</i> , 2022)..	16

Table of contents

Chapter 1: Introduction.....	1
1.1 Mendelian sampling.....	3
1.2 Genetic contributions.....	7
1.3 Decomposition of genetic trend.....	8
1.4 Decomposition of inbreeding.....	9
1.5 Inbreeding load.....	12
1.6 Dairy sheep in France.....	12
1.6.1 Conventional breeding schemes.....	15
1.6.2 Genomic selection.....	16
1.7 Objectives.....	16
1.8 Outlines.....	16
Chapter 2: Partition of the genetic trend of French dairy sheep in Mendelian samplings and long-term contributions.....	19
Chapter 3: Genetic inbreeding load and its prediction in French dairy sheep.....	37
Chapter 4: Exploring options to select based on inbreeding load via simulation.....	59
Chapter 5: Genral discussion and perspectives.....	83
5.1 Partition of the genetic trends.....	85
5.1.1 Selection pathways.....	85
5.1.2 Partitioning of the genetic trend.....	86
5.1.3 Link between partition of genetic trend and the evolution of inbreeding.....	89
5.2 Approaches to control inbreeding.....	91
5.2.1 Optimum contribution selection.....	91
5.2.2 Expected future inbreeding.....	91
5.2.3 Inbreeding load.....	92
5.3 General conclusions.....	94
References.....	97

Training, education, teaching and awards during the PhD thesis.....	103
Résumé/Abstract.....	106

Chapter 1

Introduction

1.1 Mendelian sampling

The breeding value of an animal can be expressed as the average of its parental breeding values plus a Mendelian sampling term. The recombination and segregation of parental chromosomes during meiotic division creates new variation. Mendelian sampling is the unique portion of genetic variation that an individual can bring to the population (Avendaño *et al.*, 2004; Woolliams, 2007). Mendelian sampling captures the originality of the individual with respect to its parents.

The breeding value for an animal x (u_x) is the average of the breeding values for its sire and dam (u_s and u_d), known as parental average, plus its Mendelian sampling (ϕ_x), it is as follows:

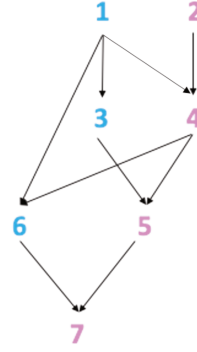
$$u_x = \frac{1}{2}u_s + \frac{1}{2}u_d + \phi_x \quad (1.1)$$

In matrix notation, the vector of breeding values is $\mathbf{u} \sim \mathcal{N}(0, \mathbf{A}\sigma_u^2)$ where σ_u^2 is the additive genetic variance and \mathbf{A} is the additive genetic relationship matrix. The matrix \mathbf{A} can be written as $\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}'$, where \mathbf{T} describes the flow of genes through the pedigree (Woolliams *et al.*, 1999), and explains the relatedness between each individual and its ancestors (García-Cortés *et al.*, 2008). \mathbf{T} draws the flow of genes from one generation to another, thereafter accounts for the direct relationships for parent-offspring. \mathbf{T} is a lower triangular matrix with one along the diagonal and all the non-zero element (to the left of the diagonal). The element of \mathbf{T} , say $t_{xx'}$, is the coefficient of relationship between animals x and x' , and is computed as follows: (1) if both x 's parents (s and d) are known, $t_{xx'} = \frac{1}{2}(t_{sx'} + t_{dx'})$; (2) if one x 's parent (s) is known, then $t_{xx'} = \frac{1}{2}(t_{sx'})$; and (3) if none of x 's parents are known, then $t_{xx'} = 0$ (Mrode, 2014).

The diagonal matrix \mathbf{D} contains Mendelian sampling variances for non-founders and genetic variances for founders; d_{xx} is the element of \mathbf{D} for each individual x . To explain more in details the matrices \mathbf{D} and \mathbf{T} , we use a small pedigree of 7 individuals (Table 1.1 and Figure 1.1).

Table 1.1. Pedigree of 7 animals.

id	sire	dam
1	0	0
2	0	0
3	1	0
4	1	2
5	3	4
6	1	4
7	6	5

**Figure 1.1.** Pedigree of 7 animals.

Using equation (1.1), we can write ϕ_x for an animal x as:

$$\phi_x = u_x - \left(\frac{1}{2}u_s + \frac{1}{2}u_d\right) \quad (1.2)$$

and $\text{var}(\phi_x) = \text{var}(u_x) - \text{var}\left(\frac{1}{2}u_s + \frac{1}{2}u_d\right)$. Being the diagonal element of \mathbf{A} for animal x equal to $a_{xx} = 1 + F_x$ where F_x is the inbreeding coefficient of animal x , we have $\text{var}(u_x) = a_{xx}\sigma_u^2 = (1 + F_x)\sigma_u^2$. Then, we can write

$$\begin{aligned} \text{var}(\phi_x) &= \text{var}(u_x) - \text{var}\left(\frac{1}{2}u_s + \frac{1}{2}u_d\right) = (1 + F_x)\sigma_u^2 - \left(\text{var}\left(\frac{1}{2}u_s\right) + \text{var}\left(\frac{1}{2}u_d\right) + \right. \\ & 2\text{cov}\left(\frac{1}{2}u_s, \frac{1}{2}u_d\right)\left.) = (1 + F_x)\sigma_u^2 - \frac{1}{4}a_{ss}\sigma_u^2 - \frac{1}{4}a_{dd}\sigma_u^2 - \frac{1}{2}a_{sd}\sigma_u^2 = (1 + F_x)\sigma_u^2 - \frac{1}{4}(1 + \right. \\ & \left. F_s)\sigma_u^2 - \frac{1}{4}(1 + F_d)\sigma_u^2 - \frac{1}{2}a_{sd}\sigma_u^2, \end{aligned}$$

where F_s and F_d are the inbreeding coefficients of animal x 's sire and dam, respectively. Then

$$\text{var}(\phi_x) = d_{xx}\sigma_u^2 = \left[(1 + F_x) - \frac{1}{4}(1 + F_s) - \frac{1}{4}(1 + F_d) - \frac{1}{2}a_{sd} \right] \sigma_u^2,$$

As $F_x = \frac{1}{2}a_{sd}$, then

$$\text{var}(\phi_x) = d_{xx}\sigma_u^2 = \left[1 - \frac{1}{4}(1 + F_s) - \frac{1}{4}(1 + F_d) \right] \sigma_u^2.$$

If both parents (s and d) are known:

$$d_{xx} = 1 - \frac{1}{4}(1 + F_s) - \frac{1}{4}(1 + F_d) = \frac{1}{2} - \frac{1}{4}(F_s + F_d).$$

If only one parent is known:

$$d_{xx} = 1 - \frac{1}{4}(1 + F_s) = \frac{3}{4} - \frac{1}{4}(F_s).$$

If neither parent is known:

$$d_{xx} = 1.$$

For the pedigree in Table 1.1, and the above calculation, the diagonal of matrix \mathbf{D} is $diag(1.0, 1.0, 0.75, 0.5, 0.5, 0.5, 0.406)$. The (co)variance matrix of Mendelian sampling is $\mathbf{D}\sigma_u^2 = diag(1.0, 1.0, 0.75, 0.5, 0.5, 0.5, 0.406) \sigma_u^2$.

The equation (1.1) is presented for individual x . However, using it recursively reveals that breeding values are linear combinations of Mendelian sampling terms. The breeding value of an individual x can be expressed (using equation 1.1) in terms of its parents' breeding values and the Mendelian sampling term. The sire's (or dam's) breeding value can also be written using equation (1.1) and so on. Thus, this equation used recursively goes back through the pedigree and we can partition each breeding value (for an individual x) as $u_x = \sum_{j=1}^m t_{xj} \phi_j$, (m total number of animals) where u_x is decomposed into a sum of j independent terms or contributions that involve the breeding value of founders and the Mendelian sampling of non-founders.

Taking our pedigree of 7 individuals (Figure 1.1), and using the equation (1.1), we partitioned the breeding value of the individuals into Mendelian sampling terms. The breeding values of founders are equal to their Mendelian sampling. Note in the example, that each colour represents an individual and it shows the flow of genes from this individual through the pedigree.

$$u_1 = \phi_1,$$

$$u_2 = \phi_2,$$

$$u_3 = \frac{1}{2}u_1 + \phi_3 = \frac{1}{2}\phi_1 + \phi_3,$$

$$u_4 = \frac{1}{2}u_1 + \frac{1}{2}u_2 + \phi_4 = \frac{1}{2}\phi_1 + \frac{1}{2}\phi_2 + \phi_4,$$

$$u_5 = \frac{1}{2}u_3 + \frac{1}{2}u_4 + \phi_5 = \frac{1}{2}\left(\frac{1}{2}\phi_1 + \phi_3\right) + \frac{1}{2}\left(\frac{1}{2}\phi_1 + \frac{1}{2}\phi_2 + \phi_4\right) + \phi_5 = \frac{1}{2}\phi_1 + \frac{1}{4}\phi_2 + \frac{1}{2}\phi_3 + \frac{1}{2}\phi_4 + \phi_5,$$

$$u_6 = \frac{1}{2}u_1 + \frac{1}{2}u_4 + \phi_6 = \frac{3}{4}\phi_1 + \frac{1}{4}\phi_2 + \frac{1}{2}\phi_4 + \phi_6,$$

$$u_7 = \frac{1}{2}u_6 + \frac{1}{2}u_5 + \phi_7 = \frac{5}{8}\phi_1 + \frac{1}{4}\phi_2 + \frac{1}{4}\phi_3 + \frac{1}{2}\phi_4 + \frac{1}{2}\phi_5 + \frac{1}{2}\phi_6 + \phi_7.$$

We can see that the breeding value of each individual is equal to the sum of the ancestral Mendelian samplings weighted, plus its individual's Mendelian sampling. In matrix notation, we can write a general expression, $\mathbf{u} = \mathbf{T}\boldsymbol{\phi}$, showing that the vector of breeding values (\mathbf{u}) is a linear combination of Mendelian sampling terms and their ancestors. The vector $\boldsymbol{\phi}$ also includes the breeding values of the pedigree founders. For our 7-animals pedigree (Figure 1.1), $\mathbf{u} = \mathbf{T}\boldsymbol{\phi}$ is equal to:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 1 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/2 & 1/2 & 1 & 0 & 0 \\ 3/4 & 1/4 & 0 & 1/2 & 0 & 1 & 0 \\ 5/8 & 1/4 & 1/4 & 1/2 & 1/2 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \phi_5 \\ \phi_6 \\ \phi_7 \end{bmatrix}.$$

To explain the \mathbf{T} matrix in the example (Figure 1.1), we take the individual 5. Individual 5 is the offspring of individuals 3 and 4. Individual 5 inherited 50% of his genes from his sire (individual 3), and 50% of his genes from his dam (individual 4), as $\left(\frac{1}{2}u_3 + \frac{1}{2}u_4\right)$. From these 50% that he inherited from his dam (individual 4), 25% is from individual 1 (individual 4's sire), and 25% from individual 2 (individual 4's dam), and 50% from individual 4's Mendelian sampling $\left(\frac{1}{2}u_4 = \frac{1}{2}\left(\frac{1}{2}\phi_1 + \frac{1}{2}\phi_2 + \phi_4\right)\right)$. Likewise, from the 50% of individual 5's genes that he inherited from his sire (individual 3), 25% is from individual 1 (the sire of individual 3) and 50% from individual 3's Mendelian sampling $\left(\frac{1}{2}u_3 = \frac{1}{2}\left(\frac{1}{2}\phi_1 + \phi_3\right)\right)$. The ancestors for individual 5 are animals 1, 2, 3 and 4. Taking into account both sides (paternal and maternal) for individual 5, individual 1 contributes $\frac{1}{2}$ from $\left(\frac{1}{2}\left(\frac{1}{2}\phi_1\right) + \frac{1}{2}\left(\frac{1}{2}\phi_1\right)\right)$, individual 2 contributes $\frac{1}{4}$ from $\left(\frac{1}{2}\left(\frac{1}{2}\phi_2\right)\right)$, individual 3 contributes $\frac{1}{2}$ from $\left(\frac{1}{2}\left(\phi_3\right)\right)$, and individual 4 contributes $\frac{1}{2}$ from $\left(\frac{1}{2}\left(\phi_4\right)\right)$. Note that these coefficients (contributions) given for the ancestor Mendelian

sampling of individual 5 are the values in the \mathbf{T} matrix (line 5). This example shows the gene flow through the pedigree, where \mathbf{T} relates the individual 5 and its ancestors (animals 1, 2, 3 and 4). The coefficients (contributions) given for the ancestors Mendelian sampling of individual 5 are $t_{51} = \frac{1}{2}$, $t_{52} = \frac{1}{4}$, $t_{53} = \frac{1}{2}$ and $t_{54} = \frac{1}{2}$. The coefficient t_{xj} is the genetic contribution of an ancestor j to an individual x , called $r_{j(x)}$ by Woolliams *et al.* (1999). Woolliams *et al.* (1999) defined the long-term genetic contribution ($r_{j(x)}$, the same as t_{xj}), as the proportion of the genes in individual x transmitted from j .

The equation (1.1) also holds for unbiased predictors, such as the Best Linear Unbiased Predictor (BLUP, Henderson, 1973) of the estimated breeding values ($\hat{\mathbf{u}}$) (García-Cortés *et al.*, 2008) as follows: $\hat{\mathbf{u}} = \mathbf{T}\hat{\boldsymbol{\phi}}$, where $\hat{\mathbf{u}}$ is a vector that contains the estimated breeding values (EBV), and the Mendelian samplings can be estimated as $\hat{\boldsymbol{\phi}} = \mathbf{T}^{-1}\hat{\mathbf{u}}$.

Animals are selected based on EBV. The focus on EBV may hide the role of Mendelian sampling in creating genetic progress (Bijma *et al.*, 2018). The role of Mendelian sampling on the selection process are expanded below.

1.2 Genetic contributions

The genetic contribution of an ancestor is defined as the proportion of all distinct genealogical pathways that travel from this ancestor to a group of descendants (Woolliams *et al.*, 1999, Woolliams, 2007). In breeding programs, some groups or individuals are prioritized over the others by their superiority (*e.g.* AI males). These superior animals may have higher contributions than the others for the upcoming generations (Wray and Thompson, 1990). This concept was used to account for the genetic gain, inbreeding and other phenomena. The genetic contribution of an individual represents the genetic contribution of the individual's Mendelian sampling to the long-term gene pool. This definition of the contribution allows to know the future gene pool represents the contributions of all ancestors (founders and non-founders) and not only founders (Woolliams, 2007). The genetic contribution ($r_{j(x)}$) of an ancestor j born at time t_1 to a descendant x born at time $t_2 (> t_1)$, is the proportion of the genes of x that are expected to be inherited from ancestor j (Woolliams *et al.*, 1999). In the long-term, the contribution of j , as $t_2 - t_1 \rightarrow \infty$, $r_{j(x)}$ tends to stabilize. The long-term genetic contribution

differs among individual ancestors depending on the lifetime breeding use of different ancestors, their EBV, and other factors (Woolliams *et al.*, 1999).

Woolliams and Thompson (1994) defined the breeding value of an individual as the weighted sum of all the Mendelian sampling of its ancestors, which led to the genetic gain expressed as $E(\Delta G) = \sum_{j=1}^m t_{x_j} \phi_j$. The genetic gain is a function of the long-term contributions and Mendelian sampling. Thus, genetic progress can be understood as the selection of positive ancestors' Mendelian samplings to their descendants and also candidates' own Mendelian samplings (Woolliams and Thompson, 1994).

1.3 Decomposition of genetic trend

From the equation: $\hat{\boldsymbol{\phi}} = \mathbf{T}^{-1}\hat{\mathbf{u}}$, García-Cortés *et al.* (2008) proposed to define a set of k partitions such that $\mathbf{P}_1 + \mathbf{P}_2 + \dots + \mathbf{P}_k = \mathbf{I}$. where \mathbf{P} is a diagonal matrix, that contains zeros and ones and it selects the corresponding columns of \mathbf{T} , or the corresponding ancestor effects for a given $\hat{\mathbf{u}}$. These partitions are based on groups or categories of interest. For example, the genetic gain can be partitioned using a category defined by sex or k categories defined by group of animals (*e.g.* AI males, dams of sires, etc.). The EBV can be partitioned into k categories as follows: $\hat{\mathbf{u}} = \hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2 + \dots + \hat{\mathbf{u}}_k$, where $\hat{\mathbf{u}}_k$ is a part of the breeding value contributed by the category k . This decomposition enables the calculation of the partial genetic trends directly from the EBV, and allows us to inspect the contributions of each category to the genetic gain. By following our example of the pedigree of 7 individuals, we can partition the EBV according to sex ($k = 2$) using \mathbf{P}_m and \mathbf{P}_f for males and females ($\mathbf{P}_m + \mathbf{P}_f = \mathbf{I}$) as

$$\mathbf{P}_m = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{P}_f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

The ones in the diagonal in \mathbf{P}_m (and \mathbf{P}_f), represent the individuals 1, 3 and 6 (2, 4, 5 and 7) in Figure 1.1 that are males (females). Then, following García-Cortés *et al.* (2008)

$$\hat{\mathbf{u}} = \mathbf{T}\hat{\boldsymbol{\phi}} = \mathbf{T}(\mathbf{P}_m + \mathbf{P}_f)\hat{\boldsymbol{\phi}} = \mathbf{T}\mathbf{P}_m\hat{\boldsymbol{\phi}} + \mathbf{T}\mathbf{P}_f\hat{\boldsymbol{\phi}},$$

and replacing $\hat{\phi}$ by $T^{-1}\hat{u}$,

$$\hat{u} = TP_m T^{-1}\hat{u} + TP_f T^{-1}\hat{u} = \hat{u}_m + \hat{u}_f,$$

where, $\hat{u}_m = TP_m T^{-1}\hat{u}$ and $\hat{u}_f = TP_f T^{-1}\hat{u}$ for males and females, respectively.

Note that by doing this, we have split, for each individual, the EBV into contributions due to Mendelian samplings of their ancestors, which, in turn, we then assign into categories (by sex in our 7-individuals pedigree). This is why García-Cortés *et al.* (2008) consistently uses the naming ‘‘Mendelian sampling’’ for this decomposition. More details about this method and its use are presented in chapter 2.

1.4 Decomposition of inbreeding

Selection in animal populations results in remarkable genetic gain but also in an unavoidable accumulation of inbreeding. The partitioning of inbreeding among the ancestors of an individual is called Mendelian decomposition of inbreeding, and involves inbreeding contributed by founders and by so-called Mendelian sampling of non-founders (Caballero and Toro, 2000). Mendelian decomposition of inbreeding traces back the specific ancestral paths through which the identical by descent (IBD) alleles are inherited. Using pedigree analysis, we can determine the proportion of total inbreeding that can be attributed to each ancestor (García-Cortés *et al.*, 2010).

The inbreeding coefficient of individual x , F_x may be decomposed in a sum of partial inbreeding coefficients each due to an ancestor j , e.g. $F_{x(j)}$ where j is an ancestor of x . To calculate partial inbreeding coefficients attributed to an animal x , we used the approach proposed by García-Cortés *et al.* (2010) which modified the conventional tabular method with a set of recursively formulas. The method operates recursively over $a_{(x,j)} = \frac{1}{2}(a_{(x,s')} + a_{(x,d')}) + \phi_{xj}$, where $a_{(x,j)}$ is the additive genetic relationship between individuals x and j (or two times the coancestry between those two individuals), and s' and d' are the sire and dam of j and ϕ_{xj} is the Mendelian sampling variation and it is related to within-family variation. For a given matrix $\phi = \{\phi_{xj}\}$, $\phi = 0$ except at the diagonal element ϕ_{xx} of the self-additive genetic relationship that are handled as one of the following options: (1) when both parents are

known, we use $\phi_{xx} = \frac{1}{4}(1 - F_s) + \frac{1}{4}(1 - F_d)$ where F_s (F_d) is the inbreeding coefficient of the sire (or dam); (2) when only one parent ($l = s, d$) is known, we use $\phi_{xx} = \frac{1}{2} + \frac{1}{4}(1 - F_l)$ and (3) when both parents are unknown we use $\phi_{xx} = 1$. Each individual in the population is considered as a partial founder where it contributes to the genetic variability. The term ϕ includes the Mendelian sampling variability and the ignorance about the knowledge of the parents (García-Cortés *et al.*, 2010). Most of ancestors do not generate partial inbreeding; only those ancestors that are common to the mother and father of x .

Let's go back to our 7-individuals pedigree (Figure 1.2), this figure is the same as Figure 1.1 but we changed the colours to show individual 7 in blue and his parents (5 and 6) in purple. Parents of individual 7 are animals 5 and 6 (presented in purple), and the sources of the coancestry between the parents of 7 are: animals 1 and 2 (founders) and Mendelian sampling of animal 4 (non-founder). Thus, the ancestors, source of coancestry, of 7 are animals 1, 2 and 4. Now we will calculate the partial inbreeding coefficients from the source of coancestry of 7 (1, 2 and 4). Founders 1 and 2 have none of their parents known, then $\phi_{11} = \phi_{22} = 1$; for animal 4, both parents are known and both parents are not inbred, then $\phi_{44} = \frac{1}{4}(1 - F_5) + \frac{1}{4}(1 - F_6) = \frac{1}{2}$. Then, the partial inbreeding coefficient of animal 7 attributed to ancestor 1 ($F_{7(1)}$) is calculated as in Figure 1.3. This partial inbreeding coefficient of animal 7 attributed to ancestor 1 is equal to the coancestry coefficient of its parents and is equal to the half of the additive genetic relationship coefficient of its parents 5 and 6 ($\frac{1}{2} a_{56(1)}$).

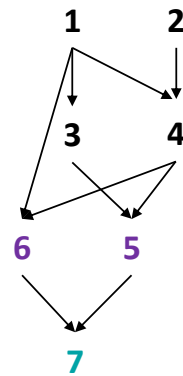


Figure 1.2. Pedigree of animal 7

	$\begin{matrix} 0 & 0 \\ \diagdown & / \\ 1 \end{matrix}$	$\begin{matrix} 0 & 0 \\ \diagdown & / \\ 2 \end{matrix}$	$\begin{matrix} 1 & 0 \\ \diagdown & / \\ 3 \end{matrix}$	$\begin{matrix} 1 & 2 \\ \diagdown & / \\ 4 \end{matrix}$	$\begin{matrix} 3 & 4 \\ \diagdown & / \\ 5 \end{matrix}$	$\begin{matrix} 1 & 4 \\ \diagdown & / \\ 6 \end{matrix}$	$\begin{matrix} 5 & 6 \\ \diagdown & / \\ 7 \end{matrix}$
1	1	0	0.5	0.5	0.5	0.75	0.625
2	0	0	0	0	0	0	0
3	0.5	0	0.25	0.25	0.25	0.375	0.3125
4	0.5	0	0.25	0.25	0.25	0.375	0.3125
5	0.5	0	0.25	0.25	0.25	0.375	0.3125
6	0.75	0	0.375	0.375	0.375	0.5625	0.46875
7	0.625	0	0.3125	0.3125	0.3125	0.46875	0.390625

$$F_{7(1)} = \frac{1}{2} a_{56(1)} = 0.1875$$

Figure 1.3. Partial inbreeding coefficient of animal 7 from founder 1

	$\begin{matrix} 0 & 0 \\ \swarrow & \searrow \\ 1 & 2 \end{matrix}$	$\begin{matrix} 0 & 0 \\ \swarrow & \searrow \\ 2 & 3 \end{matrix}$	$\begin{matrix} 1 & 0 \\ \swarrow & \searrow \\ 3 & 4 \end{matrix}$	$\begin{matrix} 1 & 2 \\ \swarrow & \searrow \\ 4 & 5 \end{matrix}$	$\begin{matrix} 3 & 4 \\ \swarrow & \searrow \\ 5 & 6 \end{matrix}$	$\begin{matrix} 1 & 4 \\ \swarrow & \searrow \\ 6 & 7 \end{matrix}$	$\begin{matrix} 5 & 6 \\ \swarrow & \searrow \\ 7 & 8 \end{matrix}$
	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2	0	1	0	0.5	0.25	0.25	0.25
3	0	0	0	0	0	0	0
4	0	0.5	0	0.25	0.125	0.125	0.125
5	0	0.25	0	0.125	0.0625	0.0625	0.0625
6	0	0.25	0	0.125	0.0625	0.0625	0.0625
7	0	0.25	0	0.125	0.0625	0.0625	0.0625

$F_{7(2)} = \frac{1}{2} a_{56(2)} = 0.03125$

Figure 1.4. Partial inbreeding coefficient of animal 7 from founder 2

	$\begin{matrix} 0 & 0 \\ \swarrow & \searrow \\ 1 & 2 \end{matrix}$	$\begin{matrix} 0 & 0 \\ \swarrow & \searrow \\ 2 & 3 \end{matrix}$	$\begin{matrix} 1 & 0 \\ \swarrow & \searrow \\ 3 & 4 \end{matrix}$	$\begin{matrix} 1 & 2 \\ \swarrow & \searrow \\ 4 & 5 \end{matrix}$	$\begin{matrix} 3 & 4 \\ \swarrow & \searrow \\ 5 & 6 \end{matrix}$	$\begin{matrix} 1 & 4 \\ \swarrow & \searrow \\ 6 & 7 \end{matrix}$	$\begin{matrix} 5 & 6 \\ \swarrow & \searrow \\ 7 & 8 \end{matrix}$
	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	0	0	0	0.5	0.125	0.125	0.125
5	0	0	0	0.25	0.125	0.125	0.125
6	0	0	0	0.25	0.125	0.125	0.125
7	0	0	0	0.25	0.125	0.125	0.125

$F_{7(4)} = \frac{1}{2} a_{56(4)} = 0.0625$

Figure 1.5. Partial inbreeding coefficient of animal 7 from animal 4

The partial inbreeding coefficient of animal 7 attributed to ancestor 2 ($F_{7(2)}$) and 4 ($F_{7(4)}$) are calculated as in Figure 1.4 and 1.5 respectively. Thereafter, Mendelian decomposition for animal 7 from ancestors 1, 2 and 4 allows to partition inbreeding coefficient of individual 7 in:

- 1 $F_{7(1)} = f_{56(1)} = 0.1875$
- 2 $F_{7(2)} = f_{56(2)} = 0.03125$
- 3 $F_{7(4)} = f_{56(4)} = 0.0625$

Inbreeding coefficient of individual 7 (F_7) is the sum of the partial inbreeding coefficients from the 3 ancestors $F_7 = 0.1875 + 0.03125 + 0.0625 = 0.28125$.

1.5 Inbreeding load

Inbreeding load (IL) is the fraction of the mutation load that is due to hidden recessive alleles in heterozygous state. This load, when exposed by inbreeding, is responsible for inbreeding depression, the decrease in performance and fitness in inbred individuals (Leroy, 2014). This load is unevenly distributed through individuals. According to population genetics theory, this load depends on the selection process, mutations, allelic recombination and crossing-overs (Bosse *et al.*, 2019). Accordingly, some individuals carry less recessive deleterious mutations than others (Gulisija *et al.*, 2006; Casellas *et al.*, 2008). For instance, one could in principle find out if different individuals carry different IL by producing *e.g.* equally inbred descendance (say mating sires to their daughters) and comparing descendants across sires. In complex pedigrees this becomes more complex because each individual possesses parts of inbreeding coming potentially from different ancestors. These fractions can however be computed using pedigree by the Mendelian decomposition of inbreeding. IL is a heritable trait, and it acts additively in the progeny from the perspective of the ancestors (Varona *et al.*, 2019). Even if IL is an additive trait, it is expressed only in inbred individuals.

Using the Mendelian decomposition of inbreeding, a linear model can estimate genetic variance and breeding values of the IL for a specific trait (*e.g.* milk yield) (Varona *et al.*, 2019). Variance of IL has been estimated in growth traits in rabbits (Casellas, 2018), pigs (Casellas *et al.*, 2008) and beef cattle (Varona *et al.*, 2019); to morphological traits in horses (Poyato-Bonilla *et al.*, 2020) and fertility traits (Perdomo-González *et al.*, 2021) in horses and dairy cattle (Martinez-Castillero *et al.*, 2021).

The theory that shows the additive nature of the IL; and the genetic correlation between IL and the additive genetic or breeding values is always negative, is present in chapter 3.

1.6 Dairy sheep in France

Dairy sheep farming has an undeniable economic, social and environmental value in France. Genetic improvement of dairy sheep breeds involves five local breeds: Basco-Béarnaise (BB), Corse (COR), Lacaune (LAC), Manech Tête Noire (MTN) and Manech Tête Rousse (MTR). These breeds are distributed in 3 geographic areas (Figure 1.6): Southern Massif Central with the LAC breed (44% of the flocks), Western Pyrenees with the BB, MTN and MTR breeds

(40% of the flocks), and Corsica with the COR breed (8% of the flocks). COR breed is not studied in this thesis.



Figure 1.6. Distribution of the 5 French dairy sheep breeds in France (Races de France, 2011).

Dairy sheep breeds benefit from genetic improvement programs started in the 1970s. Artificial insemination is used for the spread of genetic improvement and (until 2016) for the selection of elite males by progeny testing. The use of mechanical milking, the improvement of nutritional systems, the adoption of AI, and the implementation of selection schemes with genetic evaluations, incremented the size of flocks as well as the milk production by ewe (Figure 1.7) (Lagriffoul *et al.*, 2016).

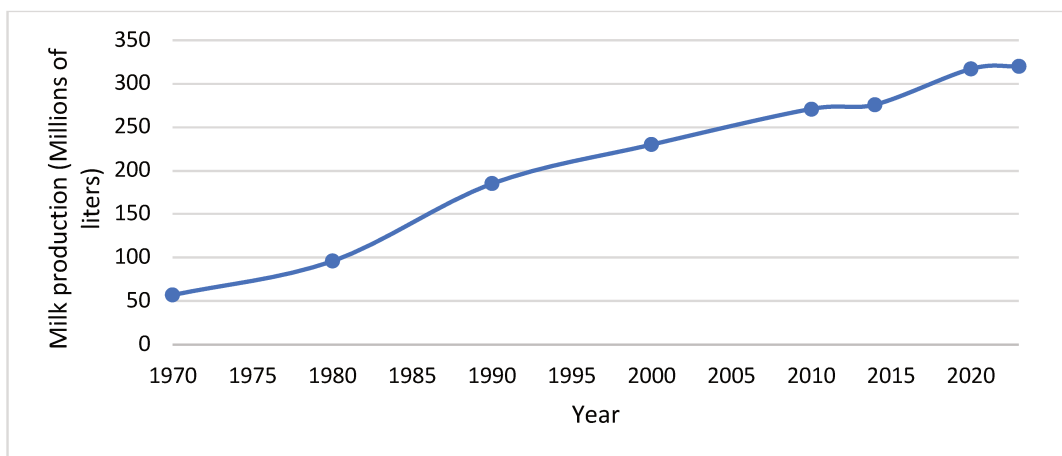


Figure 1.7. Evolution of the sheep milk production in France (Lagriffoul, personal communication).

The selection scheme has a pyramidal organisation (Figure 1.8), with the breeders of nucleus flocks (where the official milk recording is done) at the top, and the commercial flock (practicing a simplified milk recording) at the bottom. Selection flocks are limited to 15 to 25% of the total ewes of each breed and generate the genetic gain and produce rams for the whole population (Barillet, 1997). In the nucleus, pedigree and official milk recording, AI, controlled natural mating, and breeding value estimation are carried out to generate genetic progress. The genetic progress is then transferred to the commercial flocks through AI or natural-mating rams (Barillet, 1997; Carta *et al.*, 2009). The size of the selection scheme differs among breeds. LAC and MTR are the two most significant populations.

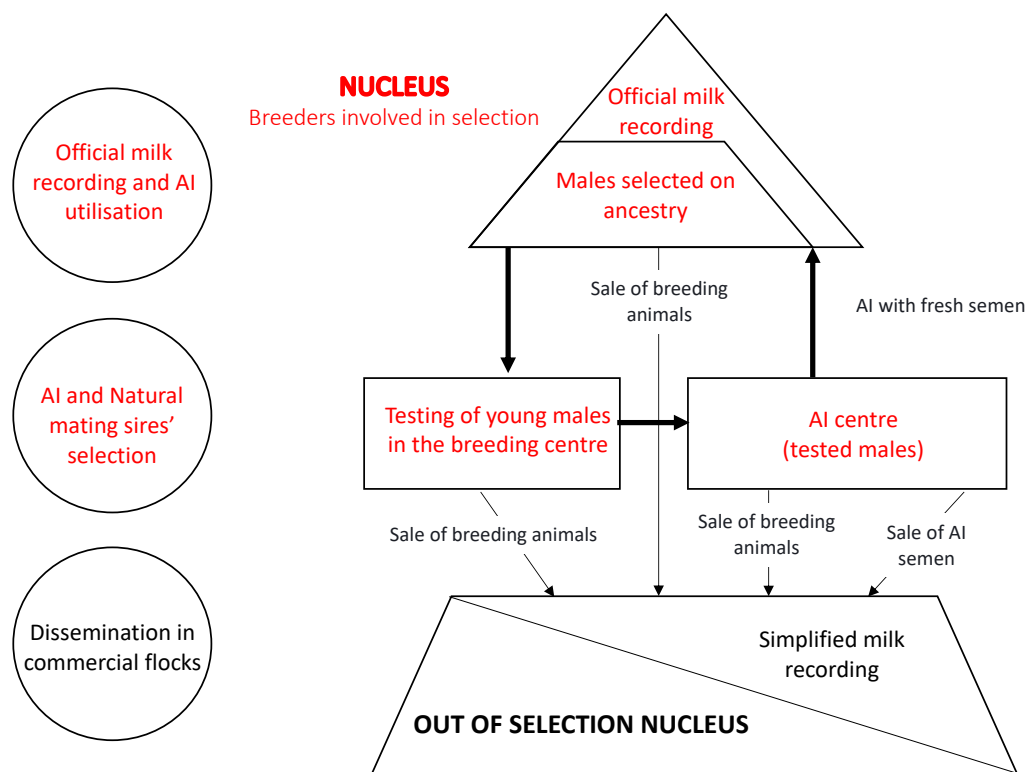


Figure 1.8. The selection scheme of French dairy sheep breeds (Barillet, 1997).

The breeding goals differ between LAC, Pyrenean and COR breeds. The breeding objectives for BB, MTN, MTR and LAC breeds include milk yield, fat and protein content and somatic cell score. Additionally, the breeding objective for the LAC breed (and only the LAC breed), includes selection for udder morphology traits. However, for the COR breed, the selection objective only includes milk yield. Additionally, LAC system still has dual purpose,

milk and meat production, whereas the other breeds are only for milk production (Astruc *et al.*, 2016).

1.6.1 Conventional breeding schemes

Conventional breeding schemes were based on performance recording and progeny testing of artificial insemination (AI) rams. All concerned breeds have performance recording according to the International Committee for Animal Recording (ICAR) rules. The AI rate in the nucleus varied from 45% (COR) to 85% (LAC). The number of yearly progeny-tested rams, in the AI center, varied from 30 rams in COR breed to 440 in LAC breed (Table 1.2). The AI center plays an important role in the selection and in the genetic dispersion progress. Rams in the progeny test had around 30 to 40 daughters in their first lactation. The progeny test and the genetic diffusion were done with the AI fresh semen which imposes some limitations (seasonal and geographical) and it led to the presence in the AI centre of a great number of rams (Astruc *et al.*, 2016). Annually, no more than 1,500 doses of fresh semen were distributed. Generation interval between sire and male offspring varied between 4 years (LAC) to 7 years (MTN) before the genomic era.

Table 1.2. Number of individuals in the selection scheme in 2014 before genomic selection (Astruc *et al.*, 2016).

Breed	Female population size	Percentage of female in the nucleus (%)	Tested rams	Rams in the AI center	Sire of sire in the AI center
BB ¹	80,000	32	50	200	17
COR ²	83,000	19	30	150	22
MTN ³	80,000	16	30	175	17
MTR ⁴	274,000	29	150	600	30
LAC ⁵	890,000	19	440	1,400	70-80

¹BB = Basco-Béarnaise; ²COR = Corse; ³MTN = Manech Tête Noire; ⁴MTR = Manech Tête Rousse; ⁵LAC = Lacaune.

The efficiency of these schemes was very good, characterized by an annual genetic gain between 0.10 (COR) to 0.23 (LAC) genetic standard deviation depending on the number of rams tested by year and their oldness (2 to 5 decades) (Astruc *et al.*, 2016).

1.6.2 Genomic selection

After genomic selection (GS) started, approximately in 2015 for the LAC, 2017 for Pyrenean breeds and 2020 for COR breed, the progeny testing was replaced by the use of the genomic rams. The GS allows to predict accurately the genetic value of young males. With GS, a quicker turnover of male is obtained by the use of AI rams without waiting for progeny testing leading to a reduction of the total number of rams in the AI center (Table 1.3) (Legarra *et al.*, 2014). The rate of AI in the nucleus flock varies between 62% (LAC) to 88% (MTN) (Astruc *et al.*, 2022). The use of GS leads also to a reduction of the gap on the genetic gain between the nucleus and the commercial flocks, and opens new opportunities for selection criteria on new traits (Duchemin *et al.*, 2012). The annual genetic gain varies between 0.11 (COR) to 0.35 (LAC) genetic standard deviation (Astruc *et al.*, 2022). Generation interval between sire and male offspring varies between 2.8 years (BB, MTN, LAC and MTR) to 5.1 years (COR) in 2022 (Astruc *et al.*, 2022).

Table 1.3. Genomic breeding schemes in French dairy sheep in 2021 (Astruc *et al.*, 2022).

Breed	First year of GS ⁶	Females in selection	New AI rams per year	Number of AI rams	Rams genotyped per year	Reference population*
BB ¹	2017	28,061	60	110	240	880
COR ²	2020	18,863	21	28	374	326
MTN ³	2017	10,547	28	68	102	585
MTR ⁴	2017	84,286	174	302	714	2,842
LAC ⁵	2015	192,928	297	674	2,974	6,397

¹BB = Basco-Béarnaise; ²COR = Corse; ³MTN = Manech Tête Noire; ⁴MTR = Manech Tête Rousse;

⁵LAC = Lacaune; ⁶GS = Genomic selection.

* Rams genotyped with daughters

1.7 Objectives

This thesis explores pedigree-based methods to partition genetic gain and inbreeding load in French dairy sheep breeds: BB, LAC, MTN and MTR.

1.8 Outlines

This thesis is organised in 5 chapters. A general introduction and the objectives of the thesis are presented first (chapter 1). The body of the thesis is a compilation of three scientific

papers or manuscripts (chapter 2 to chapter 4), followed by a general discussion, perspectives and general conclusions (chapter 5).

The chapter 2 used a retrospective analysis to decompose the genetic trend of milk yield in Mendelian samplings by categories of animals defined by sex and by selection pathways, and to similarly characterize long-term genetic contributions. We analysed genetic gain for milk yield in four French dairy sheep breeds: LAC, BB, MTN and MTR.

Chapter 3 showed that inbreeding load can be expressed as an additive genetic effect based on substitution effect under non-random matings. Second, we calculated and described the partial inbreeding coefficients in three French dairy sheep populations (BB, MTN and MTR). Third, we included these coefficients in a mixed model as random regression covariates, to predict genetic variance and breeding values of the inbreeding load for milk yield in the same breeds.

In the chapter 4, we evaluated the effectiveness of involving inbreeding load in selection strategies in a dairy sheep breeding scheme. We did this by computer simulation of 10 generations of evaluation and selection. Six scenarios that differ in the criteria of selection (only breeding values, only breeding values of inbreeding load, or both genetic and inbreeding load breeding values) and mate allocation strategies (minimising inbreeding load or minimising expected future inbreeding) were evaluated. Scenarios were compared in terms of genetic gain, pedigree-based inbreeding coefficients, rate of inbreeding, effective population size, and accuracy of selection.

Chapter 2

Partition of the genetic trend of French dairy sheep in Mendelian samplings and long-term contributions

S. Antonios^{1*}, A. Legarra^{1,2}, R. Pong-Wong³, J.M. Astruc⁴, S.T. Rodríguez-Ramilo¹
and Z.G. Vitezica¹

¹ GenPhySE, INPT, INRAE, ENVT, F-31326, Castanet Tolosan, France

² Current address: Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

³ The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush

Campus, Edinburgh, EH25 9RG, UK

⁴ Institut de l'Élevage, 149 rue de Bercy, F-75595 Paris, France

Journal of Dairy Science 106, 6275-6287 (2023)



Partitioning of the genetic trends of French dairy sheep in Mendelian samplings and long-term contributions

S. Antonios,^{1*} A. Legarra,^{1†} R. Pong-Wong,² J. M. Astruc,³ S. T. Rodríguez-Ramilo,¹ and Z. G. Vitezica¹

¹GenPhySE, INPT, INRAE, ENVT, F-31326, Castanet Tolosan, France

²The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush Campus, Edinburgh, EH25 9RG, United Kingdom

³Institut de l'Élevage, 149 rue de Bercy, F-75595 Paris, France

ABSTRACT

The genetic trend of milk yield for 4 French dairy sheep breeds (Lacaune, Basco-Béarnaise, Manech Tête Noire, and Manech Tête Rousse) was partitioned in Mendelian sampling trends by categories of animals defined by sex and by selection pathways. Five categories were defined, as follows: (1) artificial insemination (AI) males (after progeny testing), (2) males discarded after progeny testing, (3) natural mating males, (4) dams of males, and (5) dams of females. Dams of males and AI males were the most important sources of genetic progress, as observed in the decomposition in Mendelian sampling trends. The yearly contributions were more erratic for AI males than for dams of males, as AI males are averaged across a smaller number of individuals. Natural mating males and discarded males did not contribute to the trend in terms of Mendelian sampling, as their estimated Mendelian sampling term is either null (natural mating males) or negative (discarded males). Overall, in terms of Mendelian sampling, females contributed more than males to the total genetic gain, and we interpret that this is because females constitute a larger pool of genetic diversity. In addition, we computed long-term contributions from each individual to the following pseudo-generations (one pseudo-generation spanning 4 years). With this information, we studied the selection decisions (selected or not selected) for females, and the contributions to the following generations. Mendelian sampling was more important than parent average to determine the selection of individuals and their long-term contributions. Long-term contributions were greater for AI males (with larger progeny sizes than females) and in Basco-Béarnaise than in Lacaune (with the latter being a larger population).

Key words: genetic gain, genetic trend, Mendelian sampling, long-term contributions, selection scheme

INTRODUCTION

In animal genetic improvement, selection is based on EBV of candidates for selection. The methodological focus on EBV prediction may hide the role of Mendelian sampling in creating genetic progress. The breeding value of an individual can be expressed as the average of its parental breeding values plus a Mendelian sampling term. As the Mendelian sampling is the unique portion of genetic variation that the individual brings into the population (Woolliams, 2007), this term captures the “originality” of the individual with respect to its parents. In fact, the Mendelian sampling represents the deviation arising from recombination and segregation of parental chromosomes. The breeding value of an individual can be decomposed as a sum of Mendelian samplings of all its ancestors. Thus, the genetic progress can be understood as the selection of positive ancestors' Mendelian samplings to their descendants and also candidates' own Mendelian samplings (Woolliams and Thompson, 1994).

The long-term genetic contribution of an individual is its proportional contribution to the long-term genetic background of the population (Bijma, 2000). The contribution, $r_{i(j)}$, is defined as the expected fraction of genes from ancestor i (born at time t_1) transmitted to descendant j (born at a later time, t_2 ; Woolliams et al., 1993; Woolliams, 2007). When $t_2 - t_1 \rightarrow +\infty$, this distant future implies that contributions fluctuate at the beginning and stabilize in the long run. The long-term genetic contribution of an individual depends on the genetic superiority for selection of the animal; for example, when selecting based on EBV, the individuals with highest intrageneration EBV will be selected to have more offspring, which increases their long-term genetic contribution (Woolliams et al., 1999).

Received November 9, 2022.

Accepted February 28, 2023.

*Corresponding author: simona.antonios@inrae.fr

†Current address: Council on Dairy Cattle Breeding, Bowie, MD 20716.

The study of Mendelian sampling trends by groups and categories of animals allows understanding of selection decisions (García-Cortés et al., 2008), whereas analyzing long-term contributions of ancestors to descendants allows understanding of the gene flow across generations and by categories of animals (Woolliams et al., 1999). For a better understanding of the selection schemes and their implementation, the dynamics of the contributions, and the sources of “originalities,” we used a retrospective analysis of Mendelian sampling trends and long-term genetic contributions in selected populations.

Partitioning the genetic gain by categories of animals and Mendelian samplings as described above has mainly been evaluated in breeding schemes importing animals from foreign populations. Examples can be found in the literature dealing with cattle (Gorjanc et al., 2011) and pig (Škorput et al., 2015) breeding schemes. However, to date, this method has not been used in closed breeds within country. Our work focuses in 4 French dairy sheep pure breeds: Lacaune (**LAC**) and the Western Pyrenees breeds Basco-Béarnaise (**BB**), Manech Tête Noire (**MTN**), and Manech Tête Rousse (**MTR**), which are local dairy sheep breeds with no introduction from other breeds or countries. Breeding schemes of these breeds were based on performance recording and progeny testing via AI rams (Astruc et al., 2002) until 2016, from which time they have been based on performance recording and genomic selection, with early selection of rams and widespread use of AI. The breeding goals include, depending on the breed, milk yield, fat and protein yields, fat and protein contents, somatic cell score, udder morphology, and other traits (Astruc et al., 2018); however, milk yield is highly selected in all breeds. Although some of the traits (e.g., fat and protein yields) are correlated with this trait, we will focus on milk yield alone, as otherwise the analyses become too complex and difficult to compare across breeds.

The objective of this work was to fine partition genetic trends in Mendelian samplings by categories of animals defined by sex and by selection pathways, and to similarly characterize long-term genetic contributions. We analyzed genetic gain for milk yield in 4 French dairy sheep breeds: LAC, BB, MTN, and MTR.

MATERIALS AND METHODS

Because no human or animal subjects were used, this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

Decomposition of Genetic Trend in Mendelian Samplings and Long-Term Genetic Contributions

The breeding or additive genetic value for an animal i (u_i) can be expressed in terms of its parent (sire and dam) breeding values (u_s and u_d , respectively), whose average is called parent average, plus the Mendelian sampling of animal i (ϕ_i ; Thompson, 1979) as

$$u_i = \frac{1}{2}u_s + \frac{1}{2}u_d + \phi_i. \quad [1]$$

In matrix notation, the vector of breeding values (\mathbf{u}) can be written as a linear combination of Mendelian sampling terms and their ancestors as $\mathbf{u} = \mathbf{T}\boldsymbol{\phi}$ (where vector $\boldsymbol{\phi}$ also includes the breeding values of the pedigree founders). The triangular matrix \mathbf{T} describes the flow of genes through pedigree (Woolliams et al., 1999) and shows the relatedness between each individual and his ancestor (García-Cortés et al., 2008). Accordingly, $\boldsymbol{\phi}$ is the vector of Mendelian sampling terms, assumed distributed as $\boldsymbol{\phi} \sim N(0, \mathbf{D}\sigma_u^2)$ (García-Cortés et al., 2008), where \mathbf{D} is a diagonal matrix with Mendelian sampling variances for nonfounders and additive genetic variances for founders. The vector of breeding values is thus distributed as $\mathbf{u} \sim N(0, \mathbf{A}\sigma_u^2)$, where $\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}'$ is the additive genetic relationship matrix and σ_u^2 is the additive genetic variance.

Equation [1] and its matrix notation reveal that breeding values are linear combinations of Mendelian sampling terms. The breeding value of an individual i born can be expressed using Equation [1] in terms of its parents' breeding values and the Mendelian sampling term. The sire's breeding value can also be written using Equation [1] and so on. Thus, Equation [1] can be used recursively going back through the pedigree, and we can partition each breeding value (for an individual

i) as $u_i = \sum_{j=1}^m T_{ij}\phi_j$ (m total number of animals), where u_i

is decomposed into a sum of i independent terms or contributions that involve the breeding value of founders and the Mendelian sampling of nonfounders. The coefficient T_{ij} is the genetic contribution of an ancestor j to an individual i , called $r_{j(i)}$ by Woolliams et al. (1999). Woolliams et al. (1999) defined the long-term genetic contribution [$r_{j(i)}$, the same as T_{ij}] as the proportion of the genes in individual i transmitted from j . Using this definition, for a defined cohort (generation), the average of contributions $r_{i(j)}$ of each ancestor i across all individuals (descendants or not) j belonging to a defined cohort, defines its long-term contribution to that cohort, r_i (as per Howard et al., 2018), as

$$r_i = \frac{1}{n} \sum_{j=1}^n r_i^{(j)}. \quad [2]$$

The sum is computed over all individuals j ($j = 1, \dots, n$) born in the defined cohort, and n was the total number of animals in the defined cohort.

Equation [1] also holds for unbiased predictors, such as the BLUP (Henderson, 1973) of the breeding or additive genetic values ($\hat{\mathbf{u}}$; García-Cortés et al., 2008) as follows:

$$\hat{\mathbf{u}} = \mathbf{T}\hat{\boldsymbol{\phi}}, \quad [3]$$

where $\hat{\mathbf{u}}$ is a vector that contains the EBV, and the Mendelian samplings can be estimated as $\hat{\boldsymbol{\phi}} = \mathbf{T}^{-1}\hat{\mathbf{u}}$. García-Cortés et al. (2008) proposed to define a set of k partitions such that $\mathbf{P}_1 + \mathbf{P}_2 + \dots + \mathbf{P}_k = \mathbf{I}$. These partitions are based on groups or categories of interest; for instance, males and females. From this idea and using the Equation [3], we can write

$$\hat{\mathbf{u}} = \mathbf{T}(\mathbf{P}_1 + \mathbf{P}_2 + \dots + \mathbf{P}_k)\hat{\boldsymbol{\phi}} = \mathbf{TP}_1\hat{\boldsymbol{\phi}} + \mathbf{TP}_2\hat{\boldsymbol{\phi}} + \dots + \mathbf{TP}_k\hat{\boldsymbol{\phi}},$$

where \mathbf{P}_i (with $i = 1, \dots, k$) is a diagonal matrix containing zeros and ones, and it selects the corresponding columns of the \mathbf{T} matrix; in other words, the corresponding ancestor effects for a given $\hat{\mathbf{u}}$ (García-Cortés et al., 2008). Replacing the Mendelian sampling term by $\mathbf{T}^{-1}\hat{\mathbf{u}}$, we obtain

$$\hat{\mathbf{u}} = \mathbf{TP}_1\mathbf{T}^{-1}\hat{\mathbf{u}} + \mathbf{TP}_2\mathbf{T}^{-1}\hat{\mathbf{u}} + \dots + \mathbf{TP}_k\mathbf{T}^{-1}\hat{\mathbf{u}};$$

$$\hat{\mathbf{u}} = \hat{\mathbf{u}}_1 + \hat{\mathbf{u}}_2 + \dots + \hat{\mathbf{u}}_k.$$

Note that $\hat{\mathbf{u}}_k$ is a part of the breeding value contributed by the category k . From these partitions ($\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_k$), the partial genetic response can be obtained for a group or category. For example, the genetic gain can be partitioned using a category defined by sex, where \mathbf{P}_m and \mathbf{P}_f are the partition matrices defined for males and females, respectively. The matrix \mathbf{P}_m (or \mathbf{P}_f) is a diagonal matrix with ones in the male (female) positions and zeros in the female (male) positions. The method presented above enables the calculation of partial genetic trends directly from the EBV and allows us to inspect the contributions of each category to the genetic gain. Note that by doing this, we have split, for each individual, the parent average into contributions due to Mendelian sampling of their ancestors, which, in turn, then we assign into categories. This is why García-Cortés et al. (2008) consistently use the term “Mendelian sampling” for this partition.

Phenotypic and Pedigree Data

This study considered the breeding schemes of the BB, MTN, MTR, and LAC breeds until 2016, at which time genomic selection was introduced. Based on the 2021 official national genetic evaluation (single-step Genomic BLUP). In a single-step Genomic BLUP evaluation, the addition of genomic data to the evaluation can be seen as “more information,” and genomic EBV (**GEV**) can be seen as the EBV obtained at the same time but with genomic information (Macedo et al., 2020). Thus, it is possible to analyze the selection decisions in an individual and its ancestors based on GEV. The EBV or GEV for milk yield (liters per ewe per lactation) of animals born between 1985 and 2016 and their pedigree were used in the analyses (Table 1). By 1985 all breeds had ongoing, routine pedigree and milk yield recordings. The breed with the greatest population size is LAC, followed by MTR, BB, and MTN. Pedigree completeness of rams was very high: as these are elite animals, all AI rams have at least 2 parents and 4 grandparents known. Because 3% of LAC, 11% of MTR, 18% of BB, and 20% of MTN dams have unknown sires, the genetic evaluation model includes unknown parent groups (**UPG**; see Table 1), to take into account the different genetic levels of missing sires (Quaas, 1988). Use of UPG allows for realistic estimates of genetic trends. We did not attempt alternative definitions of UPG. Macedo et al. (2022) showed that an adequate genetic trend requires correct definition of UPG, which was the case here. Because EBV in models with UPG are not estimable functions, in order to obtain meaningful partitions, we shifted the EBV by a constant such that the average EBV of the first cohorts of animals was 0. To estimate animals’ Mendelian sampling, we discounted the solution of the UPG from its progeny; but otherwise, we did not include UPG solutions when presenting contributions or trends.

We decomposed the total genetic gain in 2 different manners: first, into contributions from males and females; second, into contributions from categories defined by different selection pathways. Thus, animals were classified into 5 categories (Table 2): (1) AI Males, the best males kept after progeny testing; (2) Discarded AI Males, the males discarded after progeny testing; (3) Natural Mating (**NM**) Males; (4) Dams of Males, which are the elite females chosen to generate AI Males; and (5) Dams of Females, those that were not chosen as elite females, because their estimated genetic merit was lower than that of Dams of Males. In dairy sheep selection schemes, there are females (recorded for milk yield and other traits) and males, which can be either AI or NM Males. Artificial insemination is used for spread of genetic improvement

Table 1. Number of animals in pedigree per breed

Breed ¹	Number of animals	Number of females	Number of males	Number of dams	Number of sires	Number of UPG ²
BB	145,047	142,614	2,433	66,077	2,249	32
MTN	100,764	99,328	1,436	43,822	1,353	40
MTR	489,098	480,157	8,941	230,632	8,409	40
LAC	1,474,543	1,452,032	22,511	721,567	17,346	17

¹BB = Basco-Béarnaise; MTN = Manech Tête Noire; MTR = Manech Tête Rousse; LAC = Lacaune.

²UPG = unknown parent groups.

and (until 2016) for selection of elite males by progeny testing. The AI Males produce 76% of the females, with the remaining 24% coming from NM Males (on average of all breeds). In turn, elite females are also selected to give birth to best males. All males are offspring of AI Males and AI-born females; in other words, there is no AI Male son (or grandson) of an NM Male. The categories of females and males without progeny (“Females No Progeny” and “Males No Progeny,” respectively) were present in the pedigree, but they were not analyzed, as their genes do not contribute to the genetic progress. Table 2 shows the number of animals per breed and defined category, as well as the proportion selected in the 4 classical pathways (sire of sire, sire of dam, dam of sire, and dam of dam) in the 4 breeds. Note that these categories describe correctly the pre-2016 scheme even if some (e.g., Discarded AI Males) are no longer valid in the current scheme, which is a genomic one with no formal progeny testing (although eventually daughters’ performances contribute greatly to the EBV of the AI Males).

Finally, to understand some of the results, we computed generation intervals (average age of the ancestor when its selected offspring was born) and family size per category of animals for the BB and LAC breeds. Table 3 shows the generation intervals and the family size per category of animals for BB (as representative of Western Pyrenees breeds) and LAC breeds. Generation intervals for LAC are lower (~3 yr) than the generation intervals for BB (~4 yr). Family size of the LAC AI Males is larger compared with BB. The numbers of

progenies of AI Males that are males and females are greater for LAC than for BB.

Analyzing Mendelian Samplings

Although pedigree and performance recording started earlier in some breeds (notably LAC), our analyses begin in 1985, when the dairy sheep selection schemes started to extensively use performance and pedigree recording.

The AlphaPart R package (Obšteter et al., 2021), available at <http://CRAN.R-project.org/package=AlphaPart>, which implements the method of García-Cortés et al. (2008), was used for the Mendelian sampling decomposition of EBV and genetic trends. The output of AlphaPart contains for each individual i the EBV (u_i), which is partitioned in 2 different manners. First, the EBV is split into the parent average plus the Mendelian sampling of the individual; in our notation, $\hat{u}_i = \widehat{PA}_i + \hat{\phi}_i = \frac{1}{2}\hat{u}_{sire(i)} + \frac{1}{2}\hat{u}_{dam(i)} + \hat{\phi}_i$. Second, the EBV is split into a series of values (EBV_“category” in the AlphaPart output) per each of the k categories, such that the sum across categories yields the EBV; in our notation, $\hat{u}_i = \hat{u}_{(i)1} + \hat{u}_{(i)2} + \dots + \hat{u}_{(i)k}$. In turn these $\hat{u}_{(i)1}, \hat{u}_{(i)2}, \dots, \hat{u}_{(i)k}$ are aggregated across (and not within) individuals into Mendelian sampling trends for each of the k categories (e.g., Figure 1). Note that the decomposition into trends due to Mendelian samplings sum to the overall genetic trend. In addition, it is of interest to obtain the genetic gain per year per category, in terms of Mendelian samplings. In practice this was carried

Table 2. Number of animals and selection proportions per breed and category

Breed ¹	Category							Selection proportion			
	AI Males	Discarded AI Males ²	Natural Mating Males	Dams of Males	Dams of Females	Females, No Progeny	Males, No Progeny	Sire of sire	Sire of dam	Dam of sire	Dam of dam
BB	633	866	750	1,941	64,136	76,537	184	0.47	0.66	0.26	0.94
MTN	402	507	444	1,183	42,639	55,506	83	0.46	0.65	0.48	0.89
MTR	2,843	2,923	2,643	6,979	223,653	249,525	532	0.34	0.50	0.35	0.90
LAC	8,841	6,136	2,369	17,916	703,651	730,465	5,165	0.06	0.14	0.32	0.91

¹BB = Basco-Béarnaise; MTN = Manech Tête Noire; MTR = Manech Tête Rousse; LAC = Lacaune.

²Discarded AI Males are the males discarded after progeny testing.

Table 3. Generation intervals (years) and number of offspring (males and females) per category¹

Category ²	BB				LAC			
	GIM	GIF	NPM	NPF	GIM	GIF	NPM	NPF
AI Males	4.3	3.1	8.3	86.2	3.9	3.1	13.8	109.0
Discarded AI Males	3.4	2.6	1.2	22.2	3.9	2.6	1.6	15.2
NM Males	3.2	2.7	1	4.4	3.5	2.9	1	4.5
DM	4.0	4.6	1.2	2.5	3.3	3.7	1.2	2.3
DF	—	4.2	—	1.8	—	3.4	—	1.9

¹BB = Basco-Béarnaise; LAC = Lacaune. GIM = generation interval of the individual when the male progeny was born; GIF = generation interval of the individual when the female progeny was born; NPM = number of progeny that are males; NPF = number of progeny that are females.

²NM = natural mating; DM = dam of males; DF = dam of females.

out after processing the output of AlphaPart: the contribution in terms of Mendelian sampling of year t in category j was obtained as the difference between yearly averages: $MSgain_{j(\text{year}=t)} = \hat{u}_{j(\text{year}=t)} - \hat{u}_{j(\text{year}=t-1)}$.

Analyzing Long-Term Genetic Contribution

Long-term genetic contributions (r) were computed using self-made software in BB (again, as representative of Western Pyrenees breeds) and LAC breeds. This concept (and value) of long-term genetic contribution represents the contribution of an individual’s Mendelian sampling term to the long-term genetic pool (Woolliams, 2007). Note that the gene pool of the population has contributions from all ancestors and not just the founders.

To determine the value of r_i for individual i , a scale of pseudo-generations (hereinafter “generations,” for simplicity) was used to group the individuals depending on their year of birth. In dairy sheep, the generation interval is around 4 years. Base generation (G_0) consisted of individuals born from 1985 to 1988. Seven more generations were defined, with the last generation (G_7) consisting of individuals born between 2013 and 2016. The contributions r_i are highly left-skewed, and, for ease of presentation, we use histograms or boxplots of $\log_{10}(r_i)$ (i.e., -2 means 0.01).

In addition to r_i , a selection score (x_i) was determined for all individuals (Howard et al., 2018). If the individual had offspring, $x_i = 1$, and $x_i = 0$ otherwise. Note that $x_i = 0$ implies $r_i = 0$, and therefore 3 categories can be established: unselected ($x_i = 0, r_i = 0$), selected but with no contribution to the defined generation ($x_i = 1, r_i = 0$), and selected with contribution to the defined generation ($x_i = 1, r_i > 0$). The case ($x_i = 1, r_i = 0$) occurs when individual i produces offspring in the next generation but its descendants at some point do not produce further offspring in the following generations; its genetic line goes extinct. Table 4 shows the number of individuals in each category per breed (BB

and LAC) and sex. However, the scores were not used in our further analyses for males, because males that are recorded into the pedigree files are live, breeding males (AI or NM rams) and therefore selected, whereas females are recorded if they have at least 1 lactation and may then be selected or not. Thus, pedigree recording is highly biased in males toward selected animals. For females, 2 further analyses, detailed below, were performed with scores.

The relationships between long-term genetic contributions, estimated Mendelian sampling terms ($\hat{\phi}$), and EBV were also studied. First, we studied the importance of parent average $\widehat{PA}_i = (EBV_i - \hat{\phi}_i)$ or $\hat{\phi}_i$ as source of selective advantage. Second, we examined the maintenance of contributions over time for selected individual. We used bivariate regressions to determine whether \widehat{PA}_i (corrected by year of birth to compare properly across successive years) or $\hat{\phi}_i$ were involved in the selection of an individual ($x_i = 1$ for individuals whose offspring were kept) and in the persistence of the contribution over time [which parents made a nonzero contribution ($r_i > 0$) to the population at the last, G_7 , pseudo-generation].

A generalized linear model with a binomial distribution and a logistic link function was fitted to assess the weight placed on \widehat{PA}_i versus $\hat{\phi}_i$. Let the probability of selection for individual i be $\mu_i = E[x_i]$, then if $f(\cdot)$ is the logistic link function, the following model was fitted:

Table 4. Number of females and males in each breed¹ depending on the selection score (x_i) and the long-term genetic contribution (r_i) to generation 7

Constraint		BB		LAC	
x_i	r_i	Females	Males	Females	Males
≥ 0	≥ 0	142,614	2,433	1,452,032	22,511
> 0	≥ 0	66,077	2,249	721,567	17,346
> 0	> 0	31,567	1,882	308,514	15,419

¹BB = Basco-Béarnaise; LAC = Lacaune.

$f^{-1}(\mu_i) = \alpha + \beta yr + \beta_{PA} \widehat{PA}_i + \beta_{\phi} \widehat{\phi}_i + \varepsilon_i$, where α is an intercept, yr is the year of birth, and β_{PA} and β_{ϕ} are the respective regression coefficients on \widehat{PA}_i and $\widehat{\phi}_i$. The odds ratios were computed as $\left[\exp(\widehat{\beta})\right]$. This analysis was conducted using the “binomial” option in the “glm” package of R.

RESULTS

Between 1985 and 2016, the total genetic progress for milk yield (estimated from the regression of the EBV on the year of birth) was 4.6 l/yr, 3.3 l/yr, 2.9 l/yr, and 2.7 l/yr for the LAC, MTR, MTN, and BB breeds, respectively. This implies an increment of more than 100 L for all the evaluated breeds between 1985 and 2016. The use of AI differs among breeds. In LAC flocks, AI rate reaches 85% (420 progeny-tested males by year), whereas this rate is 55% (with 30–130 males tested by year) in BB, MTN, and MTR (Astruc et al., 2002). This leads to different selection proportions among categories, as observed in Table 2; most notably, LAC has greater selection pressures in the “sire of” pathways. For all the following results, the genetic progress was stable for all 4 breeds only after approximately 1992, so some fluctuations in the results are apparent in the first years.

Figure 1 shows the decompositions of genetic trends of milk yield, in Mendelian sampling terms, contributed by males and contributed by females for the 4 breeds, in addition to the overall genetic trend. It can be verified that the genetic trend is the sum of the trends of Mendelian samplings contributed by males and females. Looking at the trends of contribution of Mendelian sampling by sex, the highest difference among breeds was between LAC and other breeds, as the contribution of the Mendelian sampling terms of females was greater than the contributions of Mendelian sampling terms of males in the LAC breed across the entire period. In 2016, LAC females and males’ Mendelian sampling contributions were (at the end of the genetic trends in 2016) 101 (69%) and 45 L (31%), respectively. The Western Pyrenees dairy sheep breeds (BB, MTN, and MTR) showed quite similar shapes of trends, where in all cases the Mendelian sampling contributions cumulated around 50% for females and males in 2016.

Figure 2 presents the decomposition of genetic trends of milk yield to the contributions of each category previously defined, for each breed. The categories’ contributions to the total genetic gain in the LAC breed differed from those of the Pyrenees dairy sheep breeds. In LAC, the contributions of the AI Males, Dams of Males, and Dams of Females were of the same order.

Looking at the year 2016, AI Males contributed 33% of the total genetic gain, and the females’ categories contributed to the total genetic gain as follows: dam of males contributed 39%, and dam of females contributed 27%. In the Western Pyrenees breeds, the AI Males’ category made the highest contribution to the total genetic gain at the end in 2016, with 47%, 47%, and 49% for the BB, MTN, and MTR breeds, respectively. The contribution to the total genetic gain by AI Males was followed by that made by Dams of Males and then by the contribution by Dams of Females. Dams of Males contributed 36%, 35%, and 39% to the total genetic gain for the BB, MTN, and MTR breeds, respectively. Finally, the dam of females category contributed 19%, 18%, and 10% to the total genetic gain in the BB, MTN, and MTR breeds, respectively. As expected, the categories that Discarded AI Males and NM Males did not contribute to the genetic gain in Mendelian sampling terms.

Figure 3 shows the average gains in terms of Mendelian samplings (selected for) by year of birth of the categories dam of males, dam of females, and AI Males. For example, in the AI Males category, the average contribution due to Mendelian sampling in 2001 is the difference between the point at the year 2001 in the trend of Mendelian samplings for AI Males in Figure 2, minus the value from the year 2000. The average contribution due to Mendelian sampling for the 2 categories of dams (dam of males and dam of females) varied between 0 and 2 L for the Western Pyrenees breeds and between 1 and 3 L for the LAC breed. The generated gain was relatively constant over time. For AI Males, the average was between –1 and 4 L. Depending on the breed, males contributed equally to or more than females, although their contribution was more irregular because at any point the number of males included in the averages shown in Figure 3 is smaller than the number of females. Similarly, less dispersion occurred in LAC compared with other breeds, due to a greater number of animals included in the averages.

The average contribution to gain due to Mendelian sampling terms for the 4 breeds varied between –1 and 1 L for Discarded AI Males (randomly, as their number is small), and it is around 0 for NM Males (with a larger number). The Discarded AI Males are not “good enough” after progeny testing, and they stop contributing their genes. Accordingly, after initial selection based on parent average, NM Males were not further selected. Thus, no contribution was obtained through their progeny to the selection response.

Now we present results on genetic contributions (measured on a scale from 0 to 1) for the BB and LAC breeds. Figure 4 shows the distribution of the \log_{10} of the contributions of G_0 to the following generations.

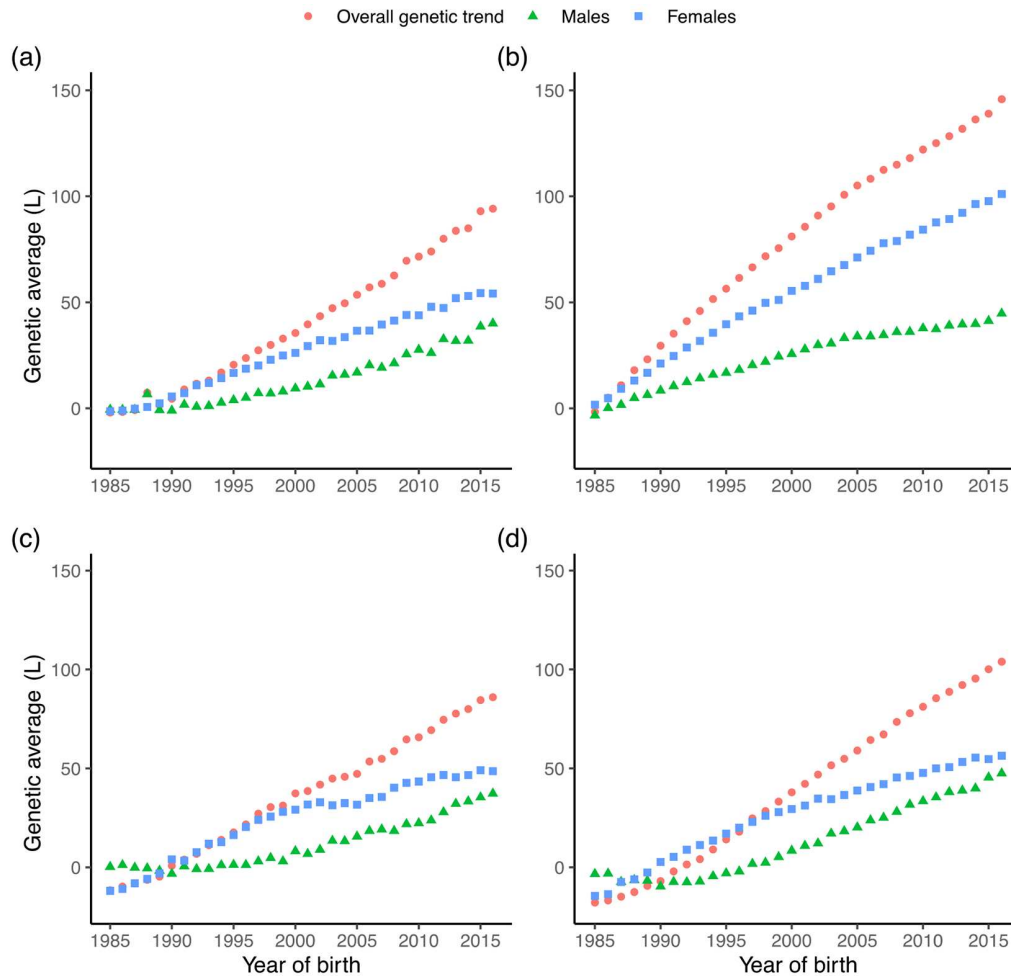


Figure 1. Total genetic trends for milk yield (L) and partitioning into contributions due to Mendelian samplings by sex. Breeds: (a) Basco-Béarnaise, (b) Lacaune, (c) Manech Tête Noire, (d) Manech Tête Rousse.

For instance, in G_7 , the boxplot describes the final contributions from individuals in G_0 (cohort 1985–1988) to individuals in G_7 (cohort 2013–2016), as obtained from Equation [2] (i.e., their average contribution ~ 28 yr later). Males had greater average individual genetic contributions to the genetic pool of the population than females, in both breeds, and that is because of males' larger numbers of offspring because of AI. Greater contributions were observed for BB, because the number of animals at each generation is around 10 times lower than the number of LAC-contributed animals. For example, in the last generation (G_7), the total number of individuals is 21,656 animals for BB and 185,247 for LAC. Therefore contributions in LAC tend to average more rapidly with time.

Genetic contributions of different categories (AI Males, Dams of Males, and Dams of Females) are

presented for both breeds in Supplemental Figure S1 (<https://doi.org/10.6084/m9.figshare.22325413.v1>, Antonios et al., 2023a). In the LAC and BB breeds, AI Males rated the highest average individual genetic contributions among all categories, followed by the genetic contributions of Dams of Males and then by the genetic contributions of Dams of Females. As mentioned above, the number of animals in the last generation (G_7) in LAC is higher (10 times) than in the BB breed.

The distribution of the \log_{10} of contributions r_i is shown in Figure 5 (with different scales on the Y-axis). The contributions of the males are more shifted to the right (the mean of r_i is approximately 10^{-4} for BB males and 10^{-5} for LAC males), which indicates higher contributions for males than for females, where the females' mean of r_i is approximately 10^{-5} for BB and 10^{-6} for LAC. In the real 0-to-1 scale, the maximal

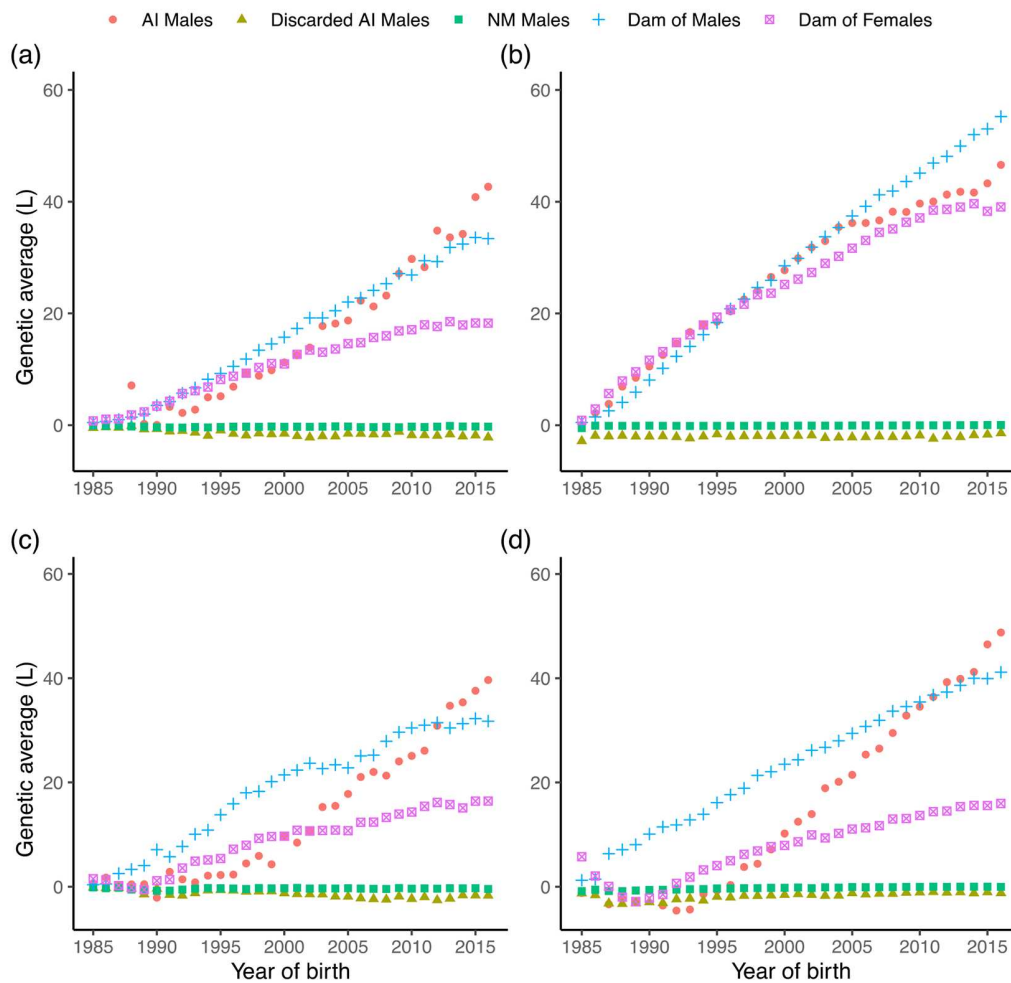


Figure 2. Yearly partition of the genetic trends for milk yield into contributions due to Mendelian samplings by categories. Breeds: (a) Basco-Béarnaise, (b) Lacaune, (c) Manech Tête Noire, (d) Manech Tête Rousse. NM = natural mating.

contributions of the males are approximately 0.09 for BB and 0.02 for LAC. For the females, the maximal contributions are approximately 0.03 and 0.01 for BB and LAC, respectively.

Long-term contributions (\log_{10} of the contributions) were plotted against the EBV of the first-generation individuals to their EBV in the BB and LAC breeds (see Supplemental Figure S2; <https://doi.org/10.6084/m9.figshare.21973358.v1>, Antonios et al., 2023b) and followed the expected relationship that individuals with highest EBV intrageneration will have more offspring, increasing their long-term contribution.

The logistic regression of the selection score x_i on \widehat{PA}_i and $\hat{\phi}_i$ for females showed that both \widehat{PA}_i and $\hat{\phi}_i$ were significant factors ($P < 0.001$) in promoting selection as a parent. Estimates of odds ratio (95% CI) for Mendelian samplings were 1.0309 (1.0299–1.0317) for BB and

1.0203 (1.0200–1.0205) for LAC and, for parent average, were 1.0154 (1.0146–1.0162) for BB and 1.0149 (1.0147–1.0151) for LAC. Odds ratio for Mendelian sampling (\widehat{PA}_i) is greater than \widehat{PA}_i (1.0154 for BB and 1.0149 for LAC). This means that for a one-unit (1-L) increase in $\hat{\phi}_i$, the odds of being selected increase 3% and 2% in BB and LAC, respectively. In males, it was not possible to perform this analysis because males registered in the pedigree are registered because they have offspring, and therefore have already been selected ($x_i = 1$), whereas the others go to the slaughterhouse.

Among selected animals, we analyzed which animals contributed over time to G_7 (if $r_i > 0$ in G_7 , then the binomial variable is equal to 1; otherwise, it is 0) in both sexes. Odds ratios from the logistic regression on \widehat{PA}_i and $\hat{\phi}_i$ are shown in Table 5. In both sexes, \widehat{PA}_i and

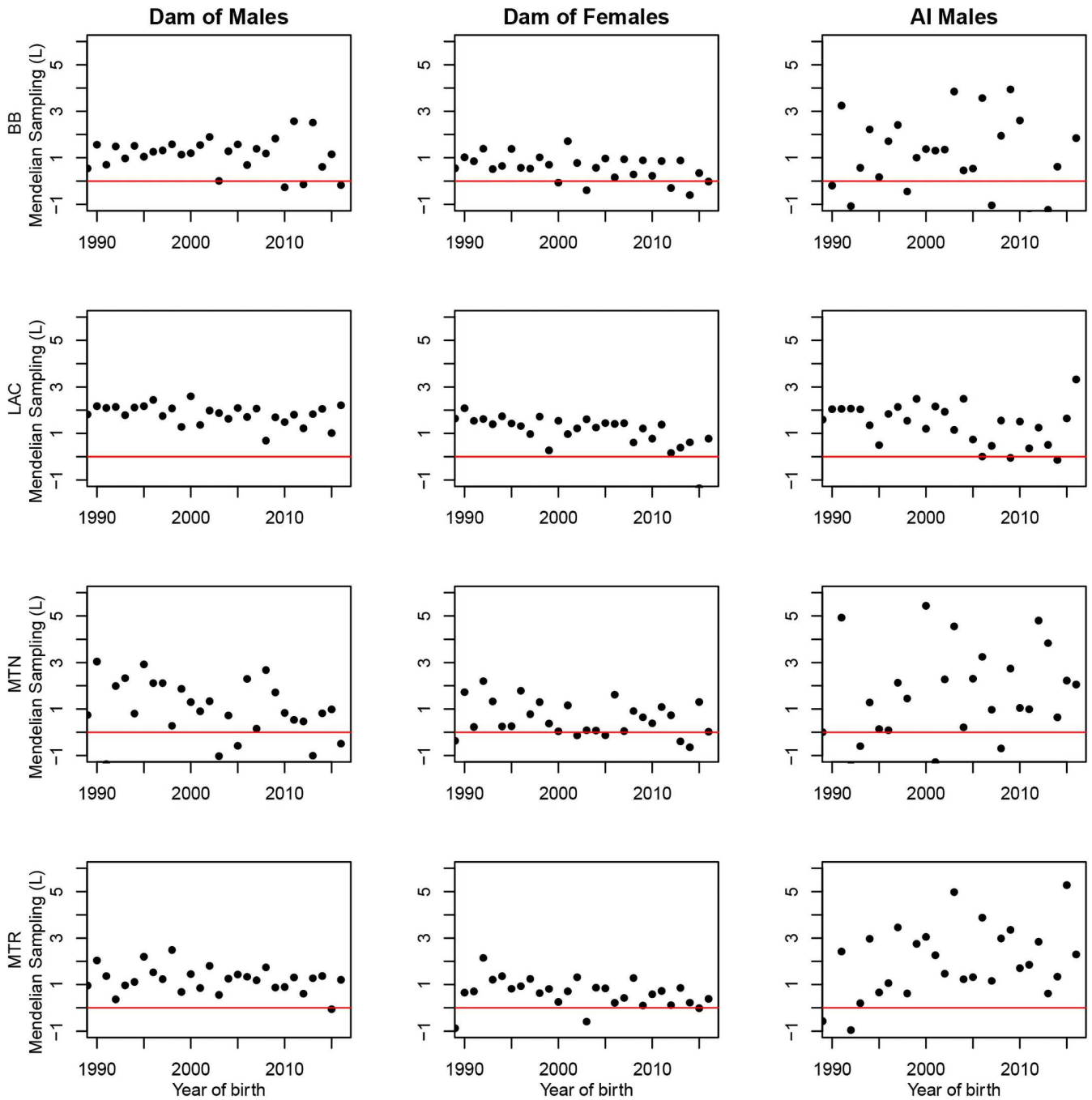


Figure 3. Average contribution due to Mendelian sampling for milk yield by year of birth of the categories Dams of Males, Dams of Females, and AI Males. The red horizontal lines indicate the zero value on the contribution due to Mendelian sampling, which occurs when no additional gain is contributed to the next generation. Breeds: BB = Basco-Béarnaise; LAC = Lacaune; MTN = Manech Tête Noire; MTR = Manech Tête Rousse.

$\hat{\phi}_i$ were positively associated with maintaining a contribution across generations conditionally on being initially selected as parent. In BB, the effect of PA_i is greater than the effect of $\hat{\phi}_i$ in both sexes. However, the

opposite was observed in LAC. This result is in agreement with Figure 1. In LAC, $\hat{\phi}_i$ increases more in females than in males, and the odds of maintaining the contributions over time are 1% higher for LAC-selected females relative to males.

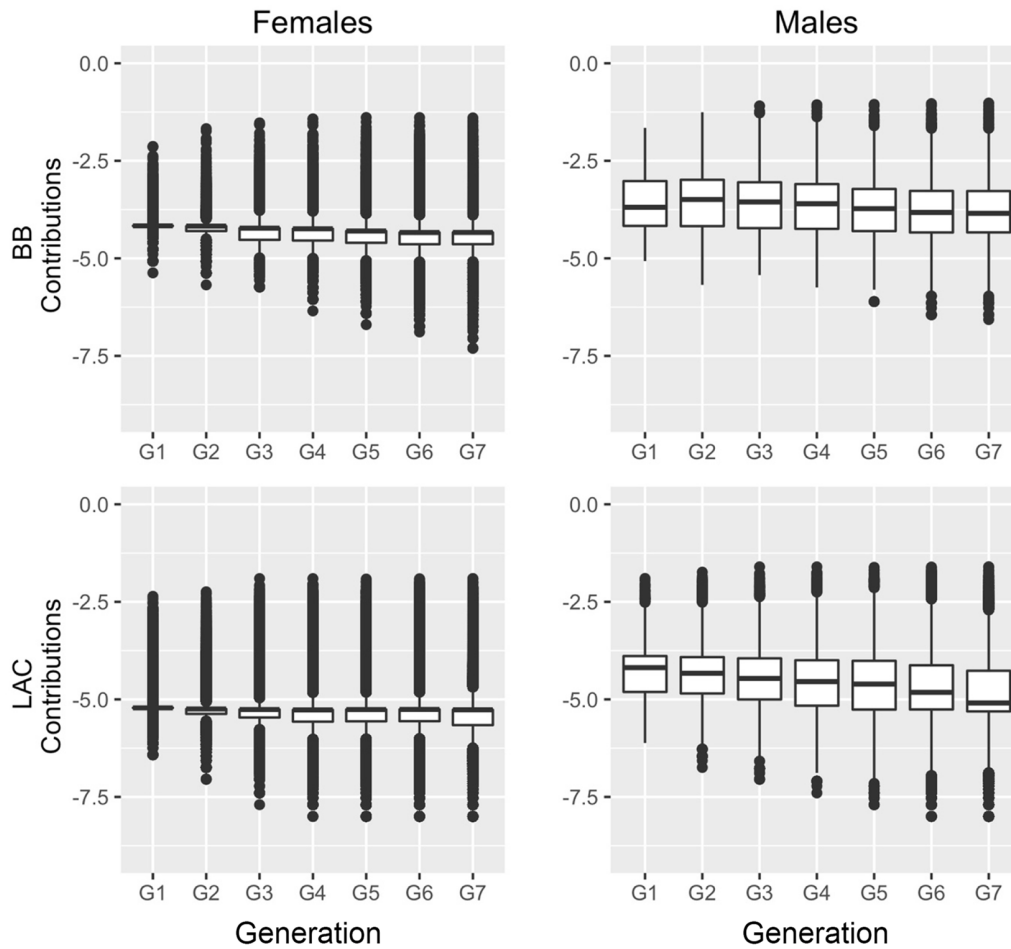


Figure 4. Boxplots of \log_{10} of contributions of males and females per generation from G_0 (base generation) to G_7 (the last generation) in Basco-Béarnaise (BB) and Lacaune (LAC) breeds. Lower whisker = minimum observation above 25% quantile $- 1.5 \times$ interquartile range (IQR); lower line of the box = 25% quantile (Q1); horizontal line inside the box = median; upper line of the box = 75% quantile (Q3); upper whisker = maximum observation below 75% quantile $+ 1.5 \times$ IQR; dots outside the box = outliers, values which either exceed $Q3 + 1.5 \text{ IQR}$ or fall below $Q1 - 1.5 \text{ IQR}$.

DISCUSSION

The decomposition of genetic trend into Mendelian sampling contributions by sex or categories showed that females play an important role in selection schemes. Across generations, females contribute to genetic progress as much as the males (BB, MTN, and MTR) and even more (LAC) to the final genetic gain, as shown in Figure 1. We interpret that, even if males are more heavily selected, half of their genetics come from their dam, whose positive Mendelian samplings are eventually selected and spread (often through AI Males) throughout the population. Indeed, this was observed by García-Cortés et al. (2008).

Most females are Dams of Males or Dams of Females (Figure 2). Dams of Males are progeny of AI Males, whereas Dams of Females can be progeny of AI or NM

Males. Dams of Males are expected to contribute more descendants than Dams of Females because only the sons of Dams of Males are selected. However, a dam of females can be the maternal grandmother of an AI male. The long-term contribution of an individual is given by the sum over its male and female selected offspring (Woolliams et al., 1999). The Mendelian sampling term in Dams of Males is estimated from their milk records, and also from granddaughter records when a son is selected as an AI male. In Dams of Females, their own phenotype, the daughters' milk records when they are selected as Dams of Males, and great-granddaughters' records (if a dam of females is the grandmother of an AI male) contribute to estimation of Mendelian sampling term. Accurate estimates of Mendelian samplings occur when the animal has a large number of progeny (e.g., Dams of Males) or even grand-progeny (e.g., Dams of

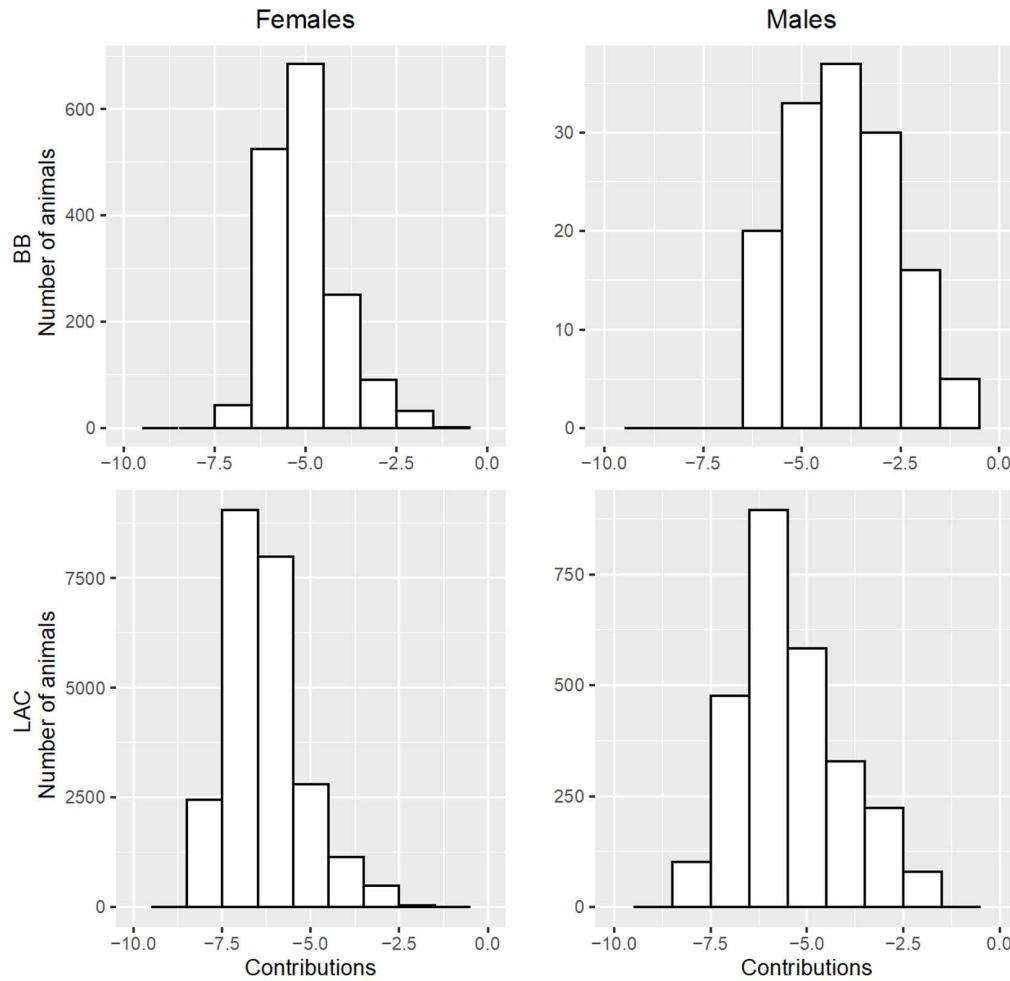


Figure 5. Histograms of the \log_{10} of the genetic contributions of males and females to the G_7 (last generation) by sex in Basco-Béarnaise (BB) and Lacaune (LAC) breeds.

Males and Dams of Females). This shows that dams are as important as AI Males in driving changes in milk yield.

The decomposition of genetic trend was different in the LAC breed compared with the Western Pyrenees

breeds. The difference between female and male contribution is greater in the LAC breed (Figure 1). In LAC, the fact that both categories of dams contribute to genetic gain in the same order as AI Males may be due to a shorter generation interval (~ 3 yr in LAC and ~ 4 yr in BB in average for dams and AI Males, as observed from averages of values in Table 3) and to the intense use of AI in the nucleus flock (greater numbers of female progenies for the AI Males in LAC than in BB, Table 3). In LAC, AI Males produced a greater number of female offspring than in the other breeds (Table 3). Thus, the EBV of LAC AI Males is more accurate, and the selection for originality (Mendelian sampling) will be more efficient. Additionally, the greater offspring numbers of LAC AI Males make possible a more intense within-family selection, as shown in Table 3. The difference in the relative importance of the Mendelian sampling and \widehat{PA}_i between breeds across generations could

Table 5. Estimates of odds ratios, with 95% CI, from the bivariate logistic regression of maintenance of nonzero contributions ($r_i > 0 \mid x_i = 1$) on estimated parent average (\widehat{PA}) and estimated Mendelian sampling terms ($\hat{\phi}$) for females and males per breed¹

Variable	BB	LAC
Females		
\widehat{PA}	1.0334 (1.0319–1.0348)	1.0057 (1.0054–1.0061)
$\hat{\phi}$	1.0303 (1.0288–1.0317)	1.0202 (1.0199–1.0205)
Males		
\widehat{PA}	1.0496 (1.0372–1.0622)	1.0035 (1.0000–1.0069)
$\hat{\phi}$	1.0061 (0.9995–1.0131)	1.0134 (1.0109–1.0158)

¹BB = Basco-Béarnaise; LAC = Lacaune.

reflect different selection strategies in practice. Mendelian sampling relies on within-family selection, whereas parent average is more related to across-family variation. A larger family size in LAC than in BB is in agreement with a greater importance of within-family selection in LAC, which could be because it is possible to exercise greater selection intensities within family when families are larger, as is the case.

To improve a population, we need certain animals to be better than their parents. For this, the originality of the animals with respect to their parents (i.e., the Mendelian sampling) has to be captured. The latter is estimated from different sources of information: (1) own phenotype (milk yield for females, collected in the selection scheme), (2) phenotypes of the progeny of males (progeny test; i.e., in the AI Males in these schemes), or (3) by genomic evaluation (not studied in this work, because genomic selection was not yet implemented in 2016). Although the precision in capturing Mendelian sampling is greater in males (by progeny testing), the final genetic progress is largely given by the originality (or genetic variability) of the females, even if their prediction (by own phenotype) is less accurate and their progeny number is smaller. This is true in this study and possibly in most ruminant selection schemes based on intensive use of AI. Note that a selected female can be the mother or grandmother of an AI male and consequently have a large number of grand-offspring. Moreover, because the number of eligible females is very high, the selection intensity to be a dam of males can be very high. Therefore, fine partitioning in Mendelian sampling trends can be interpreted as females being a large and easily accessible reservoir of genetic variability for selection. Still, AI Males are crucially important because they allow the spread of best genetics in the population (Figure 4 and Supplemental Figure S1) and genetically connect the herds. The smaller the number of the animals, the greater the contributions of those animals will be for future years. Additionally, males had greater contributions than females in both breeds because the AI Males, despite their smaller number compared with the dams, contribute more than dams to the gene pool of the population due to their larger number of offspring.

CONCLUSIONS

This study decomposed the genetic gain in Mendelian samplings by categories of animals defined by sex and by selection pathways, and explored long-term genetic contributions. This allowed identification of the different contributions of categories of individuals, and better understanding of the selection scheme. Dams contribute largely to the final genetic gain by their orig-

inality, whereas AI males are crucially important in the spread of the best genetics in the population and make a proportionally greater contribution to the gene pool of the population. The relative weights of parent average and Mendelian sampling in selection decisions seem to depend on family sizes. In this study, the Mendelian sampling term was observed to be the most important factor determining the selection of a female to become dam and maintaining genetic contributions over time.

ACKNOWLEDGMENTS

The authors thank the breeders who provided the data and the anonymous referees for useful comments on the manuscript. This study received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 772787 (SMARTER), and project ARDI (grant agreement EFA 208/16; Derio [Bizkaia], Spain) from POCTEFA funds. The GenoToul bioinformatics platform Toulouse Midi-Pyrenees provided computing and storage resources. The authors have not stated any conflicts of interest.

REFERENCES

- Antonios, S., A. Legarra, R. Pong-Wong, J.-M. Astruc, S. T. Rodríguez-Ramilo, and Z. G. Vitezica. 2023a. Supplemental Figure S1. Boxplot of log10 of contributions of AI Males, Dam of Males and Dam of Females per generation from G0 (base generation) to G7 (the last generation) in Basco-Béarnaise (BB) and Lacaune (LAC) breeds. figshare. Journal contribution. <https://doi.org/10.6084/m9.figshare.22325413.v1>.
- Antonios, S., A. Legarra, R. Pong-Wong, J.-M. Astruc, S. T. Rodríguez-Ramilo, and Z. G. Vitezica. 2023b. Supplemental Figure S2. Long-term contribution (log10 of the contributions) plotted against EBV of the first-generation (G0) individuals in Basco-Béarnaise (BB) and Lacaune (LAC) breeds. figshare. Journal contribution. <https://doi.org/10.6084/m9.figshare.21973358.v1>.
- Astruc, J. M., F. Barillet, A. Barbat, V. Clement, and D. Boichard. 2002. Genetic evaluation of dairy sheep in France. Page 45 in Proc. 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France. World Congress on Genetics Applied to Livestock Production.
- Astruc, J.-M., D. Buisson, H. Larroque, and G. Lagriffoul. 2018. New breeding goals and role of genomics on adaptation and resilience traits in French dairy sheep. Page 705 in Proc. 69th Annual Meeting of the European Federation of Animal Science, Dubrovnik, Croatia. Wageningen Academic Publishers.
- Bijma, P. 2000. Long-term genetic contributions. Prediction of rates of inbreeding and genetic gain in selected populations. PhD thesis. Wageningen University, the Netherlands.
- García-Cortés, L. A., J. C. Martínez-Ávila, and M. A. Toro. 2008. Partition of the genetic trend to validate multiple selection decisions. *Animal* 2:821–824. <https://doi.org/10.1017/S175173110800205X>.
- Gorjanc, G., K. Potočnik, L. A. García-Cortés, J. Jakobsen, and J. Dürr. 2011. Partitioning of international genetic trends by origin in Brown Swiss bulls. In Proc. 44th Interbull Meeting, Cork, Ireland. Interbull Bulletin, 81–6.
- Henderson, C. R. 1973. Sire evaluation and genetic trends. *J. Anim. Sci.* 1973(Symposium):10–41. <https://doi.org/10.1093/ansci/1973.Symposium.10>.
- Howard, D. M., R. Pong-Wong, P. W. Knap, V. D. Kremer, and J. A. Woolliams. 2018. Selective advantage of implementing optimal contributions selection and timescales for the convergence of long-

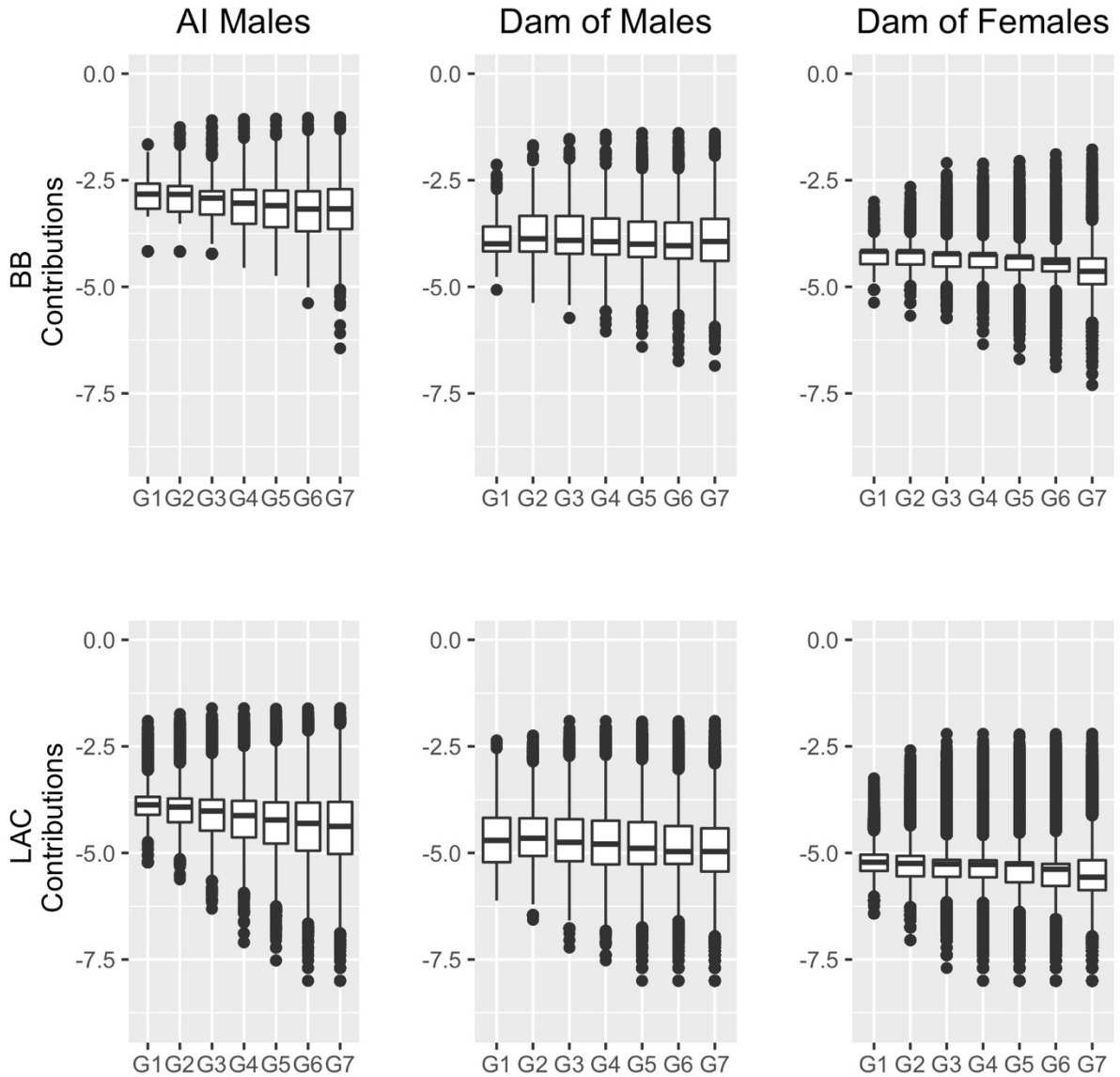
- term genetic contributions. *Genet. Sel. Evol.* 50:24. <https://doi.org/10.1186/s12711-018-0392-z>.
- Macedo, F. L., J. M. Astruc, T. H. E. Meuwissen, and A. Legarra. 2022. Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions. *J. Dairy Sci.* 105:2439–2452. <https://doi.org/10.3168/jds.2021-20860>.
- Macedo, F. L., O. F. Christensen, J.-M. Astruc, I. Aguilar, Y. Masuda, and A. Legarra. 2020. Bias and accuracy of dairy sheep evaluations using BLUP and SSGBLUP with metafounders and unknown parent groups. *Genet. Sel. Evol.* 52:47. <https://doi.org/10.1186/s12711-020-00567-1>.
- Obšteter, J., J. Holl, J. M. Hickey, and G. Gorjanc. 2021. AlphaPart—R implementation of the method for partitioning genetic trends. *Genet. Sel. Evol.* 53:30. <https://doi.org/10.1186/s12711-021-00600-x>.
- Quaas, R. L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71:91–98. [https://doi.org/10.1016/S0022-0302\(88\)79986-5](https://doi.org/10.1016/S0022-0302(88)79986-5).
- Škorput, D., G. Gorjanc, A. Kasap, and Z. Luković. 2015. Partition of genetic trends by origin in Landrace and Large-White pigs. *Animal* 9:1605–1609. <https://doi.org/10.1017/S1751731115001056>.
- Thompson, R. 1979. Sire evaluation. *Biometrics* 35:339–353. <https://doi.org/10.2307/2529955>.
- Woolliams, J. 2007. Genetic contributions and inbreeding. Pages 147–165 in *Utilisation and Conservation of Farm Animal Genetic Resources*. K. Oldenbroek, ed. Wageningen Academic Publishers.
- Woolliams, J., and R. Thompson. 1994. A theory of genetic contributions. Pages 127–134 in *Proc. 5th World Congress on Genetics Applied to Livestock Production*, Guelph, Canada. World Congress on Genetics Applied to Livestock Production.
- 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. <https://doi.org/10.1093/genetics/153.2.1009>.
- Woolliams, J. A., N. R. Wray, and R. Thompson. 1993. Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genet. Res.* 62:231–242. <https://doi.org/10.1017/S0016672300031943>.

ORCID

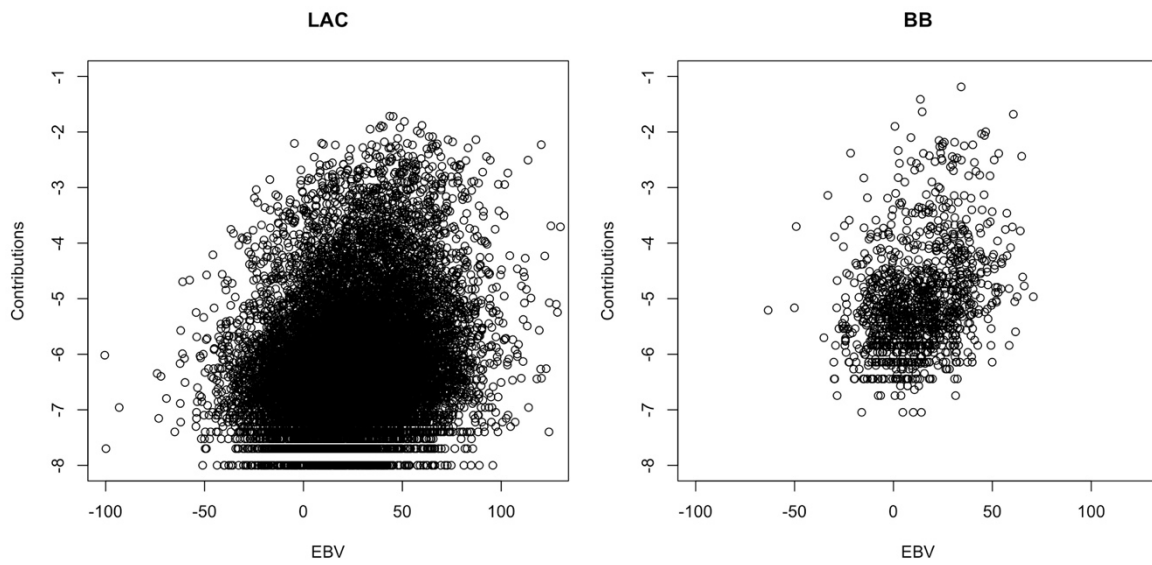
- S. Antonios  <https://orcid.org/0000-0002-2265-5118>
A. Legarra  <https://orcid.org/0000-0001-8893-7620>
J. M. Astruc  <https://orcid.org/0000-0001-7685-1301>
S. T. Rodríguez-Ramilo  <https://orcid.org/0000-0001-7150-0692>
Z. G. Vitezica  <https://orcid.org/0000-0002-3162-7258>

2.1 Appendix

Supplemental material



Supplemental Figure S1. Boxplot of log₁₀ of contributions of AI Males, Dam of Males and Dam of Females per generation from G0 (base generation) to G7 (the last generation) in Basco-Béarnaise (BB) and Lacaune (LAC) breeds.



Supplemental Figure S2. Long-term contribution (\log_{10} of the contributions) plotted against EBV of the first-generation (G0) individuals in Basco-Béarnaise (BB) and Lacaune (LAC) breeds.

Chapter 3

Genetic inbreeding load and its prediction in French dairy sheep

S. Antonios^{1*}, S.T. Rodríguez-Ramilo¹, A. Legarra², J.M. Astruc³, L. Varona⁴ and Z.G. Vitezica¹

¹ GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France

² CDCB, 4201 Northview Drive, 20716 Bowie MD, USA

³ Institut de l'Élevage, 31321 Castanet Tolosan, France

⁴ Universidad de Zaragoza, Instituto Agroalimentario de Aragón (IA2), 50013 Zaragoza, Spain

Genetic Selection Evolution (submitted on the 8th of July 2024)

Corresponding author: S. Antonios

INRAE, UMR 1388 GenPhySE

24 Chemin de Borde Rouge

31326 Castanet Tolosan Cedex, France

Phone: + 33 7 55 29 60 96

E-mail: simona.antonios@inrae.fr

Abstract

Background

The magnitude of inbreeding depression depends on the hidden (recessive) inbreeding load among ancestors. However, animals carry different alleles at potentially deleterious loci and therefore there is individual variability of this inbreeding load. Estimation of the additive genetic value for inbreeding load is possible using a decomposition of inbreeding in partial inbreeding components. Both the magnitude of variation in partial inbreeding components and the additive genetic variance of inbreeding loads are largely unknown. The objectives of this work were three. First, we showed that inbreeding load can be expressed as an additive genetic effect based on substitution effect under non-random matings. Second, we calculated and described the partial inbreeding coefficients in three French dairy sheep populations (Basco-Béarnaise, Manech Tête Noire and Manech Tête Rousse). Third, we included these coefficients in a mixed model as random regression covariates, to predict genetic variance and breeding values of the inbreeding load for milk yield in the same breeds.

Results

Pedigrees included 190,276, 166,028 and 633,655 animals of Basco-Béarnaise, Manech Tête Noire and Manech Tête Rousse, respectively, born between 1985 and 2021. A fraction of 99.1% of the partial inbreeding coefficients were lower than 0.01 in all breeds, meaning that inbreeding is created in loops that go several generations back. Less than 5% ancestors generate inbreeding, because mating is essentially between unrelated individuals. Inbreeding load estimations involved 658,731, 541,180 and 2,168,454 records of yearly milk yield from 178,123, 151,863 and 596,586 females in Basco-Béarnaise, Manech Tête Noire and Manech Tête Rousse, respectively. Adding the inbreeding load effect to the model improved the fitting (values of the statistic Likelihood Ratio Test between 132 to 383) for milk yield in the three breeds. The inbreeding load variances were equal to 11,804 and 9435 liters squared of milk yield for a fully inbred (100%) descendant in Manech Tête Noire and Manech Tête Rousse. In Basco-Béarnaise, the estimate of the inbreeding load variance was zero. The correlations between additive genetic and inbreeding load effects were -0.09, -0.08 and -0.12 in Basco-Béarnaise, Manech Tête Noire and Manech Tête Rousse.

Conclusions

The decomposition of inbreeding in partial coefficients in these populations shows that inbreeding is mostly due to several small contributions of ancestors going back several generations, which is according to the policy of avoiding close matings. There is variation of inbreeding load among animals, although its magnitude does not seem enough to warrant selection based on this criterion.

Background

Inbreeding load is the fraction of the mutation load that is due to hidden recessive alleles in heterozygous state. This load, when exposed by inbreeding, is responsible for inbreeding depression, the decrease in performance and fitness in inbred individuals [1].

Inbreeding depression is thought to be due to the presence of recessive alleles in populations or from the reduction of heterozygous loci under overdominance. Usually, inbreeding depression is more expected in fitness traits (e.g. fertility) than in traits less related to fitness (e.g. milk yield). However, there is evidence that inbreeding depression can occur in any trait [2]. For fitness traits, inbreeding depression is mainly endorsed to recessive deleterious mutations. However, for traits under directional selection such as milk yield (for increasing the mean), dominance deviation effects are on average favourable; and inbreeding depression is due to the reduced expression of dominance effects by an increase in homozygosity [2].

In livestock, inbreeding load can vary among founders, moreover if the founder families were exposed to different selection pressures on deleterious alleles [3]. Inbreeding load of individuals can be predicted in the same manner that we do for additive genetic values based on linear models [4, 5]. However, previous authors have never expressed inbreeding load in terms of simple locus effects, e.g. as a substitution effect. The inbreeding load of individuals is a heritable additive trait that is only expressed when inbreeding occurs in their offspring [5] and that can have a favorable or unfavorable effect of the studied trait (e.g. milk yield) [6]. For instance, one could in principle find out if different individuals carry different inbreeding load by producing e.g. equally inbred descendance (say mating sires to their daughters) and comparing descendants across sires. In complex pedigrees this becomes more complex because

each individual possesses parts of inbreeding coming potentially from different ancestors. The fractions attributed to each ancestor can be computed using pedigree by the Mendelian decomposition of inbreeding which traces back the specific ancestral paths through which the identical by descent (IBD) alleles are inherited. Using these fractions, a linear model can predict the inbreeding load of the individuals [4, 5] in the same manner that we do for additive genetic values.

The objectives of this work are three. First, we showed that the inbreeding load can be expressed as a genetic additive effect based on substitution effect under non-random matings. Second, we computed and described the partial inbreeding coefficients due to ancestors in 3 French dairy sheep breeds: Basco-Béarnaise (BB), Manech Tête Noire (MTN) and Manech Tête Rousse (MTR). Third, we used these partial inbreeding coefficients as covariates in a random regression mixed model to estimate genetic variance and breeding values of the inbreeding load for milk yield in the three breeds.

Theory

Under the assumption of random mating, the substitution effect of a gene is the regression of genotypic values on gene content. However, if mating is non-random, the substitution effect (α) is defined (equation 10 in Falconer [7] and equation 4.22 in Lynch and Walsh [8]) as

$$\alpha = \frac{e}{(1+F)},$$

where e is the average excess and F is the inbreeding coefficient which is the reduction of heterozygote frequencies relative to those expected in random mating. Substituting e , the substitution effect in a non-random mating can be written as

$$\alpha = a + d(q - p) \left(\frac{1-F}{1+F} \right).$$

Looking at the term involving F , we can write

$$\left(\frac{1-F}{1+F} \right) = \left(\frac{1+F-2F}{1+F} \right) = 1 - 2 \frac{F}{1+F}.$$

So, we can split α in two components, one if there is no inbreeding

$$\alpha_{noF} = a + (q - p)d,$$

and another one, that involves inbreeding

$$\alpha_F = -2 \frac{F}{1+F} (q - p)d.$$

If $F = 0$ then $\alpha_F = 0$ and we get the usual expression for $\alpha = a + d(q - p)$. If $F = 1$ then $\alpha_F = -2 \frac{F}{1+F} (q - p)d = -2 \frac{1}{2} (q - p)d = -(q - p)d$.

In which case we get $\alpha = \alpha_{noF} + \alpha_F = a$. This make sense because if $F = 1$ there are only homozygotes.

The inbreeding load (i) for an individual is therefore the gene content times the substitution effect with inbreeding, α_F , as follows:

$$i_{A_1A_1} = (2 - 2p)\alpha_F = (2 - 2p) \left(-2 \frac{F}{1+F} (q - p)d \right) = -4 \frac{F}{1+F} q(q - p)d,$$

$$i_{A_1A_2} = (1 - 2p)\alpha_F = (1 - 2p) \left(-2 \frac{F}{1+F} (q - p)d \right) = -2 \frac{F}{1+F} (q - p)^2 d,$$

$$i_{A_2A_2} = (-2p)\alpha_F = (-2p) \left(-2 \frac{F}{1+F} (q - p)d \right) = 4 \frac{F}{1+F} p(q - p)d.$$

Note that the inbreeding load is related to the analyzed trait (e.g. milk yield) and the amount of it will depend on the direction of dominance for loci affecting the trait.

At a single locus, the genetic variance due to α_F is

$$\sigma_i^2 = 2pq\alpha_F^2 = 8pq \frac{F^2}{(1+F)^2} (q - p)^2 d^2.$$

Note that this variance has to be very small, because $\frac{F^2}{(1+F)^2} \approx F^2$ which is usually very small.

The variance due to α_{noF} is the usual expression $\sigma_u^2 = 2pqa^2 + 2pq(q - p)^2 d^2$.

In addition, the covariance between i and u (the last is the usual BV) is

$$\sigma_{u,i} = -4pq \frac{F}{1+F} (q - p)^2 d^2.$$

Note that the correlation between the breeding value of the trait and its inbreeding load is always negative. Looking at the magnitude of the squared correlation

$$\frac{1}{r_{(u,i)}^2} = \frac{\left(8pq\frac{F^2}{(1+F)^2}(q-p)^2d^2\right)(2pqa^2+2pq(q-p)^2d^2)}{16p^2q^2\left(\frac{F}{1+F}\right)^2(q-p)^4d^4} = \frac{(2pqa^2+2pq(q-p)^2d^2)}{2pq(q-p)^2d^2},$$

or in other words

$$r_{(u,i)}^2 = \frac{2pq(q-p)^2d^2}{(2pqa^2+2pq(q-p)^2d^2)},$$

the squared correlation is simply the fraction of additive variance in non-inbred population ($\sigma_a^2 = 2pqa^2$) due to dominance gene action.

Methods

Data for this study were extracted from the French National dairy sheep database. Animal care and use committee approval was not necessary for this study because the data were obtained from an existing database.

Phenotypic and pedigree data

Dairy sheep selection schemes have clearly defined and consensual selection objectives that have been updated periodically. Depending on the breed, the breeding objectives include milk yield, fat and protein yields, fat and protein contents, somatic cell score, udder morphology [9]. All these traits are recorded on farm. A total of 658,731, 541,180 and 2,168,454 records of milk yield from 178,123, 151,863 and 596,586 females of BB, MTN and MTR, respectively, were included. Milk recording is performed according to the International Committee for Animal Recording (ICAR). Average milk yields (\pm SD) were 193.00 (\pm 76.25) liters, 144.31 (\pm 60.25) liters and 197.52 (\pm 83.66) liters, in BB, MTN and MTR respectively. Pedigrees included 190,276 (186,581 females and 3695 males in BB,), 166,028 (162,584 females and 3444 males in MTN) and 633,655 (622,425 females and 11,230 males in MTR) animals born between 1985 and 2021. By 1985 all breeds had ongoing, routine, pedigree and milk yield recordings. To assess pedigree completeness, the number of equivalent complete generations was computed

using PEDIG software [10]. In all breeds, inbreeding is managed through (i) avoiding mating between individuals with common grandparents and (ii) trying to keep balanced numbers of rams within family of paternal grand-sires at each step of selection.

Mendelian decomposition of inbreeding

Based on pedigree data, F can be decomposed into coefficients attributed to specific founders known as partial inbreeding coefficients. The inbreeding coefficient of individual j (F_j) can be decomposed in a sum of partial inbreeding coefficients each due to an ancestor k , e.g. $F_{j(k)}$ where k is an ancestor of j . Thus $F_j = \sum_{k \in \text{ancestors}(j)} F_{j(k)}$. Note that most ancestors do not generate partial inbreeding; only those ancestors common to both sides of the pedigree (mother and father of j) have $F_{j(k)} \neq 0$.

Partial inbreeding coefficient for animal j attributed to ancestor k , ($F_{j(k)}$), is therefore the probability that j inherit IBD alleles from ancestor k [11, 4, 5]. In this study, $F_{j(k)}$ were computed following Mendelian decomposition of inbreeding, based on the source of coancestry between the parents of each individual, with values between 0 and 1. This decomposition of inbreeding splits inbreeding among founders and the Mendelian sampling of the non-founders [12]. To calculate partial inbreeding coefficients attributed to an animal j , we used the approach proposed by García-Cortés et al. [11] which modified the conventional tabular method with a set of recursively formulas. The method operates recursively over $A_{(j,k)} = \frac{1}{2}(A_{(j,sire)} + A_{(j,dam)}) + \phi_{jk}$, where $A_{(j,k)}$ is the additive genetic relationship between individuals j and k (or two times the coancestry between those two individuals), and ϕ_{jk} is the Mendelian sampling variation and it is related to within-family variation. For a given matrix $\Phi = \{\phi_{jk}\}$, $\Phi = 0$ except at the element ϕ_{jj} of the diagonal that are handled as one of the following options: (1) when both parents are known, we use $\phi_{jj} = \frac{1}{4}(1 - F_s) + \frac{1}{4}(1 - F_d)$ where F_s (F_d) is the inbreeding coefficient of the sire (or dam); (2) when only one parent ($l = s, d$) is known, we use $\phi_{jj} = \frac{1}{2} + \frac{1}{4}(1 - F_l)$ and (3) when both parents are unknown we use $\phi_{jj} = 1$. Each individual in the population is considered as a partial founder where it contributes to the genetic variability. The term ϕ includes the Mendelian sampling variability and the ignorance about the knowledge of the parents [11].

The partial inbreeding coefficients were calculated using a Fortran program available at <https://github.com/alegarra/getPartialInbreeding>. Partial inbreeding coefficients from the Mendelian decomposition of inbreeding were used later on mixed models for the genetic analysis of milk yield.

Models

The partial inbreeding coefficients were included in a mixed model as random regression covariates, to predict genetic variance and breeding values of the inbreeding load for milk yield. The effects: flock-year-parity where parity has three classes (1, 2, 3 and more), the age at lambing within year and parity, the period of lambing within year and parity and the lambing-first test-day interval within year and parity, were included in the model as fixed effects. The model including the inbreeding load can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}b + \mathbf{Z}_u\mathbf{u} + \mathbf{Z}_u\mathbf{K}\mathbf{i} + \mathbf{Z}_p\mathbf{p} + \mathbf{e} \quad (1)$$

where \mathbf{y} is the vector of phenotypic records (milk yield), $\boldsymbol{\beta}$ is the vector of fixed effects, b is the inbreeding depression parameter per unit of inbreeding and the covariate \mathbf{f} is the vector of total inbreeding coefficients. The vectors of genetic effects, \mathbf{u} and \mathbf{i} , are the additive genetic and the inbreeding load effects, respectively; such as $\begin{pmatrix} \mathbf{u} \\ \mathbf{i} \end{pmatrix} \sim N\left(\begin{matrix} 0 \\ 0 \end{matrix}, \mathbf{G} \otimes \mathbf{A}\right)$; where $\mathbf{G} = \begin{bmatrix} \sigma_u^2 & \sigma_{u,i} \\ \sigma_{u,i} & \sigma_i^2 \end{bmatrix}$, and $\sigma_{u,i}$ is the covariance between the additive genetic and the inbreeding load effects. The genetic correlation between the breeding value of milk yield and its inbreeding load was computed as $r_{(u,i)} = \frac{\sigma_{u,i}}{\sqrt{\sigma_u^2 \sigma_i^2}}$ [6]. The models also included a random permanent effect for each animal ($\mathbf{p} \sim N(0, \mathbf{I}\sigma_p^2)$) and the residual ($\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$). The incidence matrices \mathbf{X} , \mathbf{Z}_u , and \mathbf{Z}_p relate records to fixed effects, and additive genetic and permanent environmental effects, respectively. The matrix \mathbf{K} is a lower triangular matrix, $\mathbf{K} = \mathbf{T}(\mathbf{I} - \mathbf{P})$, where \mathbf{T} contains the partial inbreeding coefficients of all individuals, \mathbf{I} is the identity matrix and the product $\mathbf{Z}_u\mathbf{K}$ links the phenotypes of those animals and their ancestors causing inbreeding. Matrix \mathbf{K} considers the fact that it is not the inbreeding load of the individual that affects the phenotype, but the combined load of their ancestors. Thus, it is an indirect genetic effect (in a way similar to maternal effects, but more complex) and needs a custom incidence matrix. The

matrix \mathbf{P} has 0 in its diagonal and its elements 0.5 connect an individual with its sire and dam [5].

The full model (FM) in equation (1) was compared to a model without the inbreeding load called the reduced model (RM): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}b + \mathbf{Z}_u\mathbf{u} + \mathbf{Z}_p\mathbf{p} + \mathbf{e}$. (Co)variance components were estimated using restricted maximum likelihood (REML) for FM and RM. The superiority of the FM over RM was tested by a likelihood ratio test, which was calculated as $-2\ln(\text{likelihood for RM}) + 2\ln(\text{likelihood for FM})$. The likelihood ratio follows a mixture of χ^2 -distributions with 0 and 1 degree of freedom [13].

The matrix \mathbf{K} was computed using a program in Julia available at <https://github.com/alegarra/getPartialInbreeding>. To avoid computational problems per numerical over/underflows, we included only absolute values of \mathbf{K} matrix higher than 0.01.

Inbreeding \mathbf{f} was calculated with the `inbupgf90` program [14] available at <http://nce.ads.uga.edu/wiki/doku.php?id=readme.inbupgf90>. Programs of the BLUPF90+ family [15] were used to estimate variance components, and are available at <http://nce.ads.uga.edu/wiki/doku.php>.

Results

Inbreeding and its Mendelian decomposition

Pedigree-based inbreeding coefficients for each breed are shown in Table 1. Low inbreeding coefficients were estimated for the whole population (less than 1%). Our inbreeding estimates agreed with values obtained in other dairy sheep breeds: Latxa Cara Negra from Euskadi (0.018), Latxa Cara Rubia (0.016), and Latxa Cara Negra from Navarre (0.018) [16]. Among inbred animals, inbreeding coefficients were higher (~ 0.03) and agreed with estimates obtained on genotyped animals on the same breeds [17]. Few animals (less than 1%) presented inbreeding coefficients greater than 0.10 (Table 1).

Table 1 Descriptive statistics for inbreeding in the three breeds

Breed	Inbred animals (%)	Fraction (%) of inbred animals with $F < 0.05$	Fraction (%) of inbred animals with $F > 0.1$	Average F among inbred animals	Average F in the whole population
BB	41	85	0.6	0.032	0.013
MTN	24	88	0.9	0.030	0.007
MTR	56	96	0.5	0.025	0.014

BB: Basco-Béarnaise; MTN: Manech Tête Noire; MTR: Manech Tête Rousse
 F : coefficient of inbreeding

The partial inbreeding coefficients from the Mendelian decomposition of inbreeding are presented in Table 2. A total of 9,775,475, 2,235,928 and 75,119,288 coefficients were generated, belonging to 3855, 3124 and 12,344 ancestors (917, 828 and 2716 sires and 2938, 2296, 9628 dams) in the BB, MTN and MTR breeds, respectively. Note that these are only some of the ancestors, i.e. most ancestors do not generate inbreeding (although they would give a very large number of generations). The total number of ancestors in the three breeds was actually 91,476, 72,467 and 308,848 individuals in BB, MTN and MTR, respectively.

Table 2 Distribution of partial inbreeding coefficients in the three breeds

Breed	Nb of coefficients	Average (SD)	Max	Nb of Ancestors involved*
BB	9,775,475	2.4×10^{-4} (9.9×10^{-4})	0.25	3855
MTN	2,235,928	5.5×10^{-4} (2.1×10^{-3})	0.25	3124
MTR	75,119,288	1.2×10^{-4} (7.2×10^{-4})	0.25	12,344

Nb: Number; SD: Standard Deviation; Max: Maximum

BB: Basco-Béarnaise; MTN: Manech Tête Noire; MTR: Manech Tête Rousse

*Ancestors involved: ancestors that generate partial inbreeding coefficients

Note that if the ancestor is distant, the partial inbreeding coefficient is small and if the ancestor is close, the partial coefficient is large. Very few individuals had a partial inbreeding coefficients of 0.25: only 4 animals in BB, 7 animals in MTN and 32 animals in MTR. In BB, the partial inbreeding coefficient of 0.25 was due to a relationship where the animal was the result of a couple mother and son. All these high values are old (<1995) and precede the current organization of artificial insemination and matings in the selection schemes.

Figure 1 shows the distribution of the log₁₀ of the partial inbreeding coefficients. Most of the partial inbreeding coefficients, 97.7% and 95.8%, were lower than 0.01 (-2 in the log₁₀ scale)

and 0.001 (-3 in the log₁₀ scale) on average in the three breeds. Partial inbreeding coefficients greater than 0.01 (-2 in the log₁₀ scale in Figure 1) were generated by 732, 519 and 2380 ancestors in BB, MTN and MTR, respectively. Among these ancestors, 13%, 22% and 12% of them were founders in BB, MTN, MTR, respectively. Among all the ancestors that generated inbreeding (Table 2), 20%, 25% and 17% of them were founders in BB, MTN and MTR, respectively.

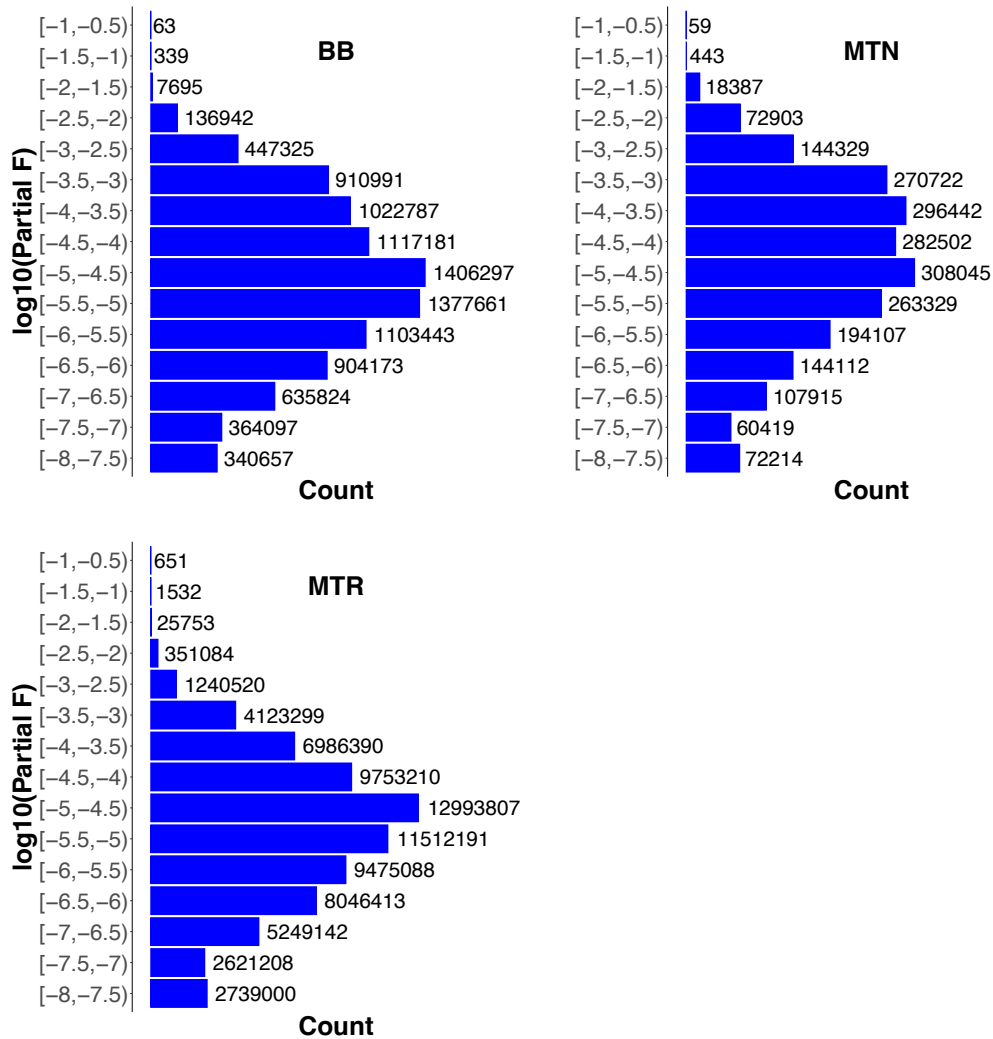


Figure 1 Distribution of the log₁₀ of the partial inbreeding coefficients for animals that generate inbreeding.
 BB: Basco-Béarnaise; MTN: Manech Tête Noire; MTR: Manech Tête Rousse; Partial F: partial inbreeding coefficient.

Figure 2 shows the number of times an animal appeared as a common ancestor from maternal and paternal lineages, which are the ones that generate inbreeding ($F_{j(k)} > 0$). A large number of ancestors appeared few times as ancestors generating inbreeding in the three breeds. There were 1456 animals in BB, 1201 animals in MTN and 5256 animals in MTR which were present less than 3 times as ancestors generating inbreeding. The number of animals which contributed to partial inbreeding more than $10^{4.5} \approx 32,000$ times was reduced: only 89, 3 and 548 ancestors in BB, MTN and MTR respectively. It is on this kind of animals (animals whose inbreeding load is expressed across several descendants) that accurate estimation of variance components relies. The number of equivalent complete generations was computed for the three breeds and it was equal to 7.04, 6.18 and 7.82 for BB, MTN and MTR breeds, respectively.

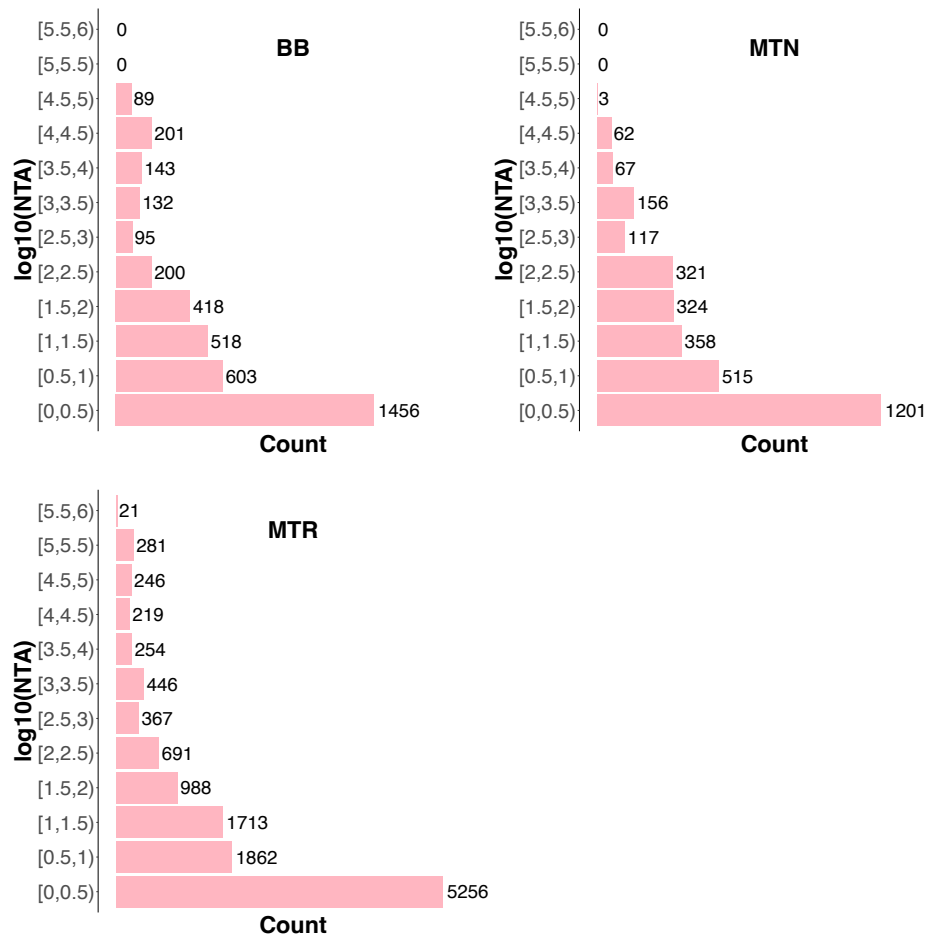


Figure 2 Distribution of the log10 of the number of times that an animal is ancestor and generates inbreeding (NTA).

BB: Basco-Béarnaise; MTN: Manech Tête Noire; MTR: Manech Tête Rousse

Genetic analysis

Genetic parameters obtained using models with (FM) and without the inbreeding load (RM) are presented in Table 3. The estimate of inbreeding load genetic variance (σ_i^2) was very inaccurate with large standard error in BB. On the contrary, σ_i^2 was highly different from zero for MTN and MTR. When we look at the likelihood ratio test (LRT) (Table 4) the perspective is somewhat different, and the null hypothesis of null inbreeding load genetic variance has to be rejected for all breeds. This, however, does not imply that σ_i^2 to be high, only that is different from 0. The obtained estimates of σ_i^2 clearly indicates variability of inbreeding load for milk yield among ancestor families in MTN and MTR breeds. For BB, it is not possible to definitely affirm that there is sizeable genetic variance of inbreeding load for milk yield.

Table 3 Parameter estimates for milk yield (liters) obtained using the two models RM and FM (SE)

Breed	Model	σ_u^2	σ_i^2	$\sigma_{u,i}$	$r_{(u,i)}$	σ_p^2	σ_e^2	b
BB	RM	847.7 (10.9)	-	-	-	417.3 (7.2)	1325.1 (2.7)	-111.8 (8.2)
	FM	847.9 (10.9)	11,804.0 (7356.2)	-289.0 (514.8)	-0.09 (0.3)	417.0 (7.2)	1325.1 (2.7)	-109.0 (11.4)
MTN	RM	681.1 (9.5)	-	-	-	363.1 (6.8)	958.5 (2.2)	-95.9 (9.3)
	FM	678.3 (9.5)	9434.7 (4089.5)	- 192.5 (436.5)	-0.08 (0.2)	364.9 (6.8)	958.5 (2.2)	-73.0 (13.9)
MTR	RM	1206.4 (8.0)	-	-	-	513.0 (4.8)	1492.9 (1.7)	-70.9 (4.5)
	FM	1205.7 (8.0)	12,923.0 (3627.1)	- 460.1 (269.6)	-0.12 (0.1)	513.2 (4.8)	1492.9 (1.7)	-50.6 (7.5)

BB: Basco-Béarnaise; MTN: Manech Tête Noire; MTR: Manech Tête Rousse

RM: reduced model (without inbreeding load effect); FM: full model (with inbreeding load effect)

σ_u^2 : additive genetic variance; σ_i^2 : inbreeding load variance; $\sigma_{u,i}$: covariance between additive genetic and inbreeding load effects; $r_{(u,i)}$: correlation between additive genetic and inbreeding load effects; σ_p^2 : permanent environment variance; σ_e^2 : residual variance; b : inbreeding depression expressed by completely inbred (100%) descendants

Table 4 Likelihood ratio test (LRT) of models included inbreeding load (FM) or not (RM)

Breed	-2 log Likelihood		LRT	
	FM	RM	χ^2	P-value
BB	6,759,011.664	6,759,143.696	132.032	7.4×10^{-31}
MTN	5,379,562.702	5,379,715.675	152.973	1.9×10^{-35}
MTR	22,626,973.729	22,627,357.574	383.845	9.1×10^{-86}

BB: Basco-Béarnaise; MTN: Manech Tête Noire; MTR: Manech Tête Rousse

FM: full model (with inbreeding load effect); RM: reduced model (without inbreeding load effect)

χ^2 : chi-square value

The inbreeding load genetic variance was larger than the additive genetic variance. This is largely a scale effect due to the small numbers involved in partial inbreeding coefficients. Note that the model gives estimates of the inbreeding load genetic effect that must be understood as the effect expressed on the phenotype (milk yield) by completely inbred (100%) descendants from a single ancestor. We rescaled the σ_i^2 to a meaningful average value of F of 0.10. Considering this value of F , the rescaled inbreeding load variances were 118.04 ($118.04 = \sigma_i^2(0.10)^2 = 11,804.0(0.10)^2$, Table 3) in BB, 94.35 in MTN and 129.23 in MTR. This rescaled variance corresponds to 4.3%, 4.5% and 4.0% of the phenotypic variance in BB, MTN and MTR, respectively.

From estimates of the genetic parameters, the genetic correlation between additive genetic and inbreeding load effects ($r_{(u,i)}$) is negative (as expected), small and with large standard errors in all cases (Table 3). An additional file includes the bivariate plot showing the relationship between additive genetic and inbreeding load effects for the three breeds [see Additional file 1].

Inbreeding depression, based on the total inbreeding coefficients, was detected for milk yield in the three breeds (Table 3). With both models, the estimate of inbreeding depression was on average equal to -110.4 liters of milk yield in BB breed. This means that a 10% increase in inbreeding would result in a reduction of 11 liters of milk yield in this breed. In MTN and MTR, estimates of inbreeding depression differed between the FM and RM models. When the model included the inbreeding load (FM), a reduction of around 26% or 2.2 liters for 10% increase in inbreeding, was observed in the estimates of inbreeding depression in MTN and MTR.

The distribution of the predicted inbreeding load was presented in Figure 3. The averages of the predicted inbreeding loads were -116.26 , 6.42 and -47.91 for BB, MTN and MTR respectively. The proportion of individuals with a positive predicted inbreeding load was 44%, 24% and 30% in BB, MTN and MTR, respectively. Additionally, 9%, 16% and 11% of the individuals in the three breeds (BB, MTN and MTR respectively) have positive predicted inbreeding loads (higher than minus the inbreeding depression estimate, $-b$), i.e. they would compensate it and even produce a positive inbreeding effect. That indicates an improving in milk yield of their inbred descendants.

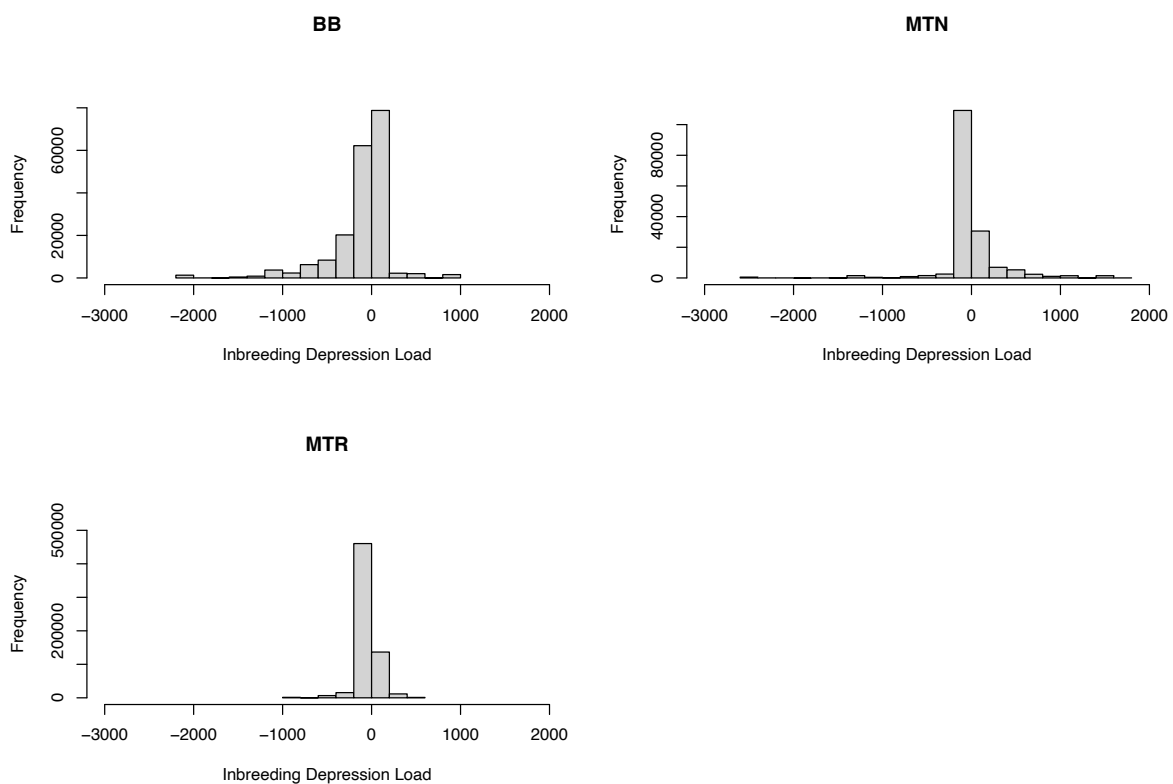


Figure 3 Distribution of the predicted inbreeding load genetic effects in the three breeds (all animals).

BB: Basco-Béarnaise; MTN: Manech Tête Noire; MTR: Manech Tête Rousse

Discussion

The Mendelian decomposition partitions inbreeding into partial inbreeding coefficients attributed to specific ancestors. Among all ancestors in the pedigrees (91,476, 72,467,

308,848), only 4.2%, 4.3%, 4.0% of them generate partial inbreeding, in BB, MTN and MTR, respectively. Among these ancestors generating inbreeding, only ~25% in MTN (and even less in the other 2 breeds) of them were founders (animals whose ancestors are unknown). Thus, inbreeding comes mainly from the Mendelian sampling of non-founders. Most partial inbreeding coefficients (~ 90%) had values lower than 0.001. These results highlight the good management of inbreeding achieved in these breeds through mating plans that avoid mating among cousins (in particular for inseminations) and through husbandry practices (for natural mating, e.g. not using rams from the same farm).

Estimates of additive genetic variance of inbreeding loads were significantly different from zero in milk yield in MTN and MTR, but not in BB. We do not find a clear explanation. The reason is not pedigree length harming the estimation - BB has a pedigree less deep than MTR (7.04 vs. 7.82 generations) but deeper than MTN (7.04 vs. 6.18 generations). An alternative hypothesis would be removal of variation in inbreeding load due to purging. In the context of inbreeding load, old inbreeding would correspond to low values of partial inbreeding and new inbreeding to high values. If selection for milk yield is strong relative to drift, purge occurs leading to elimination of deleterious alleles, and therefore there would be no variation left. However, Antonios et al. [18] could not confirm purge in BB using Kalinowski's inbreeding coefficients. Still, BB has the smallest effective population size of the three breeds (59 based on Runs of Homozygosity, compared to 81 for MTN and 109 for MTR) [17], and this could have led to purge. However, we cannot confirm this hypothesis.

Estimates of the genetic correlation between the additive genetic value and the inbreeding load were negative (as expected from the theory presented above) and low (~ -0.1). For all the breeds, the correlations were near zero and with large standard error. Compared to Varona et al. [5], our estimates were much lower than their estimate of genetic correlation for Pirenaica beef cattle (~ -0.4). Small and negative genetic correlations were also reported in Brown Swiss dairy cattle for fertility traits [6]. Milk yield is highly selected for in the breeds in this study [9]. The low values of genetic correlation between inbreeding load and breeding values imply that selection for milk yield will not cause increase inbreeding depression in milk yield in inbred animals.

Prediction of inbreeding load of individuals without progeny is possible based on relatives with inbred descendants. Artificial purging based on predicted inbreeding load effects could be

performed to reduce the effect of inbreeding depression as suggested by Varona et al. [5] and Martinez-Castillero et al. [6]. Even if this artificial purging is feasible in theory, the magnitude of inbreeding load effects predicted in this study does not seem enough to warrant selection based on this criterion.

Accurate estimates of inbreeding load effects are also an issue. Here we use pedigree information to estimate inbreeding coefficients, whereas use of SNP markers would be more accurate. Even if genomic selection was introduced in these dairy sheep breeds in 2016, there are not enough animals genotyped for that purpose (essentially all females would need to be genotyped), so the accuracy of inbreeding load prediction with genomic selection needs to be quantified [19].

Selecting individuals based on predicted inbreeding load for milk yield would basically remove recessive alleles reducing milk yield in homozygote carriers. However, recessive alleles for milk yield may have a pleiotropic effect on fitness traits, and selection to eliminate them may increase or decrease fitness. Currently, options to deal with decrease in fitness due to selection for economical traits include selection for fitness traits such as fertility, or mate allocation strategies to avoid genetic effects [20]. Thus, instead of using the inbreeding load predictions for selection, they could be used to avoid undesirable matings. Still the magnitude of these effects in this study precludes this strategy. Further research in other species and traits is needed to explore the possible, if any, benefits of these genetic management strategies.

Conclusions

We present theory that show the additive nature of the inbreeding load. The inbreeding load effect and the regular (in non-inbred population) additive genetic effect have a negative correlation depending on allele frequencies, inbreeding and dominance. There is genetic variance for inbreeding load MTN and MTR breeds, but it was not different from zero for BB. As expected, we estimated negative genetic correlations between inbreeding load and breeding values; however, estimates were close to zero in the three sheep breeds. The small magnitude of inbreeding load does not warrant selection based on this criterion.

Declarations

Ethics approval and consent to participate

Data for this study were extracted from the French National dairy sheep database. Animal care and use committee approval was not necessary for this study because the data were obtained from an existing database.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from the breeders, but restrictions apply to the availability of these data, which are not publicly available. Software is available in <https://github.com/alegarra/getPartialInbreeding>.

Competing interests

The authors declare that they have no competing interests.

Funding

This study has received funding from the project ARDI2 (grant agreement EFA 032/01) from POCTEFA funds.

Authors' contributions

SA wrote the first draft of the manuscript, performed the statistical and genetic analysis, and interpreted the results. STRR and ZGV participated in the supervision, validation and writing. ZGV and AL developed the theory, proposed the methodology and participated in the interpretation of the results, writing and review of the manuscript. AL contributed in the creation of Fortran and Julia programs used in this work. LV participated in the interpretation and review of the work. JMA provided the data and participated in editing, interpretation and review. All authors read and approved the final manuscript.

Acknowledgements

The authors thank breeders who provided the data. This study has received funding from the project ARDI2 (grant agreement EFA 032/01) from POCTEFA funds. We are grateful to the Genotoul Bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing computing and storage resources.

References

1. Leroy G. Inbreeding depression in livestock species: review and meta-analysis. *Anim Genet.* 2014;45:618–28.
2. Doekes HP, Bijma P, Windig JJ. How depressing is inbreeding? A meta-analysis of 30 years of research on the effects of inbreeding in livestock. *Genes (Basel).* 2021;12:926.
3. Gulisija D, Gianola D, Weigel KA, Toro MA. Between-founder heterogeneity in inbreeding depression for production in Jersey cows. *Livest Sci.* 2006;104:244–53.
4. Casellas J. On individual-specific prediction of hidden inbreeding depression load. *J Anim Breed Genet.* 2018;135:37–44.
5. Varona L, Altarriba J, Moreno C, Martínez-Castillero M, Casellas J. A multivariate analysis with direct additive and inbreeding depression load effects. *Genet Sel Evol.* 2019;51:78.
6. Martinez-Castillero M, Varona L, Pegolo S, Rossoni A, Cecchinato A. Bayesian inference of the inbreeding load variance for fertility traits in Brown Swiss cattle. *J Dairy Sci.* 2021;104:10040–8.
7. Falconer, D.S. A note on Fisher's 'average effect' and 'average excess'. *Genet. Res. Camb.* 1985;46:337–347.
8. Lynch, M., and B. Walsh. *Genetics and analysis of quantitative traits.* 1998. Sunderland, MA: Sinauer.
9. Astruc J-M, Buisson D, Larroque H, Lagriffoul G. New breeding goals and role of genomics on adaptation and resilience traits in French dairy sheep. In: *Proceedings of the 69th Annual Meeting of the European Federation of Animal Science: 27-31 August 2018; Dubrovnik.* 2018.
10. Boichard D. PEDIG: a fortran package for pedigree analysis suited for large populations. In: *Proceedings of the 7th world congress on genetics applied to livestock production: 19–23 August 2002; Montpellier.* 2002.
11. García-Cortés LA, Martínez-Ávila JC, Toro MA. Fine decomposition of the inbreeding and the coancestry coefficients by using the tabular method. *Conserv Genet.* 2010;11:1945–52.
12. Caballero A, Toro MA. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet Res.* 2000;75:331–43.
13. Visscher P.M. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Res Hum Genet* 2006;9:490–495.
14. Aguilar I, Misztal I. Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. *J Dairy Sci.* 2008;91:1669–72.
15. Lourenco D, Tsuruta S, Aguilar I, Masuda Y, Bermann M, Legarra A, Misztal I. Recent updates in the BLUPF90 software suite. In: *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production: 3–8 July 2022; Rotterdam.* 2022.

16. Granado-Tajada I, Rodríguez-Ramilo ST, Legarra A, Ugarte E. Inbreeding, effective population size, and coancestry in the Latxa dairy sheep breed. *J Dairy Sci.* 2020;103:5215–26.
17. Rodríguez-Ramilo ST, Elsen JM, Legarra A. Inbreeding and effective population size in French dairy sheep: comparison between genomic and pedigree estimates. *J Dairy Sci.* 2019;102:4227–37.
18. Antonios S, Rodríguez-Ramilo ST, Aguilar I, Astruc JM, Legarra A, Vitezica ZG. Genomic and pedigree estimation of inbreeding depression for semen traits in the Basco-Béarnaise dairy sheep breed. *J Dairy Sci.* 2021;104:3221–3230.
19. Varona L, López-Carbonell D, Srihi H, Ramírez M, Martínez P, Hermida M, Altarriba J, Casellas J. Genetic variability in the individual inbreeding load: genomic prediction for artificial purging. In: *Proceedings of the 12th World Congress on Genetics Applied to Livestock Production: 3–8 July 2022; Rotterdam.* 2022.
20. Bengtsson C, Stålhammar H, Thomasen JR, Eriksson S, Fikse WF, Strandberg E. Mating allocations in Nordic Red Dairy Cattle using genomic information. *J Dairy Sci.* 2022;105(2):1281–97.

3.1 Appendix

Additional file 1

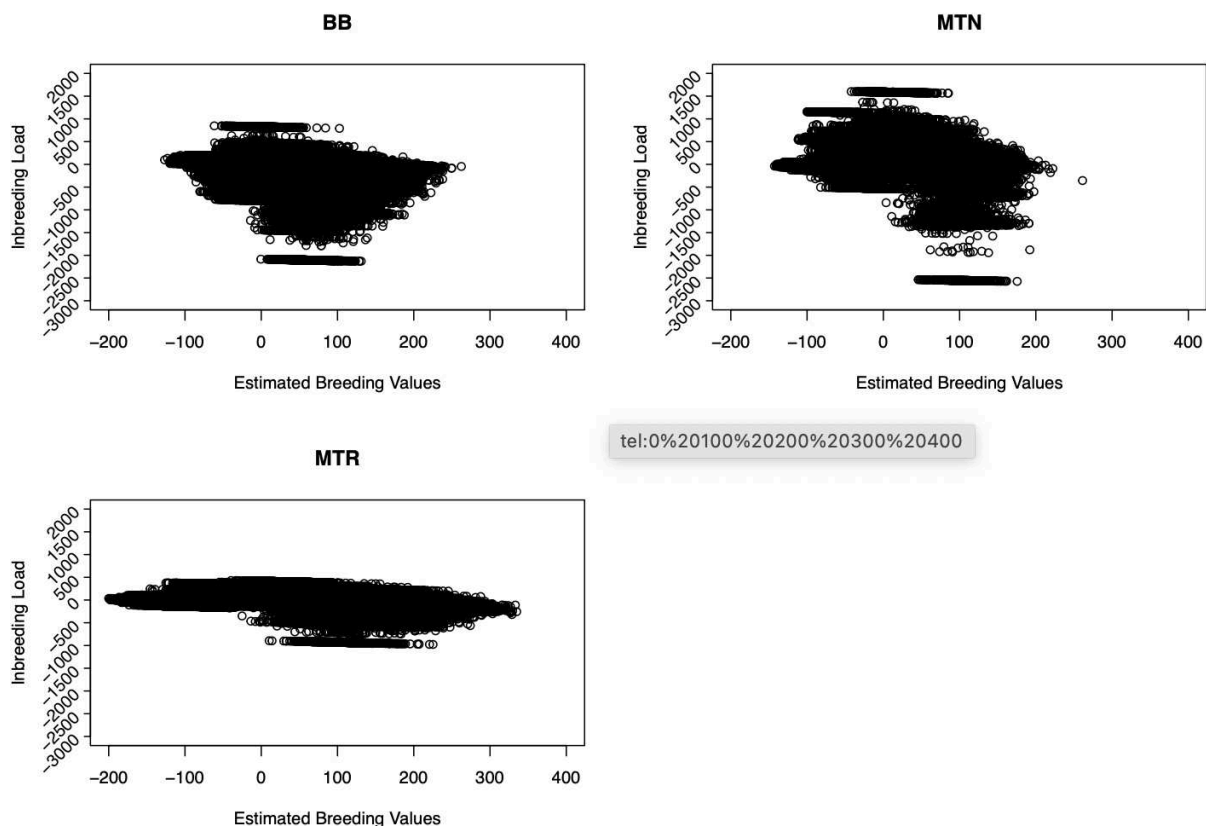


Figure S1. Bivariate plot showing the relationship between additive genetic and inbreeding load effects for Basco-Béarnaise (BB), Manech Tête Noire (MTN) and Manech Tête Rousse (MTR) breeds.

Chapter 4

Exploring options to select based on inbreeding load via simulation

S. Antonios^{1*}, G. Gorjanc², I. Pocrnic², J. Obšteter³, S.T. Rodríguez-Ramilo¹ and Z.G.
Vitezica¹

¹ GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan,
France

² The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University
of Edinburgh, Easter Bush, Midlothian, EH25 9RG, United Kingdom

³ Department of Animal Science, Agricultural Institute of Slovenia, Hacquetova ulica
17, 1000 Ljubljana, Slovenia

Manuscript in preparation

*Corresponding author: S. Antonios

INRAE, UMR 1388 GenPhySE

24 Chemin de Borde Rouge

31326 Castanet Tolosan Cedex, France

Phone: + 33 7 55 29 60 96

E-mail: simona.antonios@inrae.fr

Abstract

Individual inbreeding load, which is the fraction of the mutation load, is considered a hereditary additive trait. This load is only expressed when inbreeding occurs in the individual's offspring, and it is related to the analysed trait. Inbreeding load shows heterogeneity among individuals. Thus, inbreeding load can be used as a criterion of selection. Then individuals could be selected based on both, their estimated breeding values and their breeding values of inbreeding load. Those results led us to evaluate the effectiveness of involving inbreeding load in selection strategies in a dairy sheep breeding scheme. Where we computed a simulation of 10 generations of evaluations and selection. Six scenarios that differ in the criteria of selection (only estimated breeding values, only inbreeding load predictions, or both, estimated breeding values and inbreeding load predictions) and mate strategies minimising inbreeding load or adjusting estimated breeding values by the expected future inbreeding were evaluated. Scenarios were compared in terms of genetic gain, pedigree-based inbreeding coefficients, rate of inbreeding, effective population size, accuracy of selection. Results showed that there was no significant difference between the five scenarios where selection was based on estimated breeding values over the ten years of selection. As expected, the scenario selecting only on the inbreeding load predictions was significantly different from those five scenarios in terms of genetic gain. To conclude, selecting animals using inbreeding load was feasible. However, the magnitude of inbreeding load effects and its accuracy did not show a clear practical interest. Furthermore, the benefit of using inbreeding load in mate allocation strategies was negligible.

Introduction

Inbreeding load (IL) is the fraction of the mutation load that is due to recessive variants which can be hidden in heterozygous condition. The IL is trait specific and the amount of it will depend on the direction of dominance for loci affecting the trait. Inbreeding load of individuals is a heritable additive trait that is only expressed when the inbreeding occurs in their offspring (Varona et al., 2019).

Genetic parameters for IL had been estimated for growth traits in rabbits (Casellas, 2018), pigs (Casellas et al., 2008) and beef cattle (Varona et al., 2019); for morphological traits in horses

(Poyato-Bonilla et al., 2020); and for fertility traits (Perdomo-González et al., 2021) in horses and dairy cattle (Martinez-Castillero et al., 2021). Recently, Antonios et al. (2024) estimated IL variances for milk yield in Manech Tête Noire (MTN) and Manech Tête Rousse (MTR) breeds and showed variability in IL in both breeds.

In their previous work Antonios et al. (2024) showed that the genetic correlations between the additive genetic (breeding) values and the inbreeding load were negative and close to zero (~ -0.1) for milk yield in MTN and MTR sheep breeds. Similarly, for Pirenaica beef cattle Varona et al. (2019) also reported negative (-0.4) genetic correlations between the additive genetic values and the inbreeding load. Such negative correlations indicate that animals with high breeding values tend to cause worse inbreeding depression if their descendants are inbred (Varona et al., 2019). Furthermore, close to zero correlations imply that the ILs are not genetically correlated with the additive genetic effects, which means that individuals could be selected based on both; their standard estimated additive breeding values (EBV) and their estimated IL values. For example, by selecting individuals based on the predicted inbreeding load for milk yield we could basically remove recessive alleles that reduce the milk yield in homozygote carriers. Other possibility could be to use IL predictions for avoiding undesirable matings. Mate allocation strategies have been used in animal breeding with different purposes, for instance, for exploiting dominance (González-Diéguez et al., 2020; Toro and Varona, 2010) or managing genetic defects (Bengtsson et al., 2022). To our knowledge, inbreeding load predictions have not yet been used in mate allocation strategies. Thus, matings involving individuals with ancestors with favorable inbreeding load (positive inbreeding load for milk yield or other trait of interest) could be prioritised; while matings with ancestors with low and negative inbreeding load (that generate worse inbreeding depression in inbred offspring) could be efficiently avoided. In this context, there are no studies that explore the possible benefits of involving inbreeding load predictions in such selection strategies.

The aim of this study was to evaluate the effectiveness of including the inbreeding load in selection strategies for a dairy sheep breeding scheme via stochastic simulations. In this sense, we deployed different scenarios that differed in the selection criteria: (i) selection only on estimated breeding values (EBV), (ii) selection on both EBV and IL predictions, (iii) selection only on IL predictions, (iv) selection on both EBV and IL predictions coupled with the mate allocation strategies or (v) adjusting EBV by the expected future inbreeding. Scenarios were

compared in terms of genetic gain, pedigree-based inbreeding coefficients, rate of inbreeding, effective population size, and the accuracy of selection.

Materials and Methods

In this study, we analysed how inclusion of the inbreeding load in selection strategies impacts the genetic trends, prediction accuracies and inbreeding parameters in dairy sheep selection scheme over 20 years. The selection strategies (scenarios) included: (i) selection based on EBV, (ii) selection based on EBV and IL assuming an inbreeding of 1% or (iii) of 10%, (iv) selection based only on IL, (v) after selection based on EBV, a mate allocation strategy was used to minimize the IL in matings, (vi) selection was based on EBV adjusted by the expected future inbreeding.

The material and methods are subdivided into two sections. The first section describes the simulation of the founder population, and the second section describes the simulation of the selection scheme.

Simulation of the founder population

We used AlphaSimR (Gaynor et al., 2021) to simulate a realistic dairy sheep selection scheme that closely mimics the French dairy sheep selection scheme. Our simulation is driven by a set of parameters (Table 1) imitating the French dairy sheep selection scheme.

Historical population

We added a historical demography of the sheep population (Figure 1), by setting the effective population size (N_e) for the base population to 150 (Rodríguez-Ramilo et al., 2019) with a decrease of N_e from 7,500 at around 10,000 generations ago in line with the estimates used by Villa-Angulo et al. (2009). Successive reduction of N_e was used to reflect a progressive restriction of genetic variance due to natural and artificial selection.

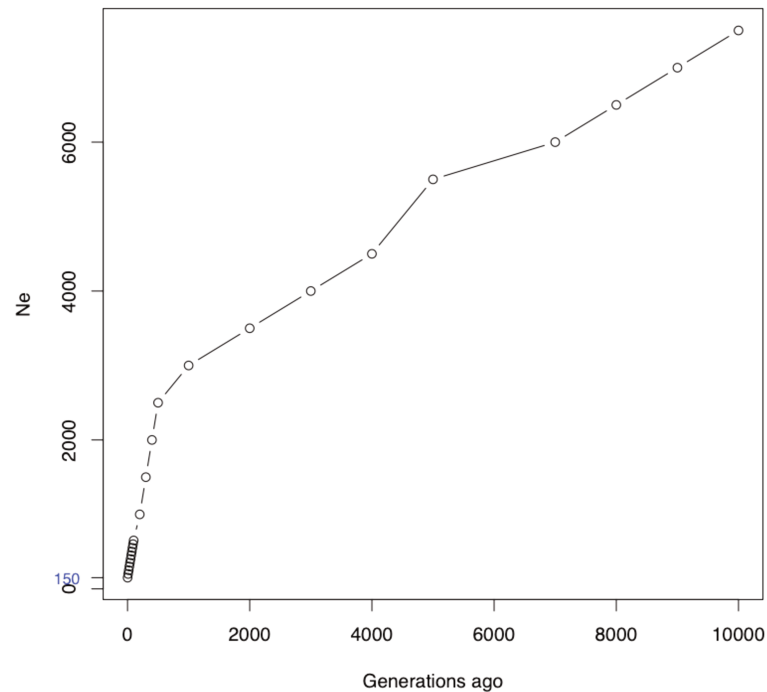


Figure 1. Effective population size (N_e) for the last 10,000 generations of simulated historical population and the actual $N_e = 150$ (blue) at time zero.

Founder genome

We used a coalescent process to simulate the founder genome comprised of 26 autosomal chromosomes, with 0.95×10^8 base pairs each; 4,000 were randomly chosen as segregation sites per chromosome. A total of 251 were randomly chosen as quantitative trait loci (QTLs) affected a trait (e.g. milk yield). In this study, we assumed that all the chromosomes have the same size, to keep the simulation parameters consistent across them. The chromosome sequences were generated using the Markovian coalescent simulator (MaCS; Chen et al., 2009; Gaynor et al., 2019). The recombination rate was 1.5×10^{-8} (Petit et al., 2017) and the mutation rate was 1.5×10^{-8} mutations per base pair (Lv et al., 2021).

Founder population

Subsequently, we initiated a base population (founder population) where the genomes of these individuals were created from the founding genomes. Those genomes were randomly allocated to animals. These animals were separated in different categories (sires of sires, sires of dams,

dams of sires, dams of dams, testing sires, natural mating sires and lambs) in order to initiate a dairy sheep scheme.

A trait, such as milk yield, controlled by additive and dominance QTL action was simulated. We used a heritability of 0.33, an overall population mean of 200 liters for milk yield and other parameters presented in Table 1. We fixed a flock size of 320 ± 129 animals, and a total number of flock of 237. We defined three fixed effects: flock, year and flock-year.

Table 1. Dairy sheep simulation parameters

σ_u^2	σ_i^2	$\sigma_{u,i}$	$r_{(u,i)}$	σ_d^2	σ_{pe}^2	σ_e^2
1000.0	12923.0	-359.5	-0.1	100.0	1500.0	400.0

σ_u^2 : additive genetic variance; σ_i^2 : inbreeding load variance; $\sigma_{u,i}$: covariance between additive genetic and inbreeding load; $r_{(u,i)}$: correlation between additive genetic and inbreeding load; σ_d^2 : dominance genetic variance; σ_{pe}^2 : permanent environment variance; σ_e^2 : residual variance

Genetic values

Genetic values for milk yield were simulated by summing the genetic effects at the 251 QTL. The additive biological QTL effects (a) were sampled from a standard normal distribution and scaled to obtain an additive genetic variance of $\sigma_u^2 = 1000$ in the founder population. Dominance variance (σ_d^2) was fixed to $\sigma_d^2 = 0.1\sigma_u^2$ based on Aliloo et al.'s (2017) paper.

Phenotypes

Only females had records (milk yield). Phenotypes were generated by adding fixed effects, random permanent environmental effect (pe) and random residual effect (e) to the genetic values. The fixed effects were flock, year, flock-year and class of parity and their values were obtained from real data. The random effects, pe and e , were sampled from $pe \sim N(0, \mathbf{I}\sigma_{pe}^2)$ and from $e \sim N(0, \mathbf{I}\sigma_e^2)$, where σ_{pe}^2 was the permanent environmental variance, σ_e^2 was the error variance and \mathbf{I} was an identity matrix. Values for σ_{pe}^2 and σ_e^2 were obtained from real data (Table 1). Milk yield phenotype was calculated by class of parity (1, 2, 3 and 4 lactation), by adding a general mean for each lactation. The milk yield's means were calculated from real data (209, 213, 207 and 180 liters in 1st, 2nd, 3rd and 4th lactation, respectively). The dams (dams of sires and dams of dams) that are in the same class of parity had the same milk yield mean.

Simulation of the selection scheme

After simulating the base population, a dairy sheep selection scheme was initiated and ran for 10 years of conventional pedigree-based selection (the burn-in phase), followed by 10 additional years of specific scenarios (described in the following subsections). The overview of the selection scheme is presented in Figure 2 while the every year of the selection scheme is detailed in Figure 3. The aim of the burn-in was to provide a common starting point for the future selection phase (scenarios) and to build-up the linkage disequilibrium (LD) in the breeding population. Both the burn-in and the scenarios were replicated 10 times each. At the end of these 20 years, the simulated dairy sheep scheme involved ~400K animals (~370K of them were females) with overlapping generations, mimicking a real French dairy sheep breeding scheme.

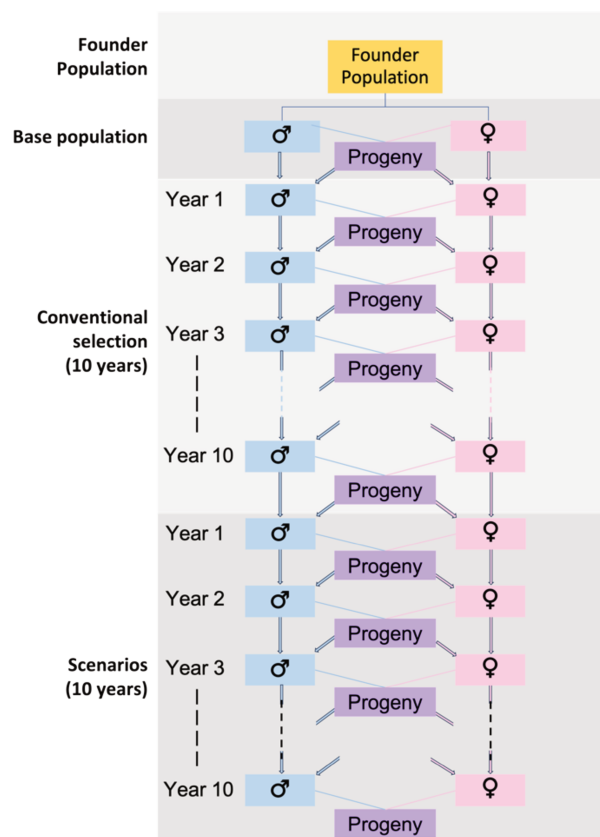


Figure 2. Schematic representation of the whole simulated dairy sheep scheme. Every year of this selection scheme (Figure 2) is presented in Figure 3.

Conventional pedigree-based selection (burn-in)

In the burn-in phase, selection was based on the EBV. Sires were selected based on their EBV estimated from their progeny testing. From matings between elite sires and elite dams, we selected the best 150 males (testing sires) based on the parents' EBV to enter to artificial insemination (AI) center at around 8 months of age. Testing sires are mated only with dams of dams (Figure 3) and each testing sire contributed 30 offspring. At 2.5 years old, we selected the best 10 males of the testing males as sires of sires and the next best 55 males as sires of dams, to be the AI sires. All of them were selected based on their breeding values estimated from their daughters records. Each AI sire stayed in service for 3 years, so in total, the number of AI sires in each year of simulation was around ~200 sires, composed of 30 AI sires of sires (the ones with the highest EBV, and selected from different families) and 165 AI sires of dams (the ones highest EBV among the remaining ones). Each AI sire of sires contributed 144 offspring per year and each AI sire of dams contributed 33 offspring per year. The natural mating (NM) sires composed of 1000 sires and were replaced every year. NM sires were selected from the progeny of AI sires, among the ones that were not selected for the testing (were not among the best 150 males). Each NM sires contributed to 19 offspring. Note that the final number of offspring per sire was calculated by taking into consideration the real breeding scheme survival rates that were equal to 0.75, a fertility of 0.6 (0.9) in AI (NM) and a prolificacy of 1.6 (1.4) in AI (NM).

Mating generated 16K female lambs in each year, and this was the final number of lambs that survived after lambing, thus accounting for the stillbirths and early deaths. The female lambs were inseminated or naturally mated at 8 months old, and became dams in the first year. In the simulation, this scheme totalled to about 33K active dams each year (Figure 3). Ten percent of the dams are dams of sires; they are the future mother of AI sires, and in this case, they are only mated with the AI sires of sires. Note that the dams of sires must be the progeny of AI sires and they are selected based on their estimated breeding values. Dams of dams were inseminated randomly from the AI sires of dams and the testing sires and the rest is naturally mated with the NM sires. In this simulation, we avoided parent offspring's mating, and mating between siblings. The replacement rate of the dams is around 23%. In this simulation, dams have 4 lactations when they are 4 years old. Thereafter, among the total number of dams of sires (3K), 35%, 26%, 23% and 18% of dams of sires are in lactation 1 (Lac 1, see Figure 3), lactation 2 (Lac 2), lactation 3 (Lac 3) and lactation 4 (Lac 4), respectively. In dams of dams' group (30K

dams of dams), we have 30%, 28%, 24% and 18% of dams of dams that are in Lac 1, 2, and 4, respectively. We considered that the prolificacy rate per dam is equal to one.

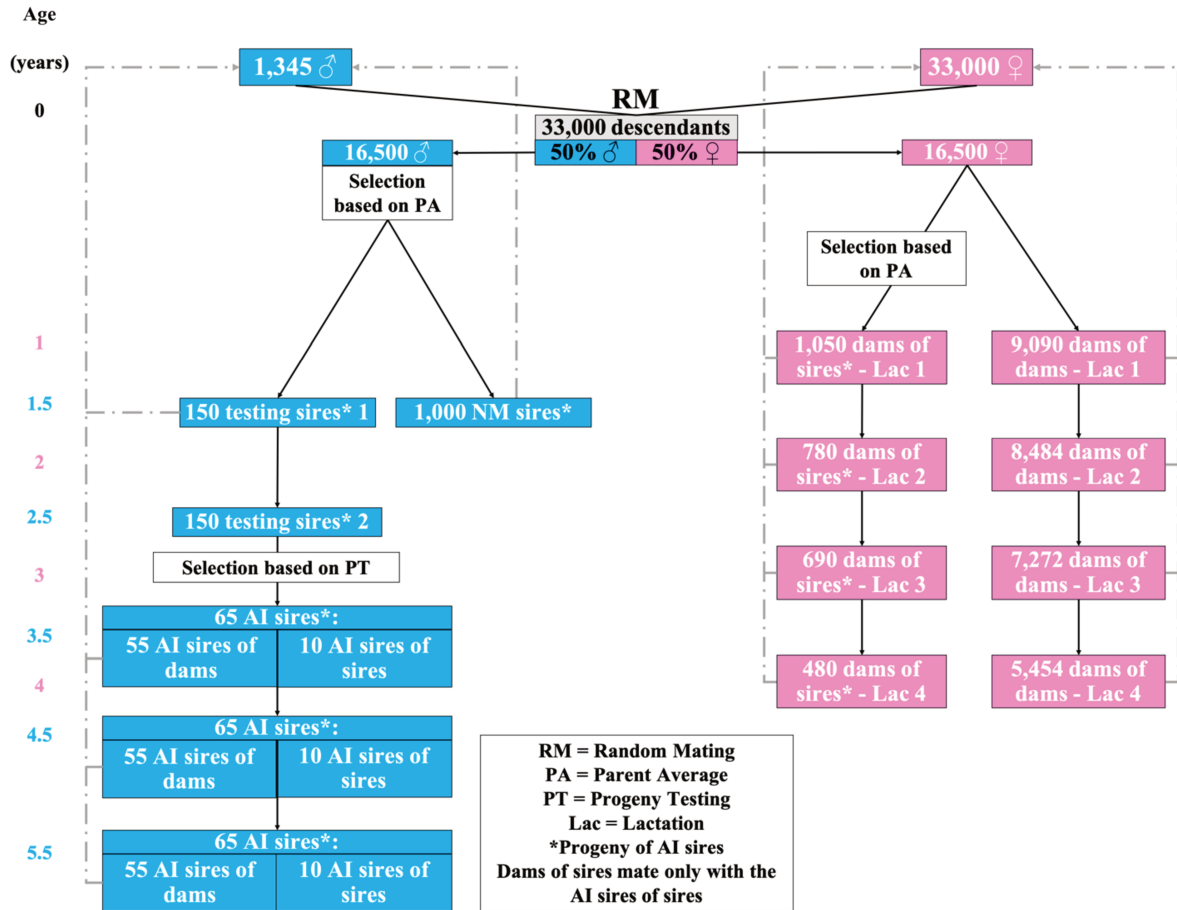


Figure 3. Schematic representation of each year in the simulated sheep selection scheme.

At the end of those 10 years of conventional pedigree-based selection, we calculated LD pattern. LD was measured by calculating the squared correlation coefficient (r^2) between all pairs of markers (Hill, 1974) assuming that 1 cM=1 Mbp. The LD decay was measured for increasing distances between markers by calculating the mean r^2 within each distance interval. The resulting average LD over replicates was 0.08 ± 0.14 for a distance of < 0.5 cM, 0.31 ± 0.35 for a distance of < 0.1 cM and 0.35 ± 0.37 for a distance of < 0.05 cM. These values are in agreement with those reported in Lacaune breed (Baloche et al., 2014) and in five other populations of domestic sheep (Meadows et al., 2008).

Scenarios

We evaluated six different scenarios that run for 10 additional years after the burn-in step (10 years of conventional pedigree-based selection). In every year, the structure of the selection scheme was the same for all scenarios as indicated in Figure 3. The scenarios differed in the selection criterion (Figure 4). Six different pedigree-based selection scenarios were evaluated: (i) selection based on EBV (sEBV), (ii) selection based on EBV and IL (Casellas, 2018) assuming an inbreeding of 1% (sEBV-IL-01) or (iii) of 10% (sEBV-IL-10), (iv) selection based only on IL (sIL), (v) after selection based on EBV, a mate allocation strategy was used to minimize the IL in matings (sMA), (vi) selection was based on EBV adjusted by the expected future inbreeding (sEFI) (VanRaden and Smith, 1999).

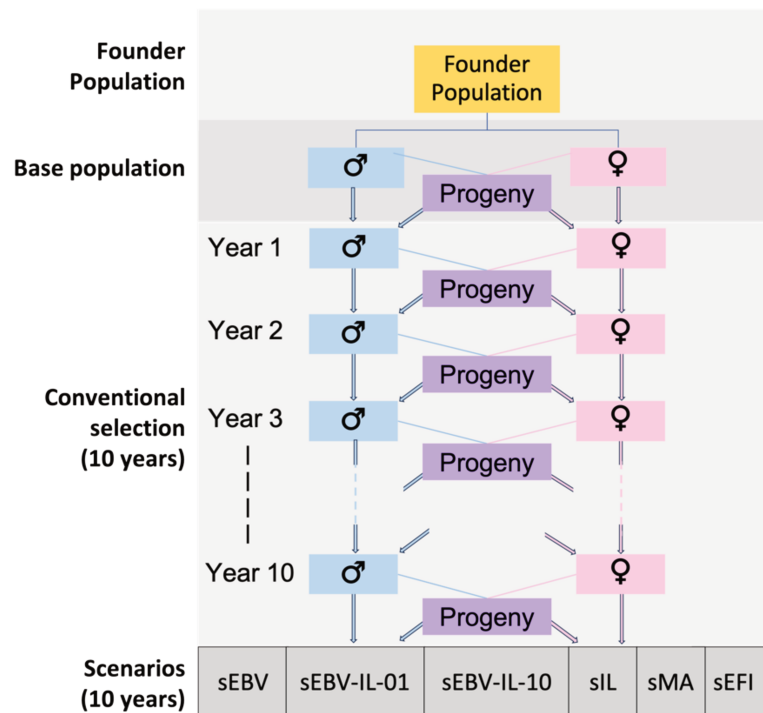


Figure 4. Schematic representation of the simulated dairy sheep scheme with six scenarios.

Conventional selection is based on estimated breeding values (EBV). sEBV: scenario selecting based on EBV; sEBV-IL-01: scenario selecting based on EBV + 0.01 of inbreeding load (IL); sEBV-IL-10: scenario selecting based on EBV + 0.1IL; sIL: scenario selecting based IL; sMA: scenario implementing mate allocation; sEFI: scenario selecting based on EBV adjusted by the expected future inbreeding.

sMA scenario took place between the 30 sires of sires and 3,000 dams of sires (90,000 possible matings). Optimization of matings was addressed via linear programming (Jansen and Wilton, 1985) using the R lpsolve package (Berkelaar et al., 2004). Three constraints were used in the optimization: (i) each sire mated with 100 dams, (ii) each dam mated only once and it gave birth to only one lamb, (iii) mating between parent and offspring as well as mating between siblings were avoided. In the MA scenario, we allowed first matings between individuals with no common ancestors, and second matings between individuals with common ancestors where the priority was for matings with a parents' average IL as high as possible (positive).

In sEFI scenario, each testing sire (150 males) was selected as AI sire (10 males), based on the best adjusted EBV (EBV_p). For each testing sire, we computed its average relationship with all the contemporary females (3,000 dams of sires) in order to obtain the EFI (expected future inbreeding) as half the average relationship of this male to the contemporary females. The EBV_p was equal to $(EBV_p = EBV + \frac{1}{2}EFI * b)$ where b was the inbreeding depression coefficient. Males that are highly related with their contemporary females were more penalized.

Prediction model

Milk yield was analysed using single-trait animal model. We estimated breeding values and inbreeding load effects using BLUPF90+ (Lourenco et al., 2022) available at <http://nce.ads.uga.edu/wiki/doku.php>. The partial inbreeding coefficients were calculated using a Fortran program available at <https://github.com/alegarra/getPartialInbreeding>. Variance components were assumed to be known from initial simulation parameters.

Pedigree-based EBV (BLUP) were obtained in scenarios sEBV and sEFI using the following model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}b + \mathbf{Z}_u\mathbf{u} + \mathbf{Z}_p\mathbf{pe} + \mathbf{e}$, where \mathbf{y} is the vector of phenotypic records (milk yield), $\boldsymbol{\beta}$ is the vector of fixed effects (flock, year, flock-year, and rang of lactation), b is the inbreeding depression parameter, \mathbf{f} is the vector of total inbreeding coefficients, \mathbf{u} is the vector of additive genetic effects, \mathbf{pe} is vector of random permanent effects and \mathbf{e} is the vector of residuals. The incidence matrices \mathbf{X} , \mathbf{Z}_u , and \mathbf{Z}_p relate records to fixed effects, and additive genetic and permanent environmental effects, respectively.

In scenarios sEBV-IL-01, sEBV-IL-10, sIL and sMA, the model included a new effet, \mathbf{i} , the vector of inbreeding load effects. This model can be written as: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}b + \mathbf{Z}_u\mathbf{u} + \mathbf{Z}_u\mathbf{K}\mathbf{i} + \mathbf{Z}_p\mathbf{p} + \mathbf{e}$, where the matrix \mathbf{K} is a lower triangular matrix, $\mathbf{K} = \mathbf{T}(\mathbf{I} - \mathbf{P})$, where \mathbf{T}

contains the partial inbreeding coefficients of all individuals, \mathbf{I} is the identity matrix and the product $\mathbf{Z}_u\mathbf{K}$ links the phenotypes of those animals and their ancestors causing inbreeding. Matrix \mathbf{P} has 0 in its diagonal and its elements 0.5 connect an individual with its sire and dam. Note that $\begin{pmatrix} \mathbf{u} \\ \mathbf{i} \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathbf{G} \otimes \mathbf{A}$; where $\mathbf{G} = \begin{bmatrix} \sigma_u^2 & \sigma_{u,i} \\ \sigma_{u,i} & \sigma_i^2 \end{bmatrix}$, \mathbf{A} is the pedigree-based additive relationship matrix, σ_u^2 is additive genetic variance, σ_i^2 is the inbreeding load variance, and $\sigma_{u,i}$ is the covariance between additive genetic and inbreeding load effects.

Measures for comparison

The six different scenarios (sEBV, sEBV-IL-01, sEBV-IL-10, sIL, sMA and sEFI), were compared in terms of genetic gain, inbreeding, rate of inbreeding, effective population size (Ne), inbreeding depression parameter (b) and selection accuracy. To make the breeding scenarios comparable, all scenarios were normalized to the last year of burn-in (year 10), so that the mean genetic value was 0 and the standard deviation was 1.

The average true total genetic value (TTGV) of animals was computed for each generation expressed relative to the TTGV at the generation 0 (last year of burn-in) to evaluate the genetic trend. For each simulated scenario, the total genetic gain for milk yield was estimated from the regression of the TTGV on the year of birth. We also calculated individual pedigree-based inbreeding coefficients using the Meuwissen and Luo (1992) algorithm as implemented in `renumf90` program (available at <http://nce.ads.uga.edu/html/projects/programs/>). From the average inbreeding coefficients per year, we computed the rate of inbreeding (ΔF) as the regression coefficient of the average inbreeding coefficients on the year of birth over all replicates. The Ne was calculated as $Ne = \frac{1}{2L\Delta F}$, where ΔF is the rate of inbreeding per year and L is the generation interval defined as the average age of the parents at the birth of their offspring. We used $L = 4$ as estimated previously for the Pyrenean breeds (Antonios et al., 2023). Selection accuracy was measured as the correlation between the estimated breeding values and the true breeding values. Results were the average of the 10 replicates of each scenario.

Results and discussion

Response to selection

Figure 5 shows the mean TTGV trends. As expected, a positive trend (genetic gain) in milk yield over the ten years of selection was observed in the five scenarios where selection was based on EBV (sEBV, sEBV-IL-01, sEBV-IL-10, sMA, and sEFI). However, in the sIL scenario, the trend of the mean TTGV was different. Note that the EBV was not included in the selection criterion in sIL scenario. The genetic trend (sIL) was positive for the first 4 years and then it started to decline. In the sIL scenario, the gain in the first four years may be attributed to animals that were previously selected based on their EBVs in the in the burn-in phase and those animals had high EBVs (high TTGV). Thus, the positive trend in the mean TTGV during four years is because there were animals, that were selected based on their EBVs (sires, dams of sires) from previous step (burn-in), they stayed for certain number of years (until they will be completely replaced) transmitting their relatively high genetic merit to their offspring.

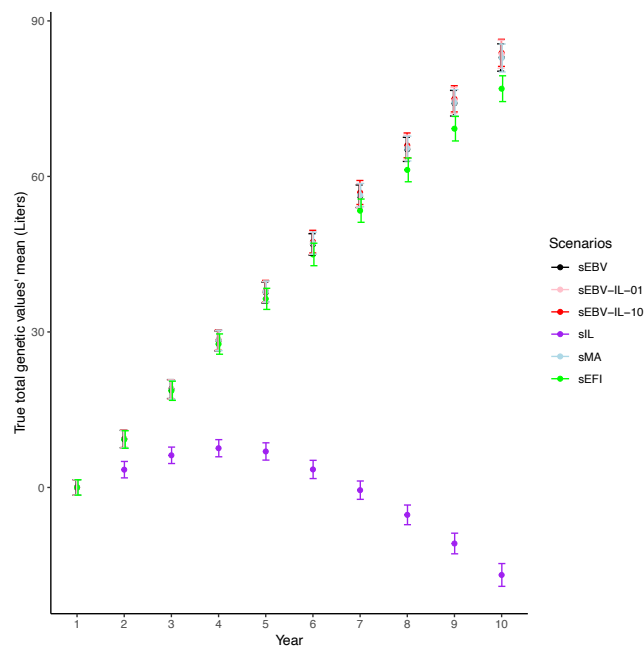


Figure 5. True total genetic value trends during 10 years of selection in the 6 different scenarios. sEBV: scenario selecting based on estimated breeding values (EBV); sEBV-IL-01: scenario selecting based on EBV + 0.01 of inbreeding load (IL); sEBV-IL-10: scenario selecting based on EBV + 0.1IL; sIL: scenario selecting based IL; sMA: scenario implementing mate allocation; sEFI: scenario selecting based on EBV adjusted by the expected future inbreeding.

All the scenarios where selection was based on EBV had similar total genetic gain for the milk yield. Five scenarios: sEBV, sEBV-IL-01, sEBV-IL-10 and sMA, had a total genetic gain of around 9 liters followed by sEFI around 8 liters (Figure 6); no significant differences ($p\text{-value} > 0.05$) were observed among them. Whereas, sIL scenario had a negative genetic gain, which means a loss in the milk yield by around 4 liters, pointing towards negative impact of selecting individuals based on their IL only. In addition, we assumed a low genetic correlation (-0.1) between the true breeding value and the inbreeding load in the simulation, thus the animals with the highest estimated IL were not necessarily those with the highest EBV which caused a loss in the milk yield.

Selection accuracy was calculated as the correlation between the true and the estimated breeding values in selection candidates. All scenarios achieved comparable accuracies of about 0.5.

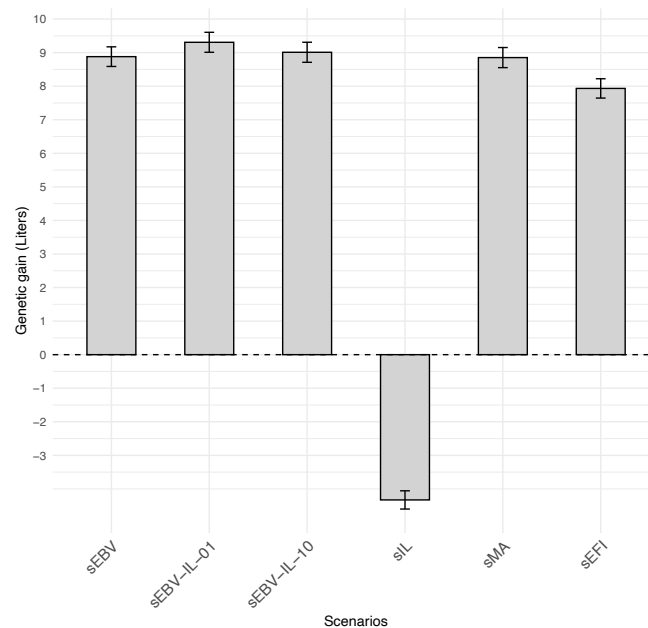


Figure 6. Total genetic gain for milk yield in the different scenarios. sEBV: scenario selecting based on estimated breeding values (EBV); sEBV-IL-01: scenario selecting based on EBV + 0.01 of inbreeding load (IL); sEBV-IL-10: scenario selecting based on EBV + 0.1IL; sIL: scenario selecting based IL; sMA: scenario implementing mate allocation; sEFI: scenario selecting based on EBV adjusted by the expected future inbreeding.

Inbreeding, inbreeding depression and rate of inbreeding

The trend of the mean of inbreeding coefficients per year are shown in Figure 7 for each scenario. Inbreeding increased from 0.3% in the last year of the the burn-in phase to around 1% in the last year of simulation. The increase of inbreeding was similar in all the scenarios with the exception of sIL (p -value > 0.05). The scenario sIL showed the lowest inbreeding. This shows that selection on IL favoured matings that are less related, leading to a reduction in the inbreeding level of the upcoming generations compared to the other scenarios. More complex mating strategies, such as sMA and sEFI, did not show any advantage in reducing inbreeding (respect to selection on EBVs).

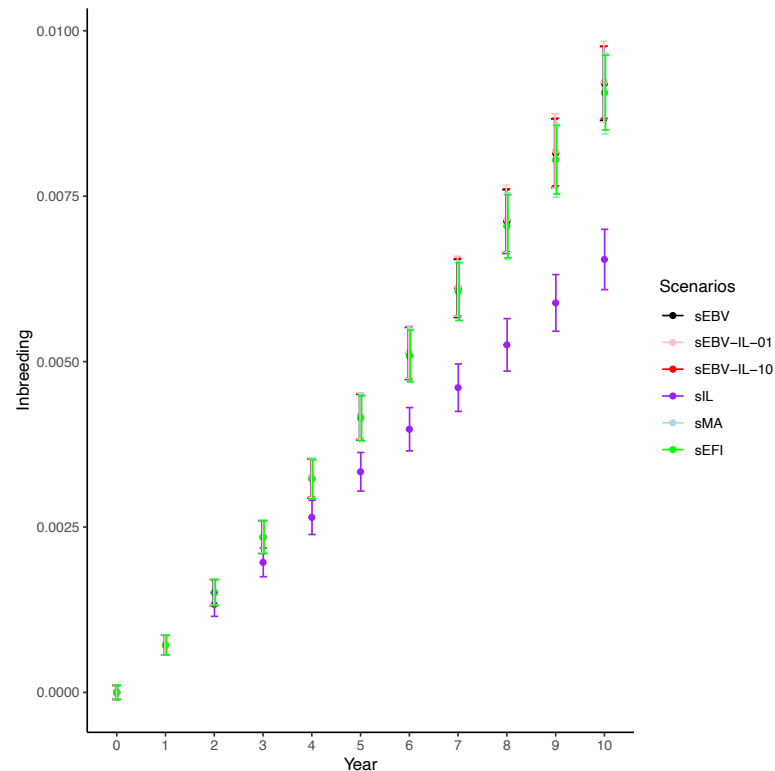


Figure 7. Evolution of the mean of inbreeding coefficients during 10 years of selection in the 6 different scenarios. sEBV: scenario selecting based on EBV; sEBV-IL-01: scenario selecting based on EBV + 0.01IL; sEBV-IL-10: scenario selecting based on EBV + 0.1IL; sIL: scenario selecting based IL; sMA: scenario implementing mate allocation; sEFI: scenario selecting based on EBV adjusted by the expected future inbreeding.

To understand more the difference in the inbreeding between the scenarios involving both, EBV and IL (sEBV-IL-01, sEBV-IL-10, sMA) and only IL (sIL), we checked the partial inbreeding coefficient and the number of ancestors generating inbreeding. Three scenarios (sEBV-IL-01, sEBV-IL-10 and sMA) had around the same number of partial inbreeding coefficients (Table 2), while sIL had the lowest number of partial inbreeding coefficients with the highest number of ancestors generating them (9K instead of 7K for the other scenarios). The number of partial inbreeding coefficient per ancestor was lower in sIL and explained the lower inbreeding recorded in the sIL (Table 2).

Table 2. Mean and number of partial inbreeding coefficients and the number of ancestors generating inbreeding in different scenarios.

Scenarios ¹	Number of partial inbreeding coefficients \pm se ²	Number partial inbreeding coefficients > 0.01 \pm se (%)	Number of ancestors generating inbreeding \pm se (%)	Average number of PF ³
sEBV-IL-01	14,400,624 \pm 272,977	28,531 \pm 1,252.7 (0.2%)	7,395 \pm 82.9	1,978
sEBV-IL-10	14,487,610 \pm 233,151	28,290 \pm 962.0 (0.2%)	7,323 \pm 104.0	1,947
sIL	11,756,042 \pm 174,884	22,373 \pm 990.3 (0.2%)	9,283 \pm 96.9	1,266
sMA	14,338,185 \pm 298,147	28,309 \pm 1,136.6 (0.2%)	7,546 \pm 36.0	1,774

¹scenarios: sEBV: scenario selecting based on EBV; sEBV-IL-01: scenario selecting based on EBV + 0.01IL; sEBV-IL-10: scenario selecting based on EBV + 0.1IL; sIL: scenario selecting based IL; sMA: scenario implementing mate allocation; sEFI: scenario selecting based on EBV adjusted by the expected future inbreeding.

²se: standard error.

³PF: partial inbreeding per ancestor.

Inbreeding depression for 100% of inbreeding is presented in Table 3. This means that a 10% increase in inbreeding would result in a reduction of 10.7, 10.6, 10.6, 10.5, 10.5 and 10.7 liters of milk yield in sEBV, sEBV-IL-01, sEBV-IL-10, sIL, sMA and sEFI, respectively. The differences between the scenarios were not significant. Selection against low IL (the fraction of the mutation load that is due to hidden recessive alleles in heterozygous state) did not reduced inbreeding depression compared to other scenarios. We expected a lower inbreeding

depression for the sIL scenario. However, the IL presented in the population, when exposed by inbreeding (even if it was lower than in the other scenarios), was responsible for inbreeding depression in milk yield. This result showed that selecting individuals on IL alone did not prevent or reduce the inbreeding depression on the trait.

The rate of inbreeding per year (ΔF_{year}) was calculated from the regression of the average inbreeding coefficient on the year of birth for the 10 years of simulation. Only the ΔF_{year} in sIL was significantly different (lower) from other scenarios (Table 3). Relying on previous researches results (Fu et al., 1998; Hedrick, 1994) they showed that the ability of a population to purge deleterious alleles is impacted by the rate at which inbreeding accumulate within the population overtime. It was proved in previous long-term selection experiments (~200 generations) in *Drosophila* strains conducted by Latter et al. (1995), they showed that the higher the rate the inbreeding the slower is the removal of the deleterious alleles. Here, we selected against hidden recessive alleles (low IL); it is as an artificial purge and for this reason the rate of inbreeding is lower in sIL. Lower ΔF compared to other scenarios was in agreement with the higher number of ancestors (9,283 ancestors generating inbreeding) found in this study (Table 2) (Woolliams, 2007). As expected, N_e in sIL was significantly different form the other scenarios.

Table 3. Inbreeding depression, rate of inbreeding per year (ΔF_{year}) and effective population size (N_e) for the six scenarios in the last year of simulation.

	sEBV	sEBV-IL-01	sEBV-IL-10	sIL	sMA	sEFI
ID*	-107.1 ± 3.6	-105.8 ± 2.6	-105.9 ± 2.3	-105.2 ± 3.5	-104.7 ± 3.5	-106.7 ± 2.6
ΔF_{year}	2.42×10 ⁻³ ± 1.51×10 ⁻⁴ ^a	2.41×10 ⁻³ ± 1.70×10 ⁻⁴ ^a	2.41×10 ⁻³ ± 1.48×10 ⁻⁴ ^a	1.28×10 ⁻³ ± 1.20×10 ⁻⁴ ^b	2.27×10 ⁻³ ± 1.69×10 ⁻⁴ ^a	2.29×10 ⁻³ ± 1.60×10 ⁻⁴ ^a
N_e	52 ^a	52 ^a	52 ^a	98 ^b	55 ^a	55 ^a

*ID: inbreeding depression expressed by completely inbred (100%) descendants.

sEBV: scenario selecting based on EBV; sEBV-IL-01: scenario selecting based on EBV + 0.01IL; sEBV-IL-10: scenario selecting based on EBV + 0.1IL; sIL: scenario selecting based IL; sMA: scenario implementing mate allocation; sEFI: scenario selecting based on EBV adjusted by the expected future inbreeding.

^{a,b} Values with different lowercase letters denote statistically significant differences between scenarios ($p < 0.05$).

Genetic correlation

In the simulation, the genetic correlation between additive genetic and inbreeding load effects was assumed to be -0.1, based on the previous estimates obtained for the dairy sheep. In the last year, the bivariate plot between EBV and estimates of the inbreeding load effects (Figure 9) permitted to see the degree and pattern of relation between both. In accordance with the theory presented in our previous work (Antonios et al., 2024), the correlation between the EBV for milk yield and the estimates of the inbreeding load was negative in all the scenarios. It was not possible to analyse the relationship between the true breeding value and the true inbreeding load, because the last one was not explicitly simulated in this work.

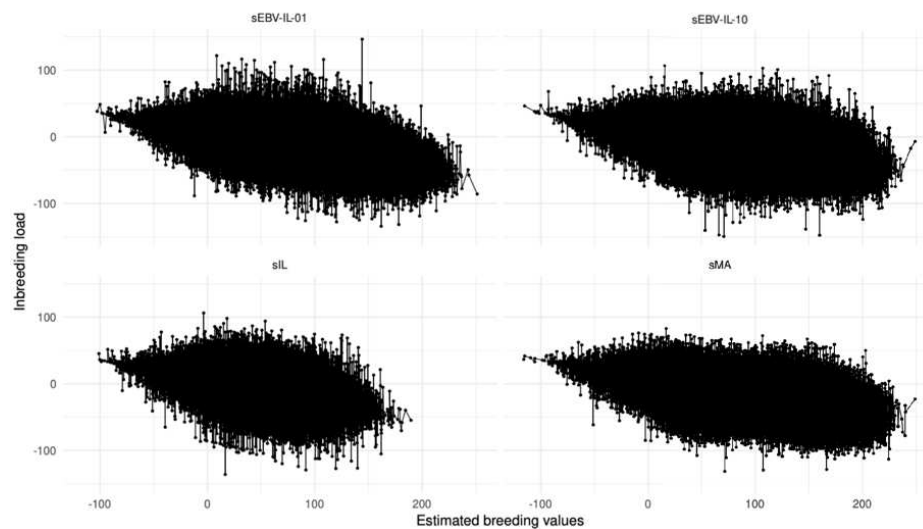


Figure 9. Bivariate plot showing the relationship between additive genetic and inbreeding load effects in four scenarios. sEBV-IL-01: scenario selecting based on EBV + 0.01IL; sEBV-IL-10: scenario selecting based on EBV + 0.1IL; sIL: scenario selecting based IL; sMA: scenario implementing mate allocation.

The accuracy of estimating inbreeding load effects was very low; for instance it was -13.46 (111.77) for sEBV-IL-01 scenario. This was in accordance with other studies (Varona et al., 2019; Varona et al., 2022; Antonios et al. 2024). Varona et al. (2022) showed in their study that the use of SNP markers can improve the accuracy of the estimation of the inbreeding load effects in selection candidates. But, in French dairy sheep breeds, even if genomic selection was introduced in 2016, up to now, there are not enough genotyped animals for that purpose, especially that all females are not genotyped.

Conclusion

To conclude, we showed that selecting animals using inbreeding load is feasible. However, the magnitude of inbreeding load effects and its accuracy did not show a clear benefit of including it in genetic evaluation models. Furthermore, the benefit of using inbreeding load in mate allocation strategies was negligible. In light of this, those results could be used as a reference for further research in other species and traits to explore the possible benefits of these genetic management strategies as well as to test the alternatives.

References

- Aliloo, H., Pryce, J. E., González-Recio, O., Cocks, B. G., Goddard, M. E., and Hayes, B. J. (2017). Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. *J. Dairy Sci.* 100, 1203–1222. doi:10.3168/jds.2016-11261
- Antonios, S., Legarra, A., Pong-Wong, R., Astruc, J. M., Rodríguez-Ramilo, S. T., and Vitezica, Z. G. (2023). Partitioning of the genetic trends of French dairy sheep in Mendelian samplings and long-term contributions. *J. Dairy Sci.* 106, 6275–6287. doi:10.3168/jds.2022-23009
- Baloche, G., Legarra, A., Sallé, G., Larroque, H., Astruc, J.-M., Robert-Granié, C., et al. (2014). Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. *J. Dairy Sci.* 97, 1107–1116. doi:10.3168/jds.2013-7135
- Bengtsson, C., Stålhammar, H., Thomassen, J. R., Eriksson, S., Fikse, W. F., and Strandberg, E. (2022). Mating allocations in Nordic Red Dairy Cattle using genomic information. *J. Dairy Sci.* 105, 1281–1297. doi:10.3168/jds.2021-20849
- Berkelaar, M., Eikland, K., and Notebaert, P. (2004). Ipsolve: Open source (mixed-integer) linear programming system. *Eindhoven U. of Technology* 63.
- Casellas, J. (2018). On individual-specific prediction of hidden inbreeding depression load. *J. Anim. Breed. Genet.* 135, 37–44. doi:10.1111/jbg.12308

- Casellas, J., Varona, L., Ibáñez-Escriche, N., Quintanilla, R., and Noguera, J. L. (2008). Skew distribution of founder-specific inbreeding depression effects on the longevity of Landrace sows. *Genet. Res.* 90, 499–508. doi:10.1017/S0016672308009907
- Chen, G. K., Marjoram, P., and Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. *Genome Res.* 19, 136–142. doi:10.1101/gr.083634.108
- Fu, Y.-B., Namkoong, G., and Carlson, J. E. (1998). Comparison of Breeding Strategies for Purging Inbreeding Depression via Simulation. *Conserv. Biol.* 12, 856–864. doi:10.1111/j.1523-1739.1998.97057.x
- Gaynor, R. C., Gorjanc, G., and Hickey, J. M. (2021). AlphaSimR: an R package for breeding program simulations. *G3 Genes, Genomes, Genet.* 11, jkaa017. doi:10.1093/g3journal/jkaa017
- González-Diéguez, D., Tusell, L., Bouquet, A., Legarra, A., and Vitezica, Z. G. (2020). Purebred and crossbred genomic evaluation and mate allocation strategies to exploit dominance in pig crossbreeding schemes. *G3 Genes, Genomes, Genet.* 10, 2829–2841. doi:10.1534/g3.120.401376
- Hedrick, P. W. (1994). Purging inbreeding depression and the probability of extinction: full-sib mating. *Heredity* 73, 363–372. doi:10.1038/hdy.1994.183
- Hill, W. G. (1974). Linkage disequilibrium between neutral genes in finite populations. *Adv. Appl. Probab.* 6, 13–15. doi: 10.2307/1426201
- Jansen, G. B., and Wilton, J. W. (1985). Selecting Mating Pairs with Linear Programming Techniques. *J. Dairy Sci.* 68, 1302–1305. doi:10.3168/jds.S0022-0302(85)80961-9
- Latter, B. D., Mulley, J. C., Reid, D., and Pascoe, L. (1995). Reduced genetic load revealed by slow inbreeding in *Drosophila melanogaster*. *Genetics* 139, 287–297. doi:10.1093/genetics/139.1.287
- Lourenco, D., Tsuruta, S., Aguilar, I., Masuda, Y., Bermann, M., Legarra, A., et al. (2022). “Recent updates in the BLUPF90 software suite”, in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production*, (Rotterdam), 1530–1533. doi:10.3920/978-90-8686-940-4_366

- Lv, F.-H., Cao, Y.-H., Liu, G.-J., Luo, L.-Y., Lu, R., Liu, M.-J., et al. (2021). Whole-Genome Resequencing of Worldwide Wild and Domestic Sheep Elucidates Genetic Diversity, Introgression, and Agronomically Important Loci. *Mol. Biol. Evol.* 39, msab353. doi:10.1093/molbev/msab353
- Martinez-Castillero, M., Varona, L., Pegolo, S., Rossoni, A., and Cecchinato, A. (2021). Bayesian inference of the inbreeding load variance for fertility traits in Brown Swiss cattle. *J. Dairy Sci.* 104, 10040–10048. doi:10.3168/jds.2020-20087
- Meadows, J. R. S., Chan, E. K. F., and Kijas, J. W. (2008). Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genet.* 9, 61. doi:10.1186/1471-2156-9-61
- Meuwissen, T., and Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* 24, 305. doi:10.1186/1297-9686-24-4-305
- Perdomo-González, D. I., Molina, A., Sánchez-Guerrero, M. J., Bartolomé, E., Varona, L., and Valera, M. (2021). Genetic inbreeding depression load for fertility traits in Pura Raza Española mares. *J. Anim. Sci.* 99, 12. doi:10.1093/jas/skab316
- Petit, M., Astruc, J.-M., Sarry, J., Drouilhet, L., Fabre, S., Moreno, C. R., et al. (2017). Variation in Recombination Rate and Its Genetic Determinism in Sheep Populations. *Genetics* 207, 767–784. doi: 10.1534/genetics.117.300123
- Poyato-Bonilla, J., Perdomo-González, D. I., Sánchez-Guerrero, M. J., Varona, L., Molina, A., Casellas, J., et al. (2020). Genetic inbreeding depression load for morphological traits and defects in the Pura Raza Española horse. *Genet. Sel. Evol.* 52, 62. doi:10.1186/s12711-020-00582-2
- Rodríguez-Ramilo, S. T., Elsen, J. M., and Legarra, A. (2019). Inbreeding and effective population size in French dairy sheep: comparison between genomic and pedigree estimates. *J. Dairy Sci.* 102, 4227–4237. doi:10.3168/jds.2018-15405
- Toro, M. A., and Varona, L. (2010). A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.* 42, 33. doi:10.1186/1297-9686-42-33

- VanRaden, P. M., and Smith, L. A. (1999). Selection and Mating Considering Expected Inbreeding of Future Progeny. *J. Dairy Sci.* 82, 2771–2778. doi: 10.3168/jds.S0022-0302(99)75534-7
- Varona, L., Altarriba, J., Moreno, C., Martínez-Castillero, M., and Casellas, J. (2019). A multivariate analysis with direct additive and inbreeding depression load effects. *Genet. Sel. Evol.* 51, 78. doi:10.1186/s12711-019-0521-3
- Villa-Angulo, R., Matukumalli, L. K., Gill, C. A., Choi, J., Van Tassell, C. P., and Grefenstette, J. J. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genet.* 10, 19. doi:10.1186/1471-2156-10-19
- Varona, L., López Carbonell, D., Srihi, H., Ramírez, M., Martínez-Cambor, P., Hermida, M., et al. (2022). “Genetic variability in the individual inbreeding load: genomic prediction for artificial purging”, in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production*, (Rotterdam), 786–789. doi: 10.3920/978-90-8686-940-4_183
- Woolliams, J. (2007). Genetic Contributions and Inbreeding. *Util. Conserv. Farm Anim. Genet. Resour.*, 147–165

Chapter 5

General discussion and perspectives

Livestock breeding is always aimed at the future and breeding programs are set up to create genetic improvement. The improvement of the average performance, in livestock populations, requires selecting individuals with average additive genetic value better than the average of the population. In most cases, the methodological focus on EBV prediction may hide the role of Mendelian sampling in creating genetic progress. A sustained genetic gain is related to the use of the Mendelian sampling variation – i.e. the new genetic variation generated at each meiosis (Woolliams *et al.*, 1999). In the first part of my PhD, we fine partitioned genetic trend into Mendelian sampling components, by categories of animals defined by sex and by selection pathways, and we also characterized long-term genetic contributions in order to identify the different contributions of categories of individuals, and for a better understanding of the selection scheme.

Even if maximizing the genetic gain is the target in selection schemes, it is important to limit the increase of coancestry and inbreeding, and different strategies have been developed to deal with this. Inbreeding, which is unavoidable in finite size populations, changes genotype frequencies by increasing the homozygosity at the expense of the heterozygous genotypes (Curik *et al.*, 2014). The inbreeding load is the fraction of the mutation load which is due to recessive variants concealed in heterozygous condition. In populations where recessive mutations (inbreeding load) are masked by the heterozygosity state, inbred matings can increase the likelihood of expressing those mutations, because they will be in a homozygote state in the offspring (Caballero, 2020). During the second part of my PhD, we proved the additive nature of the inbreeding load and developed a, to our knowledge, original theory that shows explicitly the genetic correlation of inbreeding load with the additive genetic effect. Then, we applied this theory, we estimated the genetic correlation for the trait milk yield, and evaluated the effectiveness of involving inbreeding load in selection strategies by simulation.

The objective of this chapter is to discuss the main findings of this thesis in a broader and joint perspective, highlighting the strengths and weaknesses of the models and strategies used and guiding future research.

5.1 Partition of the genetic trends

5.1.1 Selection pathways

In the process of breeding and selection, genes may be transmitted to the next generation using different pathways. Four selection pathways are taken into consideration: (1) sires of sires (SS) are the sires of the AI sires, (2) sires of dams (SD), these sires will be the fathers of dams, (3) dams of sires (DS) are the mothers of sires, and (4) dams of dams (DD). SS is the most stringent selection path to breed new SS. Within the sire category, SD, is the less stringent selection path. DS are the most stringent selection path within the dams and they produce mainly the AI sires. Among all paths, DD is the less stringent selection path and they produce dams (although they are by no means unimportant, e.g. dams can be the maternal grand-mother of an AI sire). Selection response can be divided into a number of selection paths, the number depending on the number of differences in selection intensity and the accuracy of selection.

In French dairy sheep breeds, all four selection paths can be recognised (Astruc *et al.*, 2022). SS, elite sires, are the ones with the highest total merit index (named ISOL), and they are used for AI mainly in the selection nucleus and they are mated in assortative mating with DS to produce the future AI sires. The best SD are used for AI to produce the future dams, and the rest of sires are used for the diffusion of the AI mainly in the stockbreeders (practicing, or not, simplified milk recording), the later are known as the “users of genetic progress” but do not contribute selected animals themselves. The ISOL of SD is lower than that for SS. The passage of genes from the nucleus to the “out of the nucleus” flocks is based partly on AI sires and partly on natural mating sires by sons of AI sires. To estimate the annual genetic gain (AGG), we can use the classical theory of response to selection (Rendel and Robertson, 1950) and take into account those selection pathways as follows: $AGG = \frac{\Delta G_{SS} + \Delta G_{SD} + \Delta G_{DS} + \Delta G_{DD}}{L_{SS} + L_{SD} + L_{DS} + L_{DD}}$, where L is the average age of the selected individuals in each selection pathway when their progeny was born, and ΔG is the estimated genetic superiority of the selected group over the contemporary individual born in the same year (Astruc *et al.*, 2022). The selection pathways methodology permits a vision of the flow of genes from each category to the next generation and of the effort of selection each year. Genetic improvement in dairy sheep is the result of the

joint effort of breeders and research (chapter 1, Figure 1.3). For example, in LAC, between 2000 and 2014, the estimated yearly AGG (expressed in equivalent litres) is 5.23 in ewes. After 2015 (after the implementation of genomic selection), the AGG is 7.96 in ewes (Astruc *et al.*, 2022).

5.1.2 Partitioning of the genetic trend

To improve a population, we need certain animals to be better than their parents. For this, the originality of the animals with respect to their parents (i.e., Mendelian sampling) has to be captured. In chapter 2, we showed that breeding values of any generation can be partitioned into a sum of ancestors' Mendelian sampling terms weighted by their long-term genetic contribution and individuals' Mendelian sampling terms (Bijma, 2000). This partition is useful, because it shows that long-term contribution represents the contribution of the individual's unique bit of genetic variation, Mendelian sampling, to the long-term gene pool. This helps to understand that the future gene pool has contributions from the founders and non-founders. Additionally, this chapter outlined a procedure to partition the genetic response into smaller parts that align with different categories of animals defined by sex and selection pathways. This method enabled the calculation of the partial genetic trends directly from the EBV according to different groups (e.g. selection pathway or sex) providing a way to inspect the contribution of each category to the genetic gain. Note that by doing this, we have split, for each individual, the parent average into contributions due to Mendelian sampling of their ancestors, which, in turn, then we assign into categories. Results obtained in chapter 2 highlighted the difference in the contribution of different categories of animals to better understand the selection decisions. Although the precision in capturing Mendelian sampling is greater in males (by progeny testing), the final genetic progress is largely given by the originality of the females, even if their prediction (by own phenotype) is less accurate and their progeny number is smaller. This study allows us to say that the females (dams) play an important role in the total genetic gain of the population by their originality (or genetic variabilities). Usually, this role of the dams is being neglected by the importance of the contribution of the AI males to the total genetic merit of the population. But in fact, half of the AI males' genes come from their dams and which also include half of the Mendelian sampling term of their dams, which means that dams' positive Mendelian sampling are eventually selected and spread through the AI. Additionally, the long-term genetic contribution of the animal comes from the sum of its males and females' offspring selected, and this further

justifies the importance of females in participating in the genetic progress (Woolliams *et al.*, 1999). The fine partitioning of the genetic gain in Mendelian sampling trends can be interpreted as females being a large and easily accessible reservoir of genetic variability for selection. Still, AI sires are crucially important in the diffusion of best genetics in the population and in connecting genetically the flock (nucleus and commercial flock) through the AI. With the small number of AI sires compared to the females' number, they have higher contribution for the upcoming years and that is because of the use of AI. Use of AI means that AI sires contribute more than dams to the gene pool of the population due to their larger number of offspring. Additionally, there were differences between LAC breed and the Pyrenean breeds in terms of the contribution to genetic progress of females versus males. In LAC, females had higher contribution to the overall genetic trend than the males, whereas in the Pyrenean breeds both sexes had around the same contribution to the overall genetic trend. The difference in the relative importance of Mendelian sampling and the parent average between breeds should reflect the different selection strategies in practice. In LAC breed, Mendelian sampling was more important than parent average to determine the selection of individuals and their long-term contributions. Since Mendelian sampling is related to within-family selection, we can say that a greater importance of within-family selection is given in LAC breed, and this greater importance may be due to a larger family size than in the other breeds. Additionally, the greater offspring numbers of LAC AI males make possible a more intense within-family selection. Whereas in the Pyrenean breeds, the effect of parent average is more important than the effect of Mendelian sampling, and the between family selection is more important. Long-term contributions were larger for AI males (with larger progeny sizes than females) and in BB than in LAC (with the latter being a larger population).

In the selection pathway part, we showed the formula of Rendel and Robertson (1950) for estimating the AGG and used by Astruc *et al.* (2022). This method considers the gain from selection in a single cycle of selection arising from all previous cohorts. The genetic improvement that results from a single year of selection in a group of animals may take many years to pass through the population when generations overlap. The rates of response predicted by the classical theory are therefore reached only asymptotically.

Hill (1974) pointed out that the effect of one year of selection on the successive generations is mostly erratic for many years after the selection is practised. Hill (1974) proposed an alternative approach based on a matrix which specify the passage of genes (gene

flow) between different groups. In our work, we used the gene flow approach of Woolliams *et al.* (1999), which is a modified version of the conventional gene flow proposed by Hill (1974). In the modified gene flow approach, they have defined a modified gene flow matrix (\mathbf{G}) that represents the parental origin of the genes of selected individuals, and this matrix \mathbf{G} takes into account the selection degree that is taking place changes and may vary with e.g. age. This modified gene flow accounts for individual long-term genetic contributions [by including $\beta_k(u_{x(k)} - \bar{u}_k)$ in the model for expected contributions], where β is the regression of the contribution of individual x in a category k on its breeding value $u_{x(k)}$, and \bar{u}_k the mean breeding value of selected contemporaries of x in category k . Whereas, the conventional gene flow accounts for the average genetic contribution ($\beta = 0$) (Bijma and Wolliams, 1999). Following this approach, by partitioning the breeding values into Mendelian sampling terms, Woolliams *et al.* (1999) defined the rate of genetic gain (ΔG) for an annual cohort as the product of the genes that derive from an individual and its Mendelian sampling term summed over all parents per year as:

$$\Delta G = \sum t_x \phi_x \quad (5.1)$$

where t_x is the proportion of genes deriving from an individual x (the long-term contribution of x) and ϕ_x is the Mendelian sampling of individual x , we can define the genetic gain by category (Woolliams *et al.*, 1999). This equation shows the long-term impact of the individual on the population mean, and it traces the flow of genes through the pedigree (Woolliams *et al.*, 1999). ΔG is calculated with the Mendelian sampling and not the breeding values because the breeding value of an individual contains its Mendelian sampling and its ancestors' Mendelian sampling, which means that substituting Mendelian sampling with breeding values will be double counting. Thus, with this equation the higher the genetic contribution of a group or category, the higher the genetic gain. By using this approach, we have found that the gene pool in dairy sheep populations has contributions from females and males. Males had higher contribution than females in both breeds, and this is because of their smaller number compared to the dams and they contribute more to the gene pool by their higher number of offspring. From this study, we confirm that the Mendelian sampling term was observed to be the most important factor determining the selection of the females to become dams and to maintain their long-term genetic contribution.

Why using Woolliams *et al.*'s approach (1999)? - On the light of equation 5.1, it considers that the genetic gain is coming from a cohort over all subsequent cycles of selection, unlike Rendel and Robertson's equation that considers gain from selection in a cycle coming from all previous cohorts (which results in double counting). Woolliams *et al.*'s approach (1999) shows how asymptotic proportions of genes can be predicted accurately, either on an individual or on a group level. According to Wray *et al.* (1989), accounting for the effect of selection between individuals is crucial for prediction the rates of inbreeding. Bijma and Woolliams (1999) pointed out that predicting the rate of inbreeding is not possible with the conventional gene flow method. In selected populations, the accuracy of the prediction of long-term genetic contributions is a crucial step for predicting the rate of inbreeding (Bijma and Woolliams, 1999). Thus, Woolliams *et al.*'s approach (1999) is suitable for calculating the rates of inbreeding.

5.1.3 Link between partition of genetic trend and the evolution of inbreeding.

Genetic diversity within a population is maintained by reducing the rate of inbreeding or coancestry, or by increasing the N_e of a population (de Cara *et al.*, 2013). In the ideal scenario, population genetics theory recommends keeping equal numbers of males and females and maintaining a constant population size over time. In such case, having the same contributions for males and females (each couple of animals contributes to one male and one female) would maximize the N_e . But in livestock populations, it is impossible to maintain a 1:1 sex ratio and we have far fewer males (e.g. 11,230 males in MTR) than females (e.g. 622,425 females in MTR), which impacts the N_e of a population and thereafter the rate of inbreeding (Howard, 2017). Furthermore, some families are superior concerning the selected trait (e.g. milk yield) compared to the other families. Thereafter, the contributions among groups are different and the rate of inbreeding is higher than in randomly mated populations (Wray and Thompson, 1990). Thus, it is important to monitor the rate of inbreeding in populations undergoing selection. Wray and Thompson (1990) demonstrated that there is a relation between the rate of inbreeding per generation and the long-term contributions, and this relation was extended by Woolliams and Bijma (2000) as:

$$\Delta F = \frac{1}{4}(1 - c)\sum t_x^2 \quad (5.2)$$

where c measures the departure from fully random mating; being c positive for matings between relatives and negative in the case of avoidance of relatives; and the sum is taken over contributions t of all the individuals. From this equation, we can see that to minimize ΔF , we have to increase the number of ancestors that will contribute, and thereafter the average contribution will decrease together with ΔF . The long-term contributions that will develop over time are clearly related to ΔF . This reveals the necessity to control by selection, and by good choosing and mating systems, how the long-term contributions of individuals will develop over time for better management of ΔF (Woolliams, 2007). The long-term genetic contribution is a measure of the level of relatedness between animals in a population because of a shared common ancestor. Equations (5.1 and 5.2), show the relationships between genetic gain (ΔG), the rate of inbreeding (ΔF) and long-term genetic contribution. It is important to improve the genetic gain whilst managing the rate of loss of genetic variation. Which means that in our study, the attention not to increase ΔF is mostly concentrated on the AI sires, and on both dams' categories because they have the higher genetic contribution and the higher genetic gain (chapter 2). The long-term genetic contributions of individuals is not stable, it changes with time because the genetic merit of their progeny changes with time. Moreover, if a family or category had a high contribution compared to the other families, this can be controlled in the following years of selection to limit their contribution in a way to keep the genetic gain but at the same time maintaining the rate of inbreeding in acceptable limits. Since the long-term contribution of parents is half of the long-term contribution of the offspring, a change in the contributions of those later will cause a change of the parental long-term contribution (Woolliams, 2007). To safe guard the long-term selection response and fitness, the rate of inbreeding needs to be maintained to a certain level in a way to preserve the genetic variation and keep the accumulation of mutational variation (e.g., Lynch and Hill, 1986; Keightley and Hill, 1987). With time and after many generations, more progeny will be produced, causing a change in the long-term genetic contribution of individuals which mean that for the groups that have high long-term genetic contributions (AI sires, dams) can be managed and controlled in a way to minimize the rate of inbreeding while improving the genetic gain. This helps to maintain the genetic variation of the population in the long-term, by maintaining the rate of inbreeding to an acceptable level (Woolliams, 2007). Our results showed the low rate of inbreeding in the studied breeds (average inbreeding of 1% of in the whole population).

5.2 Approaches to control inbreeding

Different approaches have been developed in livestock to control the increase in inbreeding resulting from selection programs, we will discuss about some of them.

5.2.1 *Optimum contribution selection*

The Optimum Contribution Selection (OCS, Meuwissen, 1997; Grundy *et al.*, 1998; Colleau *et al.*, 2004) method takes into consideration the genetic merit of candidates and their relationship in order to determine the optimum number of progenies for each candidate. The OCS strategy maximizes the average breeding value of the selected individuals while restricting their average genetic relationships to a predefined level. The restriction is on the average relationship between selected animals, because the increase in average relationship approximately equals the future increase in inbreeding. The aim of OCS is to increase the genetic gain to let it reach a stable rate while increasing or maintaining the genetic diversity. Therefore, the OCS could control the short and long-term effects of selection on inbreeding. Compared to the truncation selection, OCS highlights the managing between and within family (Mendelian sampling) variation (Howard *et al.*, 2018), which gives the advantage of controlling the total genetic variance. Therefore, if the average relationships of the selected parents are controlled, the change in inbreeding in the future generations is also controlled, and this is the primary mechanism behind OCS (Howard, 2017). The OCS method can control the total genetic variance, but it does not take into account recessive alleles explicitly. The reproduction in dairy sheep breeds is done based on AI with fresh semen, making it difficult to apply OCS methods (Macedo, 2020).

5.2.2 *Expected future inbreeding*

To achieve maximum genetic progress while preserving genetic diversity, it is essential to identify and select sires that are least related to the dams being mated. VanRaden and Smith (1999) introduced the concept of Expected Future Inbreeding (EFI) to achieve this. Expected Future Inbreeding is the average inbreeding produced by an animal in the next generation. The EFI is an individual based estimation of the anticipated level of inbreeding in a population. By considering the EFI, breeders can make informed decisions to minimize inbreeding and to

assess and reduce its impact on the future generation. With random mating, the EFI of an animal is the average inbreeding of all possible offspring, i.e. half of the mean relationship of the animal to its contemporary mates. The resulting estimates can then be used to adjust the breeding values for each trait of potential selection candidates that are either highly or lowly related to individuals in the whole population. The EFI provides a premium to animals which have a lower relatedness, and a penalty to more related animals. Preventing high levels of inbreeding by managing matings is important to reduce the number of lethal recessive alleles and genome homozygosity in the offspring (Howard *et al.*, 2017). The definition of EFI is simple and its implementation through the adjusted EBV is straightforward (e.g. in dairy cattle). Expected inbreeding of future progeny helps breeders to avoid inbreeding. In chapter 4, we used the concept of adjusted EBV (by adjusting EBV by the EFI) as a selection criterion, in one of the scenarios of our simulation. EFI was used on the waiting sires (potential AI sires after the progeny test) with all the contemporary females. However, in 10 generations of selection, similar inbreeding was observed in the scenarios with and without EFI.

5.2.3 *Inbreeding load*

Inbreeding Load (IL) is the fraction of the mutation load that is due to recessive variants, which can be hidden in heterozygous condition. The IL of individuals can equivalently be presented as a heritable additive trait that is only expressed when inbreeding occurs in their offspring (Varona *et al.*, 2019). This trait have a favourable or an unfavourable effect on the studied trait (e.g. milk yield) (Martinez-Castillero *et al.*, 2021). Inbreeding load of individuals can be predicted in the same manner that we do for additive genetic values, using linear models (Casellas, 2018; Varona *et al.*, 2019). This requires the split of total inbreeding coefficients into fractions due to ancestors – partial inbreeding coefficients. In chapter 3, first, we computed these partial inbreeding coefficients due to ancestors in the three Pyrenean dairy sheep breeds. Second, we used these partial inbreeding coefficients as covariates in a random regression mixed model to estimate genetic variance and breeding values of the IL for milk yield in the three breeds. Our analyses were the most important in the literature, in terms of number of animals in the pedigree and number of records. We had around 633,655 animals in MTR versus 73,246 animals in the pedigree in the paper of Martinez-Castillero *et al.* (2021) and 384,434 animals in the work of Varona *et al.* (2019). In terms of number of records, we had 2,168,454

records of milk yield in MTR versus 75,194 records in Varona *et al.* (2019) and 59,864 records in Martinez-Castillero *et al.* (2021). Additionally, an available online Fortran program was made for our study because the available programs from the literature (R program in Martinez-Castillero *et al.* (2021) and Varona *et al.*'s code (2019)) were unable to deal with our large data.

Adding the IL effect to the model improved the fitting (values of the statistic likelihood ratio test between 132 to 383) for milk yield in the three breeds. Estimates of the IL variances were different from zero except for BB. Estimates of the genetic correlation between the additive genetic value and the inbreeding load effect were negative (as expected from the theory presented in chapter 3) and low (~ -0.1) for milk yield in MTN and MTR. Our estimates were comparable to previous studies, where genetic correlations were close to zero in Brown Swiss dairy cattle for days open (Martinez-Castillero *et al.*, 2021) and for weaning weight in the Rubia Gallega beef cattle breed (Varona *et al.*, 2019). Compared to Varona *et al.* (2019), our estimates were also negative but much lower than the values obtained for Pirenaica beef cattle (~ -0.4). A negative correlation means that animals with high (desirable) breeding values for the trait tend to cause negative (undesirable) inbreeding depression affecting the same trait if their descendants are inbred (Varona *et al.*, 2019). But, in our case, our correlation estimates close to zero imply that the IL are not genetically correlated with the additive genetic effect of milk yield. This implies that selection for milk yield will not cause increase inbreeding depression in milk yield in inbred animals. Artificial purging based on predicted IL effects could be performed to reduce the effect of inbreeding depression as suggested by Varona *et al.* (2019) and Martinez-Castillero *et al.*, (2021). In this case individuals could be selected based on both their EBV for the trait and their EBV for IL. We analysed this scenario of selection by simulation in chapter 4. Because selecting individuals based on predicted IL for milk yield would basically remove recessive alleles that reduce milk yield in homozygote carriers, we expected a reduction of the inbreeding depression and higher genetic gain in those scenarios. However, this was not the case, we had the same results than in the conventional scenario (selection based only on EBV). This may be due to the fact that the prediction of IL effect was not accurate.

5.2.3.1 *Mate allocation*

In another scenario of the simulation, we included the IL in a mate allocation strategy for avoiding undesirable mating. To our knowledge, IL predictions have not been used in mate

allocation strategies elsewhere. In animal breeding, mate allocation has been proposed with different objectives (Howard, 2017): to control the inbreeding of the progeny by using pedigree information to avoid matings between related animals (VanRaden and Smith, 1999; Kinghorn, 2011), to exploit dominance variation either within-breed or/and across-breeds (DeStefano and Hoeschele, 1992; Hayes and Miller, 2000; Toro and Varona, 2010; González-Diéguez *et al.*, 2020), and to increase connectedness among herds. Here, we prioritized matings between individuals with no common ancestors, and then with ancestors with favourable IL (positive IL for milk yield); and we avoided matings with common ancestors with low and negative IL (that can generate worse inbreeding depression in inbred offspring). The genetic gain was similar as the conventional selection scenario. Selecting animals using IL was feasible. However, the small magnitude of IL effects and its low accuracy showed that there is not a clear interest in using it in genetic evaluation and selection. Furthermore, the benefit of using IL in mate allocation strategies was negligible.

5.3 General conclusions

In this study, we partitioned the genetic gain in Mendelian sampling by categories of animals defined by sex and by selection pathways, and explored long-term genetic contributions. This allowed identification of the contributions of different categories of individuals, and better understanding of the selection scheme. These results highlighted the importance of the dams in the contribution to the final genetic gain by their originality. The Mendelian sampling term was observed to be the most important factor determining the selection of a female to become dam and maintaining genetic contributions over time.

An accurate estimate of IL is an issue, in this work we used pedigree data to estimate inbreeding load whereas the use of SNP markers could be more accurate (Varona *et al.*, 2022). Further studies are needed. We did not use genomic data because genomic selection was introduced in 2016 which means that there were not enough animals genotyped for that purpose, and we needed females to be genotyped which is not the case in our dairy sheep breeds.

Estimates of variances of IL were low, and even if there was no warranty about selection criterion, we explored the possible benefits of involving inbreeding load predictions in

selection strategies. Predictions of IL effects for milk yield were used directly to select animals or in mating strategies. Mate allocation using IL effects was feasible but not interesting in terms of genetic gain or level of inbreeding. In addition, mate allocation would be complex and not easy to implement in the dairy sheep breeds. To conclude, selection based on IL (due to its variation and its magnitude) is not of practical interest.

References

- Astruc, J.M., G. Baloche, D. Buisson, J. Labatut, G. Lagriffoul, H. Larroque, C. Robert-Granié, A. Legarra, and F. Barillet. 2016. La sélection génomique des ovins laitiers en France. *INRAE Prod. Anim.* 29:41-56. <https://doi.org/10.20870/productions-animales.2016.29.1.2515>.
- Astruc, J. M., G. Lagriffoul, A. Legarra, and D. Buisson. 2022. Six years of genomic selection have increased the genetic gain in French dairy sheep. *Proc. 12th World Congr. Genet. Appl. Livest. Prod.*, Rotterdam, the Netherlands. https://doi.org/10.3920/978-90-8686-940-4_60.
- Avendaño, S., J.A. Woolliams, and B. Villanueva. 2004. Mendelian sampling terms as a selective advantage in optimum breeding schemes with restrictions on the rate of inbreeding. *Genet. Res.* 83:55-64. <https://doi.org/10.1017/S0016672303006566>.
- Barillet, F. 1997. Genetics of milk production. *The Genetics of Sheep*. Piper I., Ruvinsky A. CAB Int. 539-564.
- Bijma, P., Y. C. J. Wientjes, and M. P. L. Calus. 2018. Increasing genetic gain by selecting for higher Mendelian sampling variance. *Proc. 11th World Congr. Genet. Appl. Livest. Prod.*, Auckland, New Zealand.
- 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. <https://doi.org/10.1093/genetics/151.3.1197>.
- Bosse, M., H.J. Megens, M.F. Derks, M. A. R. de Cara, and M. A. Groenen. 2019. Deleterious alleles in the context of domestication, inbreeding, and selection. *Evol Appl.* 12:6-17. <https://doi.org/10.1111/eva.12691>.
- Caballero, A. 2020. *Quantitative genetics*. Cambridge University Press. <https://doi.org/10.1017/9781108630542>.

- Caballero, A., and M.A. Toro. 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet. Res.* 75:331-343. <https://doi.org/10.1017/S0016672399004449>.
- Carta, A., S. Casu, and S. Salaris. 2009. Invited review: Current state of genetic improvement in dairy sheep. *J. Dairy Sci.* 92:5814-33. <https://doi.org/10.3168/jds.2009-2479>.
- Casellas, J. 2018. On individual-specific prediction of hidden inbreeding depression load. *J. Anim. Breed. Genet.* 135:37-44. <https://doi.org/10.1111/jbg.12308>.
- Casellas, J., L. Varona, N. Ibáñez-Escriche, R. Quintanilla, and J.L. Noguera. 2008. Skew distribution of founder-specific inbreeding depression effects on the longevity of Landrace sows. *Genet. Res.* 90:499-508. <https://doi.org/10.1017/S0016672308009907>.
- Colleau, J.J., S. Moureaux, M. Briend, and J. Bechu. 2004. A method for the dynamic management of genetic variability in dairy cattle. *Genet. Sel. Evol.* 36:373. <https://doi.org/10.1186/1297-9686-36-4-373>.
- Curik, I., M. Ferenčaković, and J. Sölkner. 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest. Sci.* 166:26-34. <https://doi.org/10.1016/j.livsci.2014.05.034>.
- de Cara, M. A. R., B. Villanueva, M. A. Toro, and J. Fernández. 2013b. Purging deleterious mutations in conservation programmes: Combining optimal contributions with inbred matings. *Heredity.* 110:530-537. <https://doi.org/10.1111/mec.12560>.
- DeStefano, A. L., and I. Hoeschele. 1992. Utilization of dominance variance through mate allocation strategies. *J. Dairy Sci.* 75:1680-1690. [https://doi.org/10.3168/jds.S0022-0302\(92\)77925-9](https://doi.org/10.3168/jds.S0022-0302(92)77925-9).
- Duchemin, S. I., C. Colombani, A. Legarra, G. Baloche, H. Larroque, J.M. Astruc, F. Barillet, C. Robert-Granié, and E. Manfredi. 2012. Genomic selection in the French Lacaune dairy sheep breed. *J. Dairy Sci.* 95:2723-2733. <https://doi.org/10.3168/jds.2011-4980>.
- García-Cortés, L.A., J.C. Martínez-Ávila, and M.A. Toro. 2008. Partition of the genetic trend to validate multiple selection decisions. *Animal* 2:821-824. <https://doi.org/10.1017/S175173110800205X>.

- García-Cortés, L.A., J.C. Martínez-Ávila, and M.A. Toro. 2010. Fine decomposition of the inbreeding and the coancestry coefficients by using the tabular method. *Conserv. Genet.* 11:1945-1952. <https://doi.org/10.1007/s10592-010-0084-x>.
- Grundy, B., B. Villanueva, and J.A. Woolliams. 1998. Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genet. Res.* 72:159-168. <https://doi.org/10.1017/S0016672398003474>.
- Gulisija, D., D. Gianola, K.A. Weigel, and M.A. Toro. 2006. Between-founder heterogeneity in inbreeding depression for production in Jersey cows. *Livest. Sci.* 104:244-253. <https://doi.org/10.1016/j.livsci.2006.04.007>.
- Hayes, B. J., and S. P. Miller. 2000. Mate selection strategies to exploit across- and within-breed dominance variation. *J. Anim. Breed. Genet.* 117: 347-359. <https://doi.org/10.1046/j.1439-0388.2000.00252.x>.
- Henderson, C.R. 1973. Sire evaluation and genetic trends. *J. Anim. Sci.* 1973:10-41. <https://doi.org/10.1093/ansci/1973.Symposium.10>.
- Howard, J.T. 2017. Utilizing Genomic Information to Manage the Diversity and Minimize the Unfavorable Effects of Inbreeding in Livestock Populations. North Carolina State University.
- Hill, W. G. 1974. Prediction and evaluation of response to selection with overlapping generations. *Anim. Prod.* 18:117-139. <http://doi.org/10.1017/S0003356100017372>.
- Keightley, P. D., and W. G. Hill. 1987. Directional selection and variation in finite populations. *Genetics* 117:573–582. <https://doi.org/10.1093/genetics/117.3.573>.
- Kinghorn, B. P. 2011. An algorithm for efficient constrained mate selection. *Genet. Sel. Evol.* 43:4. <https://doi.org/10.1186/1297-9686-43-4>.
- Lagriffoul, G., E. Morin, J.M. Astruc, F. Bocquier, C. de Boissieu, P. Hassoun, J. Legarto, P.G. Marnet, J.L. Poulet, and F. Barillet. 2016. Panorama de la production de lait de brebis en France et son évolution depuis 50 ans. *INRAE Prod. Anim.* 29. <https://doi.org/10.20870/productions-animales.2016.29.1.2512>.

- Legarra, A., G. Baloche, F. Barillet, J. M. Astruc, C. Soulas, X. Aguerre, F. Arrese, L. Mintegi, M. Lasarte, F. Maeztu, I. Beltrán de Heredia, and E. Ugarte. 2014. Within- and across-breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech, and Basco-Béarnaise. *J. Dairy Sci.* 97:3200-3212. <https://doi.org/10.3168/jds.2013-7745>.
- Leroy, G. 2014. Inbreeding depression in livestock species: review and meta-analysis. *Anim. Genet.* 45:618-628. <https://doi.org/10.1111/age.12178>.
- Lynch, M. and W.G. Hill. 1986. Phenotypic evolution by neutral mutation. *Evolution*, 40(5), pp.915-935. <https://doi.org/10.1111/j.1558-5646.1986.tb00561.x>.
- Martinez-Castillero, M., L. Varona, S. Pegolo, A. Rossoni, and A. Cecchinato. 2021. Bayesian inference of the inbreeding load variance for fertility traits in Brown Swiss cattle. *J. Dairy Sci.* 104:10040-10048. <https://doi.org/10.3168/jds.2020-20087>.
- Meuwissen, T.H.E. 1997. Maximising the response of selection with a pre-defined rate of inbreeding. *J. Anim. Sci.* 75:934-940. <https://doi.org/10.2527/1997.754934x>.
- Mrode, R. A. 2014. *Linear Models for the Prediction of Animal Breeding Values*. 3rd ed. Cabi International. Wallingford, UK.
- Perdomo-González, D.I., A. Molina, M.J. Sánchez-Guerrero, E. Bartolomé, L. Varona, and M. Valera. 2021. Genetic inbreeding depression load for fertility traits in Pura Raza Española mares. *J. Anim. Sci.* 99:12. <https://doi.org/10.1093/jas/skab316>.
- Poyato-Bonilla, J., D.I. Perdomo-González, M.J. Sánchez-Guerrero, L. Varona, A. Molina, J. Casellas, and M. Valera. 2020. Genetic inbreeding depression load for morphological traits and defects in the Pura Raza Española horse. *Genet. Sel. Evol.* 52:62. <https://doi.org/10.1186/s12711-020-00582-2>.
- Races de France. 2011. Races laitières - Races de France. [<https://www.racesdefrance.fr/moutons/races-laitieres>] (visited on 02/07/2022).
- 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. <https://doi.org/10.1007/bf02986789>.

- Toro, M. A., and L. Varona. 2010. A note on mate allocation for dominance handling in genomic selection. *Genet. Sel. Evol.*, 42:33. <https://doi.org/10.1186/1297-9686-42-33>.
- VanRaden, P. M., and L. A. Smith. 1999. Selection and mating considering expected inbreeding of future progeny. *J. Dairy Sci.* 82:2771-2778. [https://doi.org/10.3168/jds.S0022-0302\(99\)75534-7](https://doi.org/10.3168/jds.S0022-0302(99)75534-7).
- Varona, L., J. Altarriba, C. Moreno, M. Martínez-Castillero, and J. Casellas. 2019. A multivariate analysis with direct additive and inbreeding depression load effects. *Genet. Sel. Evol.* 51:78. <https://doi.org/10.1186/s12711-019-0521-3>.
- Woolliams, J.A. 2007. Genetic contributions and inbreeding. *Util. Conserv. Farm Anim. Genet. Resour.* 147-165. https://doi.org/10.3920/9789086865925_008.
- Woolliams, J.A., and P. Bijma. 2000. Predicting rates of inbreeding in populations undergoing selection. *Genetics* 154:1851–1864. <https://doi.org/10.1093/genetics/154.4.1851>.
- 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. <https://doi.org/10.1093/genetics/153.2.1009>.
- Woolliams, J.A., and R. Thomson. 1994. A theory of genetic contributions. *Proc. 5th World Congr. Genet. Appl. Livest. Prod.*, Guelph, Canada.
- Wray, N.R., and R. Thompson. 1990. Prediction of rates of inbreeding in selected populations. *Genet. Res.* 55:41-54. <https://doi.org/10.1017/s0016672300025180>.
- Wray, N.R., J.A. Woolliams, and R. Thompson. 1990. Methods for predicting rates of inbreeding in selected populations. *Theoretical and Applied Genetics*, 80:503-512. <https://doi.org/10.1007/BF00226752>.

Training, education, teaching and awards during the PhD thesis

Year	Training course	Hours
2024	<ul style="list-style-type: none"> • Course on programming and computer algorithms in animal breeding with focus on single-step GBLUP and genomic selection in practice (Modules A & B). University of Georgia, Athens, USA, 13th – 31th May 2024. • Researcher Connect Online Programme. British Council, 11th March 2024 – 21th March 2024. 	75 24
2023	<ul style="list-style-type: none"> • Abroad Internship, The Roslin Institute, university of Edinburgh, 1st April – 26th August 2023. • TOEFL, score 88. 5th April 2023. • Breeding programme modelling with AlphaSimR. University of Edinburgh, MOOC – EdX, 19th January 2023. • The fundamentals of quantitative genetics: concepts and in-depth methodology. Sète, France, 4th– 8th December 2023. • Scientific integrity in the research field (on line). • Training on methodology for biology-driven selection. Centre for Quantitative Genetics and Genomics at Aarhus University, Aarhus, Denmark, 25th September 2023 – 28th September 2023. 	- - 25 35 6 20
2022	<ul style="list-style-type: none"> • SMARTER Short course, Population Genomics and data analyses. Universidad de León, León, Spain, 16th – 17th Mai 2022. • TOEIC, score 815. 21th April 2022. • Algorithm base. INRAE, Toulouse, France, 7th– 8th April 2022. • Papyrus, training course for doctoral students to help optimize their collection, monitoring, management and exploitation activities of bibliographic information and then facilitate the writing, publication and communication of their results. INRAE, 14th January 2022 – 11th March 2022. 	12 - 14 35
2021	<ul style="list-style-type: none"> • Linux & cluster. Genotoul bioinformatics platform, INRAE, Toulouse, France, 11th – 12th October 2021. 	14

	<ul style="list-style-type: none"> • First step in AWK programming. Genotoul bioinformatics platform, INRAE, Toulouse, France, 14th October 2021. • Genetic Models for Animal Breeding. Wageningen University and Research (MOOC – EdX), 8th November 2021. 	7
		30
	Total hours of training courses	297
Year	Teaching	Hours
2022-2024	<p>Animal breeding and genetics</p> <p>Practical courses: Population effective size and inbreeding (ENSAT – Licentiate 3rd year); Relatedness, genetic evaluation and genetic progress (ENSAT – Master 1st year).</p> <p>Practical courses: Allele and genotype frequencies under the Hardy-Weinberg equilibrium, or migration and genetic drift scenarios (ENSAT – Licentiate 3rd year).</p>	34
Year	Awards	Place
2023	AGBT Ag – poster (Partition of the genetic trend of French dairy sheep breeds in Mendelian sampling) (prize recipient)	Texas, USA
2022	180 seconds presentation on breeding durability (prize recipient)	Rennes, France

Titre : Méthodes basées sur la généalogie pour partitionner le gain génétique et le fardeau génétique chez les ovins laitiers

Mots clés : Originalité génétique, Fardeau génétique, Sélection, Ovins laitiers

Résumé : Cette thèse explore des méthodes basées sur la généalogie pour partitionner le gain génétique et le fardeau génétique (FG) dans les races ovines laitières françaises : Lacaune (LAC), Basco-Béarnaise (BB), Manech Tête Noire (MTN) et Manech Tête Rousse (MTR).

Le Chapitre 2 a utilisé une analyse rétrospective pour affiner la partition de la tendance génétique dans les échantillonnages mendéliens par catégorie d'animaux définies par le sexe et par la voie de sélection, ainsi que pour caractériser les contributions génétiques à long terme. Nous avons analysé le gain génétique pour la production laitière dans quatre races : LAC, BB, MTN et MTR. Les mères à béliers (MAB) et les mâles d'insémination artificielle (IA) ont été les sources les plus importantes de progrès génétique, comme l'a montré la décomposition des tendances de l'échantillonnage mendélien. Les contributions annuelles étaient plus variables pour les mâles d'IA que pour les MAB, étant donné que ces contributions ont été calculées en moyenne sur un plus petit nombre d'individus. En termes d'échantillonnage mendélien, les femelles ont contribué davantage que les mâles au gain génétique total, et nous interprétons cela comme étant dû au fait que les femelles constituent un plus grand réservoir de diversité génétique. En outre, nous avons calculé les contributions à long terme de chaque individu aux pseudo-génération suivantes. L'échantillonnage mendélien était plus important que la moyenne des parents pour déterminer la sélection des individus et leurs contributions à long terme. Ces contributions étaient plus significatives pour les mâles d'IA (dont la descendance est plus importante que celle des femelles) et en BB qu'en LAC (étant une race de taille plus importante).

Au Chapitre 3, la théorie qui montre la nature additive du FG est présentée. L'effet du FG et l'effet génétique additif (dans une population non consanguine) ont une corrélation négative dépendant de la fréquence des allèles, de la consanguinité et de la dominance. Nous avons calculé et décrit les coefficients de consanguinité partielle dans trois races : BB, MTN et MTR. Ensuite, nous avons inclus ces coefficients dans un modèle mixte en tant que covariables de régression aléatoire pour estimer la variance et les valeurs génétiques du FG pour la production laitière. Il existe une variance génétique pour le FG dans les races MTN et MTR, mais elle n'était pas différente de zéro pour BB. Comme attendu, nous avons estimé des corrélations génétiques négatives entre le FG et les valeurs génétiques estimées ; cependant, elles étaient proches de zéro dans les trois races. La faible magnitude du FG ne justifie pas une sélection fondée sur ce critère.

Dans le Chapitre 4, nous avons évalué l'efficacité de l'intégration du FG dans les stratégies de sélection chez les ovins laitiers. Nous avons simulé 10 générations de sélection. Six scénarios qui diffèrent par les critères de sélection (uniquement les valeurs génétiques additive estimées du caractère, uniquement les valeurs génétiques estimées du FG, ou à la fois les deux) et les stratégies d'accouplement (minimiser le FG ou la consanguinité attendue dans la descendance) ont été évalués. Les scénarios ont été comparés en termes de gain génétique, coefficients et taux de consanguinité, taille efficace et précision de la sélection. Il est possible d'utiliser les prédictions des effets du FG pour sélectionner les animaux directement ou dans le cadre de stratégies d'accouplement. Cependant, la sélection basée sur le FG (en raison de sa variation et de sa magnitude) ne présente pas d'intérêt pratique. À la lumière de nos résultats, l'inclusion d'animaux génotypés pourrait améliorer la précision de la prédiction des FG individuelles. D'autres recherches sont nécessaires.

Title: Pedigree-based methods to partition genetic gain and inbreeding load in dairy sheep

Key words: Genetic originality, Inbreeding load, Selection, Dairy sheep

Abstract: This thesis explores pedigree-based methods to partition genetic gain and inbreeding load in French dairy sheep breeds: Lacaune (LAC), Basco-Béarnaise (BB), Manech Tête Noire (MTN) and Manech Tête Rousse (MTR).

The Chapter 2 used a retrospective analysis to fine partitioning genetic trend in Mendelian samplings by categories of animals defined by sex and by selection pathways, and to similarly characterize long-term genetic contributions. We analysed genetic gain for milk yield in four dairy sheep breeds: LAC, BB, MTN and MTR. Dams of males and Artificial Insemination (AI) males were the most important sources of genetic progress as observed in the decomposition in Mendelian sampling trends. The yearly contributions were more erratic for AI males than for dams of males as they are averaged across a smaller number of individuals. Overall, in terms of Mendelian sampling, females contributed more than males to the total genetic gain, and we interpret that this is because females constitute a larger pool of genetic diversity. In addition, we computed long-term contributions from each individual to the following pseudo-generations. Mendelian sampling was more important than Parent Average to determine the selection of individuals and their long-term contributions. Long-term contributions were larger for AI males (with larger progeny sizes than females) and in BB than in LAC (with the latter being a larger population).

In Chapter 3, we presented theory that show the additive nature of the inbreeding load. The inbreeding load effect and the regular (in non-inbred population) additive genetic effect have a negative correlation depending on allele frequencies, inbreeding and dominance. We calculated and described the partial inbreeding coefficients in three French dairy sheep populations: BB, MTN and MTR. Then, we included these coefficients in a mixed model as random regression covariates, to predict genetic variance and breeding values of the inbreeding load for milk yield in the same breeds. There is genetic variance for inbreeding load in MTN and MTR breeds, but it was not different from zero for BB. As expected, we estimated negative genetic correlations between inbreeding load and breeding values; however, estimates were close to zero in the three sheep breeds. The small magnitude of inbreeding load does not warrant selection based on this criterion.

In Chapter 4, we evaluated the effectiveness of involving inbreeding load in selection strategies in a dairy sheep breeding scheme. We did this by simulation of 10 generations of evaluations and selection. Six scenarios that differ in the criteria of selection (only breeding values, only breeding values of inbreeding load, or both genetic and inbreeding load breeding values) and mate allocation strategies (minimising inbreeding load or minimising expected future inbreeding) were evaluated. Scenarios were compared in terms of genetic gain, inbreeding coefficients, rate of inbreeding, effective population size, and accuracy of selection. The use of predictions of inbreeding load effects to select animals directly or in mating strategies is feasible. However, selection based on inbreeding load (due to its variation and magnitude) is not of practical interest. In light of our results, the inclusion of genotype animals could improve the accuracy of predicting individual inbreeding loads. Further research is needed.