TESIS DE LA UNIVERSIDAD DE ZARAGOZA
2021
275

Itsasne Granado Tajada

Assessment of the usefulness of genomic information in the Latxa dairy sheep breeding program.

Director/es
Ugarte Sagastizabal, Eva

http://zaguan.unizar.es/collection/Tesis

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

© Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

ASSESSMENT OF THE USEFULNESS OF GENOMIC INFORMATION IN THE LATXA DAIRY SHEEP BREEDING PROGRAM.

Autor

Itsasne Granado Tajada

Director/es

Ugarte Sagastizabal, Eva

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Producción Animal

2021

Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es

Assessment of the usefulness of genomic information in the Latxa dairy sheep breeding program

Itsasne Granado Tajada

2021

1542
Universidad
Zaragoza

Neiker, Instituto Vasco de Investigación y Desarrollo Agrario

Departamento de Producción Animal

Universidad de Zaragoza

Facultad de Veterinaria

Departamento de Producción Animal y Ciencia de los Alimentos

ASSESSMENT OF THE USEFULNESS OF GENOMIC INFORMATION IN THE LATXA DAIRY SHEEP BREEDING PROGRAM

ITSASNE GRANADO TAJADA

Marzo, 2021

Assessment of the usefulness of genomic information in the Latxa dairy sheep breeding program

Evaluación de la utilidad de la información genómica en el programa de mejora de la raza ovina de leche Latxa

Doctoral thesis presented by

Itsasne Granado Tajada

Directed by

Dr. Eva Ugarte Sagastizabal

in order to obtain the doctoral degree within the

Animal Production Doctorate Program

of the University of Zaragoza, Faculty of Veterinary Medicine

Zaragoza, March 2021

Neiker, Instituto Vasco de Investigación y Desarrollo Agrario

Departamento de Producción Animal

Dr. **Eva Ugarte Sagastizabal**, investigadora principal del Departamento de Producción Animal de Neiker, Instituto Vasco de Investigación y Desarrollo Agrario.

Certifica:

Que la presente memoria de tesis titulada "**Assessment of the usefulness of genomic information in the Latxa dairy sheep breeding program**", que se corresponde con el proyecto de tesis doctoral aprobado el 23 de febrero de 2018, y de la que es autora **Itsasne Granado Tajada**, ha sido realizada bajo su dirección y reúne las condiciones requeridas para optar al grado de Doctor por la Universidad de Zaragoza.

Fdo. Eva Ugarte Sagastizabal

Vitoria-Gasteiz, 12 de marzo de 2021

Conforme al acuerdo de 25 de junio de 2020, del Consejo de Gobierno de la Universidad, por el que se aprueba el Reglamento sobre Tesis Doctorales de la Universidad de Zaragoza publicado en el Boletín Oficial de la Universidad de Zaragoza 07-2020 de 13 de julio de 2020, se presenta la Tesis "**Assessment of the usefulness of genomic information in the Latxa dairy sheep breeding program**" en la modalidad de **compendio de publicaciones** tratándose de trabajos previamente publicados, al cumplir los requisitos exigibles para ser defendida de este modo. Se han publicado, con fecha posterior a la del inicio de los estudios de doctorado, cuatro artículos con unidad temática, tres de ellos en una revista científica cuyo índice de impacto se encuentra incluido en la relación de revistas del Journal of Citation Reports (JCR).

Los trabajos incluidos se detallan a continuación:

Granado-Tajada, I., S. T. Rodríguez-Ramilo, A. Legarra, and E. Ugarte. 2019. Consanguinidad y parentesco en la raza ovina de leche Latxa. Albeitar. 229:16-18.

Granado-Tajada, I., S. T. Rodríguez-Ramilo, A. Legarra, and E. Ugarte. 2020. Inbreeding, effective population size and coancestry in the Latxa dairy sheep breed. Journal of Dairy Science. 102:1–11. https://doi.org/10.3168/jds.2019-17743

Granado-Tajada, I., A. Legarra, and E. Ugarte. 2020. Exploring the inclusion of genomic information and metafounders in the Latxa dairy sheep genetic evaluations. Journal of Dairy Science. 103:6346–6353. https://doi.org/10.3168/jds.2019-18033

Granado-Tajada, I., Varona, L. and Ugarte, E. 2021. Genotyping strategies for maximizing genomic information in evaluations of the Latxa dairy sheep breed. Journal of Dairy Science. *In Press*. https://doi.org/10.3168/jds.2020-19978

La revista científica **Journal of Dairy Science** tiene un factor de imparto de 3,333 según el JCR del 2019, se posiciona en el primer cuartil dentro del área Agriculture, Dairy & Animal Science (Q1, 5/63) y en el segundo cuartil del área Food Science & Technology (Q2, 37/139). Los trabajos se han realizado en coautoría con **investigadores doctores** e **Itsasne Granado Tajada** ha participado plenamente tanto en la conceptualización, tratamiento de datos, desarrollo de cálculos y análisis de resultados; como en la escritura y edición del manuscrito.

Neiker, Instituto Vasco de Investigación y Desarrollo Agrario

Departamento de Producción Animal

Dr. **Eva Ugarte Sagastizabal**, investigadora principal del Departamento de Producción Animal de Neiker, Instituto Vasco de Investigación y Desarrollo Agrario.

Certifica:

Que la presente memoria de tesis titulada "**Assessment of the usefulness of genomic information in the Latxa dairy sheep breeding program**", de la que es autora **Itsasne Granado Tajada**, ha sido realizada bajo su dirección, y que considerándola finalizada, autorizan su presentación en la modalidad de **Compendio de Publicaciones** para que sea juzgada por la comisión correspondiente.

Fdo. Eva Ugarte Sagastizabal

Vitoria-Gasteiz, 12 de marzo de 2021

“Todos somos seres especiales y únicos. Cada uno de nosotros es una colección personal de éxitos evolutivos que viene al planeta como una hoja prácticamente en blanco en la que cabe el universo entero. Las primeras palabras de esa página las escriben nuestros progenitores en el lenguaje genómico de las cuatro letras de la vida; el resto debemos completarlo nosotros mismos usando los otros lenguajes de la vida y de la felicidad mientras navegamos por las agitadas e incontrolables aguas del azar.”

Carlos López Otín, *La vida en cuatro letras*

El presente trabajo se ha realizado gracias a la financiación del proyecto **ARDI** (Ardi Research Development and Investigation; EFA208/16) con fondos de INTERREG POCTEFA (Programa INTERREG V-A España-Francia-Andorra), proyecto cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER).

Itsasne Granado Tajada ha recibido una beca del **Gobierno Vasco** a través de una ayuda de formación de jóvenes investigadores en formación y tecnólogos del Departamento de Desarrollo Económico e Infraestructuras.

Este trabajo no hubiera sido posible sin la colaboración de la Confederación de Asociaciones de Criadores de Ovino de razas Latxa y Carranzana (**CONFELAC**), por su aportación de datos de control lechero, registros genealógicos e información genómica de las poblaciones de Latxa.

Cuatro años han dado para mucho trabajo, pero también me han llevado a lugares que ni me imaginaba poder conocer y me han permitido descubrir a tantas personas que llevaré conmigo. Sin ellas, este trabajo no habría sido posible.

En primer lugar agradecer a Eva su gran ayuda, paciencia, sabio criterio y optimismo contagioso. Confiaste en que terminaría la tesis, antes de que yo me lo pudiera creer.

Neiker también me ha aportado compañeros que en un momento u otro me han ayudado a resolver mis dudas sobre el sistema de producción y el programa de mejora de la raza Latxa. También son parte de esta familia Lourdes y Larraitz, que me han acompañado en muchas dudas y cafés. Y en especial Carolina, que empezó escuchándome con paciencia y sin entender de qué me quejaba y ha acabado siendo compañera de cursos, congresos y tardes de cupcakes. Ahora ya comprende el problema de las ovejas hermafroditas y llora conmigo.

Una parte importante de la vida y andanzas en Neiker lo ha formado el grupo de becarias y amigas, que han enriquecido cafés y logrado que los jueves tengan otro color. Chicas, nos debemos un finde en León. Además, Vitoria siempre me recordará a las mejores compañeras de piso que podría haber, gracias.

Los meses pasados en Toulouse también me aportaron personas cuyo apoyo profesional y personal hicieron que la estancia fuera rodada. Eso sí, francés poco en el pasillo hispanohablante del INRA.

Otra pieza clave del viaje fueron los meses invertidos en Valencia, aquella inmersión en la genética cuantitativa fue la base sobre la que construir todo lo que he ido aprendiendo en este tiempo. Mientras que los amigos que me llevo de aquel tiempo me acompañaron en mi primera independencia, descubrir las fallas y a mí misma.

Hay personas que ya estaban conmigo antes de que este proyecto despegase y otras que se han ido sumando, mis chicas de la ikastola, las Biogirls, la kukuadri y demás amigos que han formado parte de mi vida y me han acompañado en las subidas y bajadas de forma inestimable. Gracias por dejar que me queje aunque no entendáis nada.

Este viaje ha permitido un crecimiento personal y mi compañía ha sido maravillosa e inspiradora.

Por supuesto, el comienzo de todo fue mi familia. Que me han apoyado en mis propósitos y han seguido preguntando por mis números.

Gracias a todos.

INDEX OF CONTENTS

INDEX OF TABLES

GENERAL INTRODUCTION

CHAPTER 1. INBREEDING, EFFECTIVE POPULATION SIZE AND COANCESTRY IN THE LATXA DAIRY SHEEP BREED

CHAPTER 2. INCLUSION OF GENOMIC INFORMATION IN LATXA DAIRY SHEEP GENETIC EVALUATIONS

CHAPTER 3. GENOTYPING STRATEGIES FOR MAXIMIZING GENOMIC INFORMATION IN EVALUATIONS OF THE LATXA DAIRY SHEEP BREED

INDEX OF FIGURES

GENERAL INTRODUCTION

CHAPTER 1. INBREEDING, EFFECTIVE POPULATION SIZE AND COANCESTRY IN THE LATXA DAIRY SHEEP BREED

CHAPTER 2. INCLUSION OF GENOMIC INFORMATION IN LATXA DAIRY SHEEP GENETIC EVALUATIONS

CHAPTER 3. GENOTYPING STRATEGIES FOR MAXIMIZING GENOMIC INFORMATION IN EVALUATIONS OF THE LATXA DAIRY SHEEP BREED

ABSTRACT

The development of molecular genetics and bioinformatics has transformed the animal breeding field, bringing about the possibility of including genomic information of individuals in genetic evaluation models. The implementation of genomic selection (**GS**) in dairy sheep breeding programs has been limited, mainly due to the presence of a wide range of breeds with small population sizes, lower use of artificial insemination (**AI**), heterogeneous performance recordings and a lower economic value per individual. Therefore, the implementation of GS in dairy sheep has followed different strategies depending on the breed and its characteristics.

The Latxa breed is autochthonous to the Basque Country and Navarre (Spain); three populations are distinguished (Latxa Cara Negra from Euskadi: **LCNEUS**; Latxa Cara Rubia: **LCR** and Latxa Cara Negra from Navarre: **LCNNAF**) and a separate breeding program exists for each one. The inclusion of genomic information in evaluations was previously studied and no conclusive results were found. Hence, the main objective of this doctoral thesis is to analyze the potential implementation of GS and the use of genomic information in the Latxa dairy sheep breed.

The first study analyzed the estimation of inbreeding, coancestry and effective population size ($\boldsymbol{N_e}$) coefficients, based on pedigree and genomic information and by different estimation methods. The inbreeding increase obtained was moderate, reflecting that the breeding program has done a suitable mating control even though decisions were taken considering pedigree-based inbreeding estimates. Different estimates depending on the information source and methodologies were expected. However, results were consistent and showed that black populations (LCNEUS and LCNNAF) have higher inbreeding increases and smaller N_e, whereas in LCR population the importation of genetic material from the French Manech Tête Rousse (MTR) has reduced the rates of inbreeding, resulting in larger N_e and an increase in coancestry between both blonde populations.

The second study addressed the effect of including genomic information in milk yield evaluations, by comparing prediction accuracies of pedigree (BLUP) and genomic (ssGBLUP) evaluations by cross-validation. In a first step the available genomic information of AI rams was used and the missing pedigree was modeled with metafounders. The results were slightly

biased and with small dispersion, but the inclusion of genomic information did not increase or decrease the accuracy of predictions for any of the breeds.

Afterwards, in a second step the analysis was updated due to the availability of a bigger number of genotyped animals, which in addition to AI rams also included natural service rams and females. The same methodology was applied (except for the use of genetic groups instead of metafounders) and the three Latxa populations were assessed. The bias and slope of the regression results were similar for both evaluation methodologies. However, unlike in the previous study, including the new genomic information improved the accuracies of genomic prediction. Specifically, LCNEUS and LCR populations showed a 14 % increase in accuracy, with a genotyped population near to 1,000 individuals, whereas LCNNAF had a 50 % lower genotyped population and a 5 % accuracy increase was found.

Finally, the third study was a simulation work where different genotyping strategies to create the reference population were tested and prediction accuracy of pedigree (BLUP) and genomic (ssGBLUP) evaluations for milk yield was compared. The simulation was carried out based on real phenotypic, genomic and pedigree structure of Latxa breeds, and the maximum number of genotyped individuals per year was limited given the economic restrictions of each breeding program. As expected, larger number of genotyped animals resulted in greater prediction accuracy; however, the population structure had an important influence. Thus, the genotyping of males not connected with the population did not affect accuracy, even if a large number of males were genotyped. The inclusion of females in the reference population was confirmed to be highly beneficial; between 10 and 50 % increases in accuracy were found. Different proportions of the genotyped number of males and females did not result in different accuracy estimates.

Therefore, the results found in this thesis would lead one to conclude that the inclusion of genomic information in milk yield genetic evaluations of Latxa breed is beneficial, and given the characteristics of the breed, including females in the genotyped population would be positive. Moreover, the availability of genomic information has enabled the obtention of accurate inbreeding and coancestry estimates of Latxa populations, showing it to be a highly useful tool for population management.

RESUMEN

Los avances acaecidos en el área de la genética molecular y de la informática han revolucionado el campo de la mejora genética animal, ya que abren la posibilidad de incorporar la información genómica de cada individuo en los modelos de evaluación genética. La implementación de la selección genómica (**SG**) en programas de ovino lechero ha sido limitada, principalmente debido a la presencia de una gran variedad de razas con tamaños poblacionales pequeños, un menor uso de la inseminación artificial (**IA**), controles de rendimientos heterogéneos y un menor valor económico. Por todo ello, la implementación de la SG en ovino lechero ha seguido diferentes estrategias en función de cada raza y sus características.

La raza Latxa es una población autóctona de la Comunidad Autónoma Vasca y Navarra, está dividida en tres variedades (Latxa Cara Negra de Euskadi: **LCNEUS**; Latxa Cara Rubia: **LCR** y Latxa Cara Negra de Navarra: **LCNNAF**) y cada una de ellas dispone de su propio programa de selección. Estudios previos sobre la inclusión de información genómica en las evaluaciones arrojaron resultados no concluyentes. Por lo tanto, el principal objetivo de esta tesis doctoral es analizar la potencial implantación de la SG y el uso de la información genómica en la raza de ovino lechero Latxa.

El primer estudio analizó la estima de coeficientes de consanguinidad, parentesco y censo efectivo ($\mathbf{N_e}$) en base a información de pedigrí e información genómica, empleando dos metodologías diferentes en el último caso. El incremento de la consanguinidad obtenido fue moderado en todos los casos, reflejando el correcto manejo realizado dentro del esquema para controlar la consanguinidad. Como es de esperar, los resultados fueron diferentes en función de la fuente de información y metodología empleadas. Sin embargo, los resultados fueron consistentes y mostraron que las poblaciones negras (LCNEUS y LCNNAF) tienen un mayor aumento de la consanguinidad y menores $\mathbf{N_e}$ que la población LCR. Esto se atribuye a que esta última variedad, desde el año 2010 está importando material genético de la raza francesa Manech Tête Rousse (**MTR**) y en consecuencia presenta un mayor $\mathbf{N_e}$, un menor aumento de consanguinidad y simultáneamente un aumento significativo del parentesco entre ambas poblaciones.

El segundo estudio abordó el efecto de incluir información genómica en las evaluaciones para producción de leche comparando mediante validación cruzada la precisión obtenida en evaluaciones por pedigrí (BLUP) frente a la obtenida en evaluaciones genómicas (ssGBLUP). En una primera fase se utilizó la información genómica disponible que correspondía a machos de IA y se modeló la genealogía faltante a través de metafundadores. Aunque los resultados mostraron sesgos y dispersiones favorables, la inclusión de la información genómica no aumentó ni disminuyó la precisión de predicción para ninguna de las razas.

Posteriormente, en una segunda fase se repitió el análisis al disponer de un mayor número de animales genotipados que además de machos de IA incluía machos de monta natural y hembras. Se aplicó la misma metodología (excepto por el uso de grupos genéticos en lugar de metafundadores) y se evaluaron las tres poblaciones de Latxa. Los resultados fueron similares para ambas metodologías de evaluación en términos de sesgo y pendiente. Sin embargo, a diferencia del estudio anterior, incluir la nueva información genómica mejoró la precisión de las predicciones genómicas. Concretamente, el aumento de precisión en las poblaciones de LCNEUS y LCR estuvo en torno al 14 %, con poblaciones genotipadas cercanas a 1000 individuos, mientras que en LCNNAF, con una población genotipada 50 % menor, el aumento de precisión se situó en torno al 5 %.

Finalmente, el tercer estudio fue un trabajo de simulación desarrollado con el fin de comparar diferentes estrategias de genotipado para conformar la población de referencia. Dicha comparación se realizó en términos de precisión de predicción de evaluaciones genéticas (BLUP) y genómicas (ssGBLUP) para producción de leche. La simulación se realizó en base a la estructura fenotípica, genómica y genealógica real de las poblaciones de raza Latxa, limitando el número máximo de individuos genotipados al año teniendo en cuenta las restricciones económicas del esquema. Como se esperaba, el aumento del tamaño de la población genotipada aumenta la precisión de predicción. Sin embargo, la estructura de dicha población influye de forma importante. Así, el genotipado de machos no conectados genealógicamente con la población no influyó en la precisión, por mayor que fuera el número. También se constató que la inclusión de hembras en la población de referencia es muy beneficiosa, obteniéndose aumentos de precisión de entre 10 y 50 %. Diferentes proporciones del número de machos y hembras genotipadas no reportaron diferencias de precisión.

Por lo tanto, teniendo en cuenta los estudios realizados en el contexto de esta tesis, se puede concluir que el uso de la información genómica en las evaluaciones genéticas para producción de leche en la raza Latxa es beneficioso y que dadas las características de la raza, incluir

hembras en la población genotipada de referencia sería una buena opción. Además, disponer de información genómica ha permitido obtener estimas de consanguinidad y parentesco precisas de las poblaciones de Latxa, mostrándose como una herramienta de gran utilidad para la gestión de las poblaciones.

LABURPENA

Azkenengo hamarkadetan biologia molekularreko tekniketan eta informatikan lortutako aurrerapenak animalien hobekuntza genetikoaren iraultza ekarri du, animali bakoitzaren informazio genomikoa ebaluazio genetikoetan kontuan hartzeko aukera ekarri duelako. Esne arditako programetan hautaketa genomikoaren (**HG**) ezarpena mugatua izan da, gehienbat populazio tamaina txikiko hainbat arraza daudelako, intseminazio artifizialaren (**IA**) erabilera mugatuarekin, hainbat ekoizpen kontrol motekin eta balio ekonomikoa baxua izanik. Guzti honegatik, HG ezarpenak esne arditan bide ezberdinak jarraitu ditu arraza bakoitzaren ezaugarrien arabera.

Latxa, Euskal Autonomi Erkidegoko eta Nafarroako bertako arraza da, hiru azpi-populazio bereizten dira (Latxa mutur gorria: **LCR**; eta Latxa mutur beltza, Euskal Autonomi Erkidegoko eta Nafarroako populazioak daudelarik: **LCNEUS** eta **LCNNAF**, hurrenez hurren) eta bakoitzak bere hobekuntza programa dauka. Informazio genomikoa ebaluazioetan gehitzearen inguruan aurretiaz egindako ikerketek ez zuten emaitza erabakigarririk ondorioztatu. Hortaz, doktoreko tesi honen helburu nagusia HGaren ezarpen aukerak eta informazio genomikoaren erabilera Latxa esne ardi arrazan aztertzea da.

Lehen ikerlanean odolkidetasun, ahaidetasun eta populazio eraginkorraren ($\mathbf{N_e}$) koefizienteen estimazioa aztertu zen datu genealogikoetan zein informazio genomikoan oinarrituta, bi metodologia ezberdin erabiliz azken honetan. Aurkitutako odolkidetasunaren haziera neurrizkoa izan zen hiru arrazatan, programak odolkidetasuna ekiditeko jarraitutako jardunbideak egokiak izan direla adieraziz. Espero izatekoa denez, emaitzak ezberdinak izan ziren erabilitako informazio iturria eta metodologiaren arabera. Alabaina, emaitzak sendoak izan ziren eta populazio beltzek (LCNEUS eta LCNNAF) LCR populazioak baino odolkidetasun handiagoa eta $\mathrm{N_e}$ txikiagoak dituztela erakutsi zuten. Izan ere, 2010. urteaz geroztik, azken populazio hori Manech Tête Rousse (MTR) arraza frantseseko material genetikoa inportatzen ari da eta ondorioz $\mathrm{N_e}$ handiagoa du, odolkidetasun-igoera txikiagoa eta aldi berean, bi populazioen arteko ahaidetasuna nabarmen handitu da.

Bigarren ikerlanean, esne ekoizpeneko ebaluazioetan informazio genomikoa gehitzearen efektuari heldu zitzaion, balioztatze gurutzatuaren bidez alderatuz pedigri bidezko ebaluazioen (BLUP) zehaztasuna, ebaluazio genomikoetan (ssGBLUP) lortutakoarekin. Lehen fase batean, erabilitako informazio genomikoa IAko ahariei zegokien eta falta zen genealogia

metafundatzaileen bidez modelatu zen. Emaitzek alborapen eta sakabanatze txikiak erakutsi zituzten arren, informazio genomikoa gehitzeak ez zuen iragarpen zehaztasuna handitu edo gutxitu arraza bakar batentzat ere.

Ondoren, bigarren fase batean lana errepikatu zen genotipatutako animalia gehiago zeudelako eta IAko ahariez gain, artaldetako ahariak eta ardiak ere bazeudelako. Metodologia bera aplikatu zen (metafundatzaileak erabili beharrean talde genetikoak erabiltzeagatik izan ezik) eta hiru Latxa populazioak aztertu ziren. Emaitzak antzekoak izan ziren bi ebaluazio-metodologietarako, alborapenari eta emaitzen sakabanatzeari dagokionez. Hala ere, aurreko azterlanean ez bezala, informazio genomiko berria gehitzeak iragarpen genomikoen zehaztasuna hobetu zuen. Zehazki, LCNEUS eta LCR populazioen zehaztasuna % 14 inguru altuagoa izan zen, 1000 animali inguruko populazio genotipatuekin, eta LCNNAFan genotipatutako populazioa % 50 txikiagoa izanik, zehaztasuna % 5 altuagoa izan zen.

Azkenik, hirugarren ikerlana simulazio-lan bat izan zen, erreferentziazko populazioa osatzeko genotipatze-estrategia desberdinak alderatzeko. Esne ekoizpeneko ebaluazio genetikoen (BLUP) eta genomikoen (ssGBLUP) iragartzeko zehaztasuna alderatu zen. Simulazioa Latxa arrazako populazioen egitura fenotipikoan, genomikoan eta genealogikoan oinarrituta egin zen, urtean genotipatutako gehieneko banakoen kopurua mugatuz eskemaren ahalmen ekonomikoa kontuan hartuta. Espero bezala, genotipatutako populazioaren tamaina handitzeak iragarpenaren zehaztasuna handitzen du. Hala ere, populazio horren egiturak eragin handia du. Horrela, genealogikoki populazioarekin konektatu gabeko aharien informazio genomikoak ez zuen zehaztasunean eraginik izan. Halaber, egiaztatu zen ardiak sartzea oso onuragarria dela, % 10 eta% 50 arteko doitasun-igoerak lortu zirelarik. Genotipatutako ahari eta ardi kopuruaren proportzio ezberdinek ez zuten zehaztasun desberdintasunik eman.

Beraz, tesi honen testuinguruan egindako ikerketak kontuan hartuta, Latxa arrazan esne ekoizpeneko ebaluazio genetikoetan informazio genomikoa erabiltzea onuragarria dela ondoriozta daiteke, eta arrazaren ezaugarriak kontuan hartuta, erreferentziazko populazio genotipatuan ardiak sartzea aukera ona izango litzateke. Gainera, informazio genomikoa edukitzeak aukera eman du Latxa populazioen odolkidetasun- eta ahaidetasun-estimazio zehatzak lortzeko, tresna oso erabilgarria izanik populazioak kudeatzeko.

GENERAL INTRODUCTION

GENERAL INTRODUCTION

1. Latxa dairy sheep breed

The Latxa breed is a dairy sheep breed which is autochthonous to the Western Pyrenees (Spain), characterized by a good daily aptitude, high rusticity and a good adaptation to a rainy and medium-cold environment. Two ecotypes are distinguished: Latxa Cara Rubia (**LCR**) and Latxa Cara Negra (LCN). These are mainly differentiated according to their blond or black head colour, respectively. Moreover, the black strain is divided into two non-contiguous regions with slight morphology differences: Latxa Cara Negra from Euskadi (**LCNEUS**) and Latxa Cara Negra from Navarre (**LCNNAF**). Currently, the LCR population is constituted by approximately 182,000 ewes distributed across 4,000 flocks. Regarding the black populations, there are 92,000 LCNEUS ewes in 3,800 flocks, while LCNNAF is the smallest population, with 62,000 ewes in 418 flocks. Figure 1 shows the geographical distribution of the three Latxa populations, although overlapping and mixed-flocks exist.

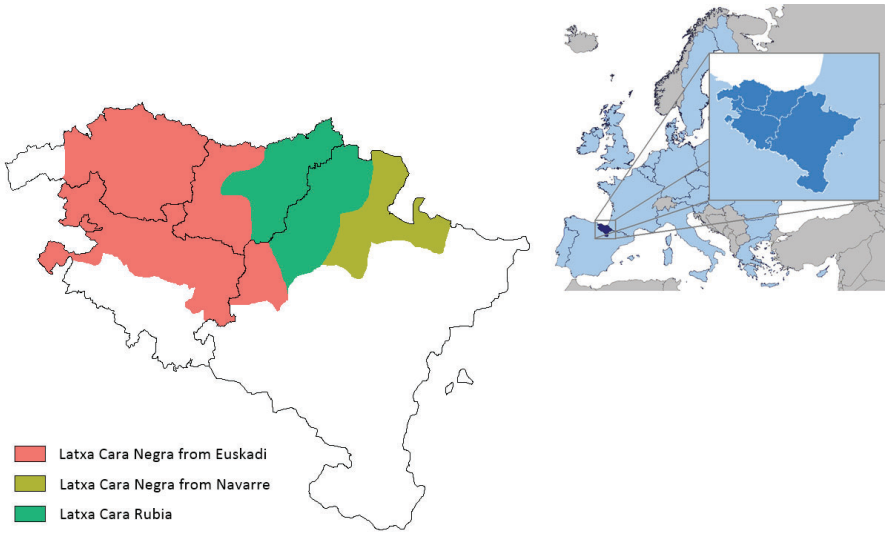

Figure 1: Geographic distribution of the three Latxa populations.

The Latxa production system is characterized as being one birth per year, maximizing the use of environmental resources, communal pastures and mountain passes. The milk yield is the most relevant aspect in the economic output of Latxa breed (Gabiña et al., 2000), because 95 % of the milk produced is transformed into cheese. Moreover, 60 % of this cheese is under the Protected Designation of Origin Idiazabal, which establishes some constraints, like the breed and the geographic area of production allowed. There is also a label mark that protects milking lambs but, in this case, only 5 % of the lambs are marked under this label.

The same breed is located on the other side of the border in France. This breed, named Manech, is also subdivided into two strains according to head colour: Manech Tête Rousse (**MTR**) and Manech Tête Noire (**MTN**). Due to geographical proximity and common use of pastures of the Pyrenees Mountains, animal exchanges within each colour are customary over both sides of the borderline. Moreover, importation of semen from the French to the Spanish schemes has been systematic from 2010 in LCR and 2018 in LCNNAF, to take advantage of breeding programs that started earlier and they are currently successful (Barillet and Roussely, 1987; Larroque et al., 2014). This connection and differences between breeds have been studied before (Ugarte et al., 1996; Legarra et al., 2014) and there is a consensus to consider that Latxa and Manech black and blond breeds are the same breed but separated geographically and with independent breeding schemes. A visual representation of relationships across populations based on principal components of genomic information was found to agree with this conception, as may be seen in Figure 2 from Legarra et al. (2014).

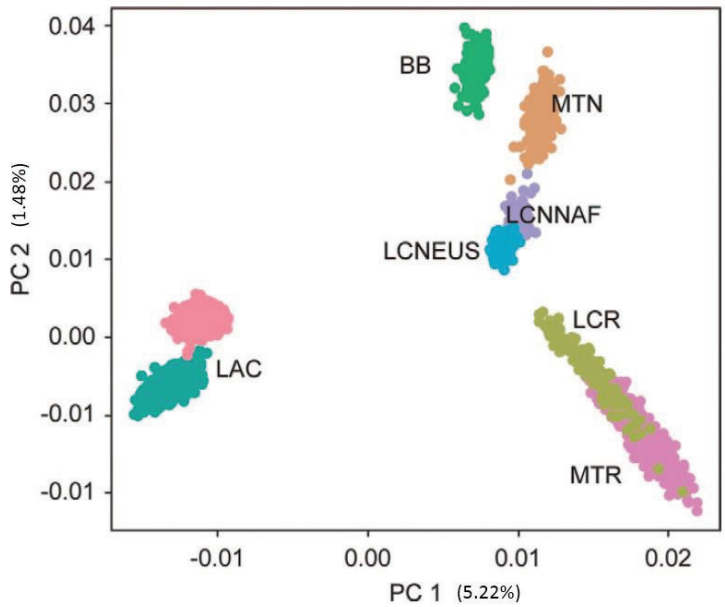

Figure 2: Principal component (PC) analysis showing the location of Western Pyrenees dairy sheep breeds: Basco-Béarnaise (BB), Lacaune (LAC), Manech Tête Noire (MTN), Manech Tête Rousse (MTR), Latxa Cara Rubia (LCR), Latxa Cara Negra Navarra (LCNNAF), and Latxa Cara Negra Euskadi (LCNEUS). Figure from Legarra et al. (2014).

1.1. The breeding program

The Latxa breeding program is based on pure breed selection and was started in the Basque Country (or Euskadi) in 1984 and in Navarre in 1986. A separate breeding program exists, one for each population: LCR, LCNEUS and LCNNAF. Historically, the breeding program has advanced by focusing on the introduction of new traits and implementation of new

methodologies. The first recorded trait was milk yield; the milk recording program was implemented in 1982 and simplified in 1985 (Gabiña et al., 1986). In 1998 genotyping for Scrapie resistance selection (*PRNP* gene) of artificial insemination (**AI**) males was started, to remove susceptible allele carriers. Data recording for milk composition started in 1999 and recording for udder morphology started in 2001. Finally, the latest new data collected has been the genomic information for genomic selection, which started in 2011.

To select for new traits in the breeding program, new methodologies have to be implemented. The evaluation by Best Linear Unbiased Prediction (BLUP) methodology was introduced in 1986 (Gabiña et al., 1991). Later in 1995 unknown parent groups were included in the genetic evaluation (Ugarte et al., 1996, 1997), and the contemporary groups used in the current milk yield evaluation model were defined in 2002 (Legarra et al., 2005). In 1997, exploratory studies were done to analyze the inclusion of milk composition (Legarra and Ugarte, 2001) and udder morphology (Legarra et al., 1999, 2001; Ugarte et al., 2001) traits, whose addition to the scheme started in 2005. Afterward, a combined index for milk yield, milk composition and udder morphology traits was developed in 2008. Recently, in 2014 exploratory analysis in terms of prediction accuracy of genomic selection were done (Legarra et al., 2014).

Regarding the historical close relationship between Latxa and French Manech breeds, several exchanges of individuals and semen have been done between the breeding programs. For blond populations, in 1990 10 French MTR rams were acquired to be used in the LCR scheme (Ugarte et al., 1996), and semen was imported occasionally in 1985, 1988, 1992 and 1997; and since 2010 the importation of semen from MTR to LCR has been systematic. For black populations, the exchange of genetic material between LCNNAF and MTN breeds was done from 1997 to 2000 and restored in 2018. In spite of genetic closeness, exchanges between LCNEUS and MTN have never been done, due to morphological differences.

1.2. Selection objectives and criteria

The purpose of the Latxa breeding program is to improve productive efficiency to increase the economic profitability of breeders, thereby keeping their traditional production system, rusticity and adaptation to local environmental conditions. Firstly, the main objective of the selection program is milk yield, to increase the milk production per ewe. For this trait the evaluated trait is 120 days standardized milk yield, estimated as the milk production from lambing to 120 days of lactation.

Cheese transformation capacity is directly related with fat and protein contents. However, the estimates of genetic correlation with milk yield have been reported as negatives (Legarra and Ugarte, 2001). So, milk composition traits are also selection objectives of the improvement program, whose selection criteria are kilograms and percentage of fat and protein in milk.

The presence of udders with undesirable morphology can affect the milking routine, as well as cause higher perinatal mortality of lambs due to suckling problems and increased losses as the productive life of these sheep is shortened. Therefore, udder morphology is also a selection objective, whose relevance increased after a greater adoption of milking machines. The selection criteria were based on the classification system defined by De la Fuente et al. (1996), where udder depth and attachment and teat length and verticality are scored on a scale from one to nine.

Given the relevance of the previously described selection objectives, a combined index was developed with a unified objective: to increase milk yield while maintaining fat and protein content at an acceptable level and improving or maintaining the udder morphology of the population (Legarra, 2002). The selection criteria combining all the traits give a relative weight of 48 % to milk yield, 12 % to fat percentage, 14 % to protein percentage, 9 % to udder attachment and 17 % to teat verticality.

Moreover, the breeding program also takes into account the resistance to transmissible spongiform encephalopathies (TSE), by the removal of susceptible allele carriers (Hurtado et al., 2002; Alfonso et al., 2006). Studies have also been done on somatic cell score, but it has not been included as a selection objective because the heritability and the genetic variances are low and in a first step management changes could be more feasible (Legarra et al., 2007).

1.3. Key components of the breeding program

1.3.1. Pedigree recording

Maternity assignment is performed at lambing by the shepherd, who has to record the date, identification of the ewe, number and sex of born lambs and their registration in the herd book.

Regarding the paternity assignment of lambs, when lambing is from artificial insemination the ram used to inseminate the ewe is considered as the father. This lambing has to be between 142 and 157 days after the insemination, and the ewe must have stayed separated from any other male at least a week before and after insemination. Natural service rams are recognised

as fathers when the mating has been controlled and with good conditions to ensure the reliability of the information (small mating sets with a ram and a few ewes).

When male lambs are selected to enter the artificial insemination centre, paternal filiation controls by molecular marker analysis are compulsorily done to ensure the filiation of the future progeny tested rams. Moreover, filiation controls could be done in other animals to assign maternities or paternities. However, filiations are not systematic, which leads to heterogeneous pedigrees with a considerable proportion of individuals with unknown pedigrees. As a result, based on the average of the full data set, 37 % of the ewes with milk records have unknown sire and dam, even though this percentage has decreased to 21 % since 2000.

Traditionally, inbreeding has been estimated and managed based on pedigree information, but taking into account the lack of complete knowledge this estimation may not be an accurate measure.

1.3.2. Performance control

The performance recording for milk production is based on milk yield recording and is carried out following the current regulations of the Official Milk Control and the rules of the International Committee for Animal Recording (ICAR, 2018). Lactations are calculated using the Fleischmann method and the real lactation (from lambing to the end of lactation), the milked lactation (real lactation minus the produced during the first 30 days of sucking) and the 120 days standardized lactation (from lambing to 120 days of lactation) are calculated. The trait used as selection criterion is the 120 days standardized milk yield, because it is less affected by the management than the other two measures and highly correlates with the milked lactation which is the economic objective (Gabiña et al., 1989, 2000).

For milk composition traits, fat and protein contents are measured by qualitative milk control, for which purpose individual milk samples are collected the control day. Regarding udder morphology, assessment is done based on the classification system defined by De la Fuente et al. (1996) as previously described. Both milk composition and udder morphology controls are done by authorized and qualified assessors to at least all the first lambing individuals only from flocks with high AI use, for economic reasons.

1.3.3. Genetic evaluation model

The genetic evaluation for milk yield is carried out considering the information of all the ewes with calculated lactations into historical data. The evaluation is done independently for each Latxa population, using the BLUP (Best Linear Unbiased Predictor) methodology of mixed models with an animal model with repeatability:

$$Y_{ijklm} = FYS_i + A_j + L_k + I_l + u_m + p_m + e_{ijklm}$$

Where flock-year-season (***FYS***), age-parity number (***A***), number of lambs born alive (***L***) and interval from lambing to first milk recording (***I***), are included as fixed effects (Legarra et al., 2005); and additive genetic (***u***) with unknown parent groups (Ugarte el at., 1996) and individual random environmental or permanent (***p***) are considered as random effects. Analyses are performed using the BLUPf90 software suite (Misztal et al., 2002).

The genetic parameters for milk yield are shown in Table 1, and the estimated heritability is between 0.18 and 0.25 depending on the population with a repeatability of around 0.40.

Table 1: Variance components, heritability (h^2) and repeatability (R) for milk yield in Latxa breeds for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR) and Latxa Cara Negra from Navarre (LCNNAF).

Breed	σ_a^2	σ_p^2	σ_e^2	h^2	R
LCNEUS	307	372	983	0.18	0.40
LCR	454	401	1206	0.22	0.41
LCNNAF	515	311	1215	0.25	0.40

As the work developed in this thesis manuscript is focused on milk yield evaluations, the genetic evaluation model and genetic parameters for milk composition and udder morphology traits are described in the Annexe I.

1.4. Phenotypic and genetic trends

Currently, the breeding program is well implemented and shows consolidated results with an annual genetic gain (**AGG**) for milk yield between 0.19 and 0.23 standard deviations, depending on the population. Regarding the other selected traits, the mean AGG for fat content is 0.04 and for protein content 0.10, while for udder attachment this is 0.01 and 0.07 for teat verticality.

1.5. Structure of the breeding program

The Latxa breeding program is based on pure breed selection, for female selection in flocks as well as for male selection to be assessed by progeny test. The selection scheme is described in Figure 3, and the process is as follows.

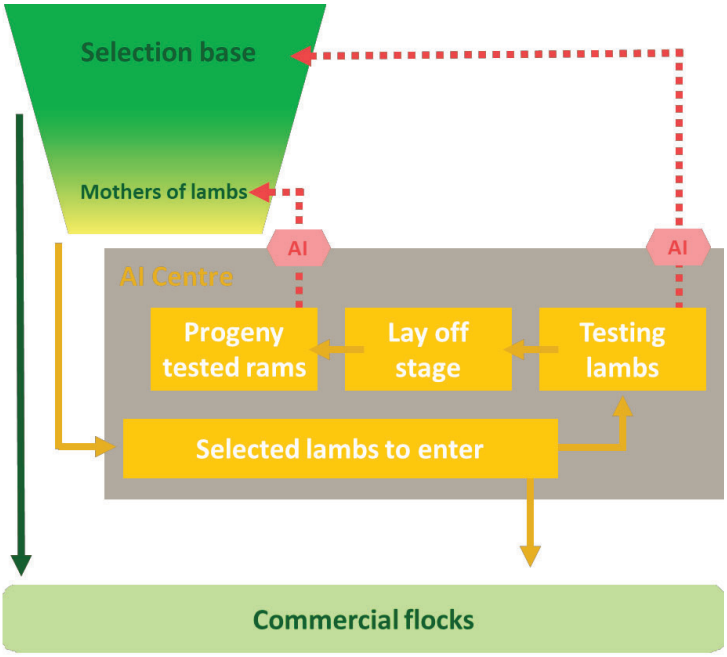

Figure 3: Scheme of the structure of the Latxa breeding program.

The selection of lambs to be progeny tested is made based on their pedigree index, from 10 % of the best ewes inseminated with a progeny tested ram. The best lambs are selected for the AI centre, and the remaining individuals are used for natural service. In this way also the genetic progress of natural service rams is also being worked with, although it is smaller due to lower selection intensity and accuracy of pedigree index evaluations.

At the AI centre there is another selection and those individuals with conformational defects, problems when collecting the semen or unsuitable physiological developments are rejected. Moreover, a selection committee comprised of breeders and technicians decides which individuals are going to be progeny tested based on pedigree index, breed standard and functional characteristics. So, at the end of the process approximately the 50 % of the selected lambs are tested.

Regarding the testing stage, each male is used to inseminate a minimum of 120 ewes in at least 10 flocks. Afterwards, there is a lay off stage where rams remain until their progeny test results are calculated, only when the ram has no less than five daughters with calculated lactation in a minimum of five flocks.

According to genetic results, rams are classified as progeny tested rams and remain in the artificial insemination centre. These rams are used based on their genetic value, to the extent possible and bearing in mind that fresh semen is used and that the diffusion capacity is limited. The best rams are the more used ones to diffuse the genetic progress, always avoiding inbreeding problems at mating.

2. Genomic Selection

2.1. Historical background: Classical selection

Historically animals have been selected based on their own phenotypic performance in order to improve one or several breeding objectives. During the last century, advances in population and quantitative genetics have supplied useful tools to estimate breeding values (**EBV**) of candidates to selection. These EBV are used to select the individuals with the highest genetic merit, which have allowed increased genetic progress in livestock populations. The first revolutionary step was the development of selection indexes (Hazel, 1943). This multiple regression method made it possible to combine all the information available on the animal and its relatives, and also the phenotypic information of genetically correlated traits. This selection index is the best linear predictor (BLP), but it lacks some properties such as the presence of bias due to not taking into account genetic differences among contemporary groups or only calculation of random effects because it assumes that observations are pre-adjusted to correct environmental effect.

The next relevant advance was the development of mixed model equations (**MME**) which allowed the estimation of breeding values by the best linear unbiased prediction (**BLUP**) methodology (Henderson, 1975). BLUP improves the selection index prediction by the simultaneous estimation of fixed effects and breeding values. However, it was necessary to develop a strategy to obtain the inverse of the numerator relationship matrix (**A**), done by Henderson (1976) and Quaas (1976), before this methodology was widespread. The inclusion of the relationship matrix in the MME of BLUP accounts for the changes in genetic variances resulting from selection, inbreeding, genetic drift and assortative mating (Sorensen and Kennedy, 1983; Kemp, 1985). Since the development of this methodology, models such as sire, animal, maternal or multiple-trait have been applied in genetic evaluations of many breeding programs. So during the last decades BLUP has been the main method for prediction of breeding values based on pedigree and phenotypic data for most of the livestock populations under genetic selection.

2.2. From the beginning of molecular markers to genomic selection

The advances in molecular biology techniques brought about the availability of molecular information. This new source of information became an important tool to increase the accuracy of evaluations and enhance the response to selection by including it in the genetic evaluation procedure, together with pedigree and phenotypic information. Molecular markers are interesting because they can capture genetic relationships between individuals and they are linked to quantitative trait loci (**QTL**). The gains in accuracy from incorporating molecular information depend on the amount of genetic variation explained by the marker and this information is especially useful for the evaluation of traits that present difficulties by classical genetic selection (Dekkers, 2004), such as those with low heritability or difficulty in obtaining phenotypes (e.g. sex-limited, expensive recording, measured late in life or at culling). The use of molecular markers in genetic evaluations has led to the development of new selection strategies.

When the genes controlling interesting traits are identified, these are combined with classical information in gene assisted selection (**GAS**). If many genes are identified and a large part of the genetic variance is controlled, GAS can be the most accurate methodology because direct selection could be done (Shumbusho, 2014). Several major genes affecting important traits have been discovered in livestock species (Rothschild et al., 1996; Elsen et al., 1999; Davis, 2005; Ogorevc et al., 2009) and in most cases they are normally considered (e.g. Boroola, Casein, Myostatin or PRNP genes), although the use of GAS is not very common in practice (Larzul et al., 1997; Dekkers, 2004). In case of not identified genes, markers in linkage with genes can be used in a marker assisted selection (**MAS**) that considers genetic markers as proxies of the QTL. For this selection strategy, the biggest and statistically significant QTL of the trait has to be found and combined with the polygene (the genetic part not influenced by detected QTL) (Shumbusho, 2014). Although many large QTLs have been located (Hu et al., 2013), they explain less than 10 % of genetic variance (Mouresan, 2016) because generally most quantitative traits appeared to be determined by many additive genes of infinitesimal effect at unlinked loci (Fisher, 1918; Bulmer, 1980). Therefore, the difficulty of detecting genes and the small proportion of genetic variance explained by those detected ones led to few breeding programs successfully implementing MAS (Boichard et al., 2002; Bennewitz et al., 2004; Dekkers, 2004).

The first draft of the Human Genome Project (Sachidanandam et al., 2001) revealed that the most common type of genome sequence variation can be attributed to single nucleotide

polymorphisms (**SNP**). Based on the idea of incorporating marker information into BLUP evaluations (Fernando and Grossman, 1989) and extending the concept to the whole genome (Lande and Thompson, 1990; Haley and Vischer, 1998), Meuwissen et al. (2001) developed the genome-wide selection or genomic selection (**GS**). This method simultaneously combines pedigree and phenotypic data with genomic information from the whole genome in the estimation of livestock's genetic potential. A key aspect of SNP markers is that they are abundant and located throughout the whole genome (Schork et al., 2000), so it is expected that some of the markers are in linkage disequilibrium (**LD**) with the QTLs of traits of interest. If the marker density is high enough to cover the whole genome, the additive effect of the QTLs would be captured without having to locate them (Mouresan, 2016).

At first, SNP effects were estimated by Bayesian methods (Meuwissen et al., 2001) or by the equivalent method named genomic BLUP (**GBLUP**) (Habier et al., 2007), based on a reference population with genomic information and phenotypes (or pseudo-phenotypes from non-genotyped relatives). The GBLUP method is similar to the classic BLUP (Henderson, 1975), but instead of the standard pedigree relationship matrix (**A**) it uses a genomic relationship matrix (**G**) built from molecular data (VanRaden, 2008). Genomic EBVs (GEBV) of genotyped candidates to selection were predicted based on genomic relationships. Afterwards, traditional BLUP evaluation was still needed to account for pedigree relationships. As several steps were needed to estimate GEBV, this methodology was called multistep. Although multistep methods were largely implemented for genomic evaluations worldwide, only a portion of the pedigree is genotyped and genomic information cannot be extended to non-genotyped animals (Lourenco et al., 2020); so genotyped animals have GEBV and non-genotyped animals EBV, which are not directly comparable. Moreover, when genotyped individuals had no phenotypic records the calculation of pseudo-phenotypes was needed (e.g. daughter yield deviation), so there was a loss of information due to weighting caused by different amount of information in the original data set and potential bias caused by selection (Mouresan, 2016).

In order to cope with these disadvantages and to perform a joint evaluation using all phenotypic, pedigree and genomic information, Misztal et al. (2009) proposed the single-step genomic BLUP (**ssGBLUP**) methodology. This method involves the replacement of pedigree relationship matrix (**A**) used in traditional BLUP by a realized relationship matrix (**H**), to account for genomic relationships. Additionally, Legarra et al. (2009) suggested considering pedigree relationships as a priori relationships and genomic relationships as the observed relationship. The derivation of the joint distribution of both relationships would allow the extension of

genomic information to non-genotyped animals. This means that in ssGBLUP pedigree relationships for non-genotyped animals are enhanced by the genomic information of their relatives (Lourenco et al., 2020). The combined pedigree and genomic relationship matrix (**H**) showed convergence problems for large data set (up to 1 million), but Aguilar et al. (2010) and Christensen et al. (2010) found an inverse of H that allows drastically simpler computations. This methodological development along with the automation of SNP genotyping at relatively low cost has enabled the implementation of GS in livestock populations.

2.3. Why is genomic selection such a huge revolution?

The expected genetic gain (Δ**G**) from a breeding program can be predicted using the Rendel and Robertson formula (Rendel and Robertson, 1950), as:

$$\Delta G = \frac{i r_{A,\hat{A}} \sigma_g}{L}$$

Where selection intensity (***i***), prediction accuracy ($\boldsymbol{r_{A,\hat{A}}}$) and additive genetic standard deviation ($\boldsymbol{\sigma_g}$), are divided by generation interval (***L***). It is broadly accepted that the implementation of GS allows bigger Δ**G** than traditional progeny testing selection, mainly due to higher prediction accuracy and reduced generation interval (Schaeffer, 2006; Obšteter et al., 2019).

The accuracy of genomic prediction is influenced by several factors. Among them there are the specific features of the trait of interest (heritability, genetic architecture, number and distribution of genes, and linkage disequilibrium between prediction markers and QTL), the characteristics of the population and the design of the genotyped reference population (size, sex ratio or the relationship of genotyped individuals within them, with candidates to selection and with the overall population) (Lund et al., 2016; Schöpke and Swalve, 2016; van den Berg et al., 2019). The latter factor is the only feature under the control of the breeding program and where GS plays an important role.

Genotyped populations composed of only males has been the standard approach since the beginning of GS (VanRaden, 2020) due to limited economic resources, because the selection pressure exerted on this sex is greater and the diffusion of the genetics of rams achieved by AI is wider. To improve this common issue, a large amount of research focuses on the design of genotyping strategies to optimise the selection of individuals to be genotyped. In this sense, there are studies about the impact of selecting individuals genetically related with candidates to selection (Hayes et al., 2009b; Habier et al., 2010; Clark et al., 2012), including genotyped

females in the reference population (Jimenez-Montero et al., 2012; Thomasen et al., 2014; Perez et al., 2019), or increasing the genotyped population size with data from other populations of the same breed or different but related breed (Lund et al., 2016; Schöpke and Swalve, 2016).

The generation interval is directly affected by improved prediction accuracy of non phenotyped and early genomically tested sires. Despite the lower accuracy of genomically selected sires compared to conventional progeny tested selection, the availability of GEBV at birth allows the use of young sires and faster turnover of increased genetic gain (Obšteter et al., 2019).

Moreover, compared with classical selection methods, GS has the potential to control the rate of inbreeding per generation better through a more accurate estimation of Mendelian sampling term, thereby differentiating among siblings and decreasing their co-selection (Daetwyler et al., 2007). In addition to considering Mendelian sampling, the use of genomic information to estimate inbreeding coefficients avoids underestimation due to incomplete or erroneous pedigree information and takes into account the LD caused by selection (Oliehoek and Bijma, 2009; Forutan et al., 2018). Consequently, lower inbreeding rates (Forutan et al., 2018) and higher genetic variability (Obšteter et al., 2019) can be achieved.

2.4. Genomic Selection in livestock populations

With the development of molecular techniques, bioinformatics and evaluation methodologies, the implementation of GS in breeding programs has been widespread, although at variable rates depending on the species and breed (Ibañez-Escriche and Gonzalez-Recio, 2011; Jonas and de Koning, 2015; Meuwissen et al., 2016).

Dairy cattle breeding programs were the first and most successful in implementing GS (VanRanden, 2020). The first official genomic evaluation for Holstein was released in January 2009 by the US Department of Agriculture (USDA), and since then evaluations for other breeds (Jersey, Brown Swiss or Aryshire) and in many countries have been done (Hayes et al., 2009a; Wiggans et al., 2017). A key factor was the collaboration between countries to establish consortiums like EuroGenomics, to create a common reference population which allowed increased prediction reliability (Lund et al., 2011). Genomic predictions showed over 0.8 accuracies for production traits and 0.7 for fertility, longevity, somatic cell count, and other traits (Lund et al., 2011; Wiggans et al., 2011). These high accuracies reflected the large reference populations that were made-up with many progeny-tested bulls with very accurate

phenotypes, which were often used to predict GEBV of close relatives (Meuwissen et al., 2016), and provided an alternative to traditional progeny testing that allowed an important reduction of the generation interval (Hayes et al., 2009a). Currently, more than five million dairy cattle have been genotyped for the purpose of GS (Council on Dairy Cattle Breeding, 2020), and more emphasis is expected on selection for environmentally friendly production, including reduced waste production and gas emission (Schöpke and Swalve, 2016). In addition, as genetic variance for economic traits is maintained by increase in frequency of rare alleles, new mutations, and changes in selection goals and management, it is unlikely to reach a selection plateau in the near future (Weller et al., 2017).

Regarding beef cattle, GS has not been adopted as widely as in dairy cattle. The main constraints to implementation are the segmented nature of the industry with multiple breeds and crossbreds, the low use of AI and the relatively low levels of phenotyping (Berry et al., 2016). As a consequence, although more than 500,000 Angus animals are genotyped in the USA (Mefford, 2018), the reported accuracies are between 0.3 and 0.7 (Van Eenennaam et al., 2014). These lower accuracies are influenced by small reference population size, with few high-accuracy sires, and the relatedness between reference population and candidates to selection. Moreover, due to the lack of connection between breeds, the use of multi-breed reference population showed slight increases in accuracy (0.33 to 0.38, by Bolormaa et al., 2013).

In pig breeding programs, the selection structure is quite different from cattle. The improvement is done on the nucleus, where elite boars are selected, and has an important impact on commercial populations (Ibañez-Escriche et al., 2014; Van Eenennaam et al., 2014). The pyramidal structure, together with the systematic exploitation of crossing and heterosis are factors that can make GS economically feasible (Simianer, 2009). As boar test recoding is generally done before the selection, opportunities to reduce generation intervals are limited (Meuwissen et al., 2016). The benefits of GS are therefore directed at traits such as meat quality traits measured only on the carcass (Samore et al., 2015), maternal traits observed only on females and obtained after reproductive age (Lillehammer et al., 2011) or crossbred performance traits not recorded on purebred animals (Esfandyari et al., 2015).

The poultry industry has comparable breeding schemes to that of pigs, and comparable traits where GS is promising for genetic improvement (Ibañez-Escriche and Gonzalez-Recio, 2011). The main objective of implementing GS is for traits that can only be directly recorded for one sex such as egg quality and production traits in layers, to improve crossbred performance in

commercial environment in broilers or to select for traits not recordable in the nucleus as disease resistance (Meuwissen et al., 2016). Greater responses to selection have been shown from GS (Heidaritabar et al., 2014; Wolc et al., 2015), and the availability of improved accuracy enabled the redesign of breeding programs to optimize the selection intensity applied to young animals (Wolc et al., 2015), and the avoidance of hierarchical mating structure (Hsu et al., 2015). In addition, due to the recent progress in sequencing, it is likely that turkeys will be the next poultry species implementing GS (Dalloul et al., 2010).

With regard to the implementation of GS in small ruminant breeding programs there are fewer references and gains in accuracy of genomic schemes are generally low (Mrode et al., 2018). Mainly it is due to the small population sizes, low LD, multi-breed evaluations, limited phenotype recording and small genotyping cost-benefit comparing with cattle (Rupp et al., 2016). In sheep meat quality traits, an average gain in genomic prediction accuracy between 0.05 and 0.27 has been described (Daetwyler et al., 2012; Auvray et al., 2014). As will be shown below, the accuracy of genomic evaluations in dairy sheep was found to be between 0.10 and 0.20 higher (Baloche et al., 2014b; Legarra et al., 2014), and increased accuracy was also described in goat genomic evaluations (Carillier et al., 2013; Molina et al., 2018). Based on modelling versus traditional selection methods, Shumbusho et al. (2013) showed that annual genetic gains were up to 17.9 % greater for combined meat and maternal traits in meat sheep, 51.7 % in dairy sheep, and 26.2 % in dairy goats.

Finally, in other species such as rabbit or aquaculture, the implementation of GS is still under development. The availability of commercial SNP chips was an important milestone for species of smaller economic value (Bertolini et al., 2014; Palti et al., 2015). Moreover, in aquaculture only a few species have breeding programs where GS could be relevant for selection of traits like resistance to infections, growth or meat condition (Wang et al., 2017; Silva et al., 2019; Kjetså et al., 2020).

Therefore, the genomic revolution started in dairy cattle and through the last two decades has spread through livestock breeding programs as a very useful tool which can lead to interesting benefits. However, the schemes implemented should not be imitated, although common traits of interest exist. The specific requirements of each breeding program should be considered when adapting the GS scheme before implementation (Jonas and de Koning, 2015; Mrode et al., 2018).

2.5. Genomic Selection in dairy sheep breeds

The International Sheep Genomic Consortium (ISGC, 2002) is a partnership among 20 countries that has facilitated the development of genomic tools. Among them, the OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) was released in 2009, bringing new perspectives for GS implementation in sheep breeding programs. In comparison with dairy cattle, dairy sheep breeding has several differences that should be considered. The presence of a wide range of breeds, with environmentally adapted production systems, small population sizes, very heterogeneous data recording systems and a lower economic value per individual are some intrinsic characteristics that influence the implementation of GS in sheep breeding programs.

The French Lacaune was the first dairy sheep breed to implement GS. Previously several studies were done to analyze the effect of GS in prediction accuracy based on real data. Different genomic evaluation methods were assessed (multistep GBLUP, Bayes Cπ, Partial Least Squares and Sparse Partial Least Squares) for milk yield, fat content and somatic cell scores, and no important differences were found among methods, GEBV being always more accurate than pedigree EBV (Duchemin et al., 2012). Applying ssGBLUP method with only one evaluation step, the GEBV for milk yield were 0.15 more accurate than pedigree based estimates (Baloche et al., 2014b), and estimates for fat and protein content, somatic cell scores and udder traits were between 0.10 and 0.20 more accurate. An experiment considering progeny test results of rams selected using either EBV or GEBV confirmed the gain in AGG (Baloche et al., 2014a).

In Western Pyrenees breeds, the implementation of GS was studied for the first time in 2014. Legarra et al. (2014) carried out a study with 6 breeds of dairy sheep: Basco-Béarnaise, the two strains of French Manech (Manech Tête Rousse, **MTR**; and Manech Tête Noire, **MTN**), and the three populations of Spanish Latxa. Daughter yield performances of rams were used for pedigree and genomic (by pseudo-ssGBLUP method) evaluations. A gain in accuracy was shown when genomic information was included in predictions of French breeds, GEBV for milk yield being more accurate than pedigree based evaluations (0.16 for MTR, 0.11 for MTN and 0.06 for Basco-Béarnaise). As Manech and Latxa are highly related breeds, multi-breed predictions were also tested by Legarra et al. (2014), pooling blond breeds (MTR and LCR) and black breeds (MTN and LCNNAF). However, no significant gains in accuracy were found compared with single breed predictions, in spite of their strong connection.

Among French dairy sheep breeds, Lacaune and MTR showed the highest gains in accuracy from using genomic information. The genotyped population sizes were of around 1,900 and 1,000 AI rams for Lacaune and MTR respectively, which are over the minimum of 1,000 individuals set by Shumusho et al. (2013) to achieve benefits from GS over classical selection in dairy sheep. With those prediction accuracies and an already short generation interval (4 years), GS in dairy sheep depends critically on the ability to accurately select young rams. The specific requirements of the breeding program should be considered when adapting the implementation of GS (Jonas and de Koning, 2015; Mrode et al., 2018). In these dairy sheep breeds the use of fresh semen and the seasonality of AI are key factors, so a high number of live rams are required to face them.

Considering these constraints, different genomic breeding schemes were modelled in Lacaune and MTR (Buisson et al., 2014). With the same cost as the classical breeding scheme, in Lacaune a GS scheme with genomic selection rate at lambs of 0.3 and progeny selection rate of 0.8 was found to yield an extra 15 % gain in AGG and a 40 % reduction of the number of AI rams. However, in MTR AI rams only were reduced by 25 %, so the same selection rates increased by 40 % the costs of GS scheme, which could be worthwhile with a 20 % increase in AGG. Significantly reducing the number of rams in the AI centre is essential to achieve a profitable GS scheme. In progeny-testing dairy sheep schemes using fresh semen, a large number of rams are waiting for first crop after their first use on AI, whereas a genomic scheme could avoid this waiting period because lambs are selected based on more precise estimates. Therefore, by modifying the management to a GS scheme there could be fewer rams in the AI centre, reducing costs and with no loss of genetic diversity. Finally, the breeding organizations decided to move toward a GS scheme in 2015 in French Lacaune (Buisson et al., 2014) and in 2017 in French Western Pyrenees breeds (Legarra et al., 2014).

Italian Sarda is other Mediterranean dairy sheep breed that faced the limitations in implementing GS with a different approach: a female reference population. A female nucleus was developed, which represented the genetic variability of the whole population; ewes were routinely phenotyped and ewes and their sires were genotyped (Salaris et al., 2018). The usefulness of this approach for milk yield evaluations was assessed by Usai et al. (2018), and genomic predictions of rams were found to be 0.13 more accurate than pedigree based predictions. It was concluded that genomic predictions of Sarda rams based on a female reference population can be used for selective breeding.

Regarding Spanish dairy sheep breeds, in 2018 studies to assess the state of GS implementation were started in four breeding programs: Assaf, Churra, Manchega and Latxa (National Institute for Agricultural and Food Research and Technology or INIA, 2020a), although all these breeding programs had started to genotype previously and some genomic evaluation studies had already been done. These works focused on analysing the selection criteria to make up the genotyped population and the development of genomic evaluations. Currently, routine genomic evaluations are done in Assaf and Manchega breeds, while Churra and Latxa are in an initial stage. Moreover, the benefit on genomic evaluations of constituting a metapopulation with these four dairy sheep breeds was assessed (INIA, 2020b). However, increasing the genotyped population size with individuals of other breeds did not show higher accuracies in genomic predictions. The genetic differentiation between studied breeds was likely to have been a decisive factor in the results obtained, as may be seen in Figure 4.

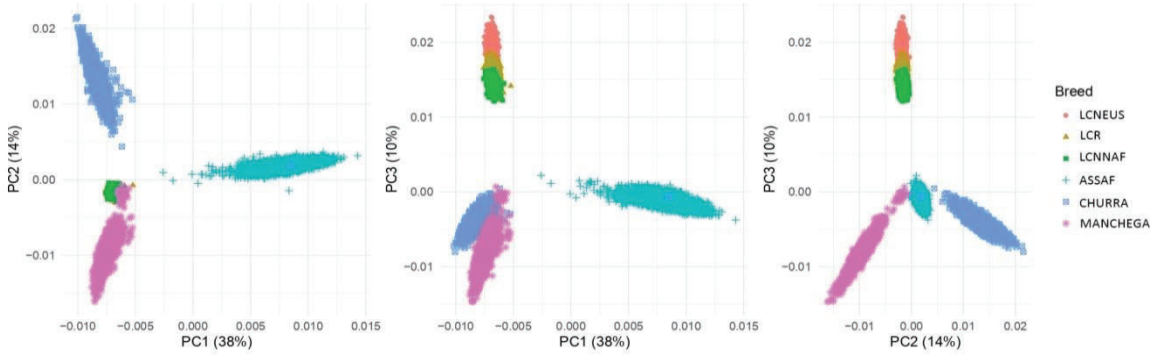

Figure 4: Principal component (PC) analysis of the included Spanish dairy sheep breeds: Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), Latxa Cara Negra from Navarre (LCNNAF), Assaf, Churra and Manchega. Figure from INIA (2020b).

2.6. Genomic selection in Latxa Breeds

Legarra et al. (2014) did the first study to assess genomic predictions in the three populations of the Latxa breed. However, no clear advantages of GS in terms of prediction accuracy were found and some results were inconsistent. The bias found was high for genetic and genomic predictions in the three breeds (from 18 to 146), and in all cases higher than for French breeds. The dispersion of EBV was far from one for LCNEUS and LCR (between 0.19 and 0.37), while LCNNAF showed slopes of 0.73 and 0.88 for pedigree and genomic evaluations respectively. The inclusion of genomic information was shown to worsen prediction accuracy for LCNEUS (-0.18) and improve it for LCR (0.30), while for LCNNAF accuracy was near zero (0.04). These results did not agree with the genetic and phenotypic progress observed in the breeding program. Only LCNNAF showed coherent results, even though the gain in accuracy of GEBV was not statistically significant.

Authors attributed these results to the genomic data available, which was limited and had an unbalanced distribution across the population, being few genotyped individuals in some years. Also, a weak relationship was described between genotyped individuals with phenotypic data and candidates to selection. Poor modeling may have affected the results as well, because missing pedigree was modelled using unknown parent groups and pseudo-single-step GBLUP methodology was used with daughter yield deviations, which could have transmitted inaccuracies to genomic predictions.

So, the results by Legarra et al. (2014) for Latxa breeds were not conclusive and the benefit of including genomic information into genetic evaluations was not clear. Since then, the Latxa breeding program has collected more phenotypic and pedigree data and has genotyped systematically all the new males used in recent years in AI.

3. Framework of the thesis manuscript

The work presented in this thesis has been developed in this context. The studies included focuses on analysing the feasibility of incorporating genomic information in the breeding program of small-sized populations such as the Latxa sheep breed, with three main aspects:

- Compare inbreeding, coancestry and effective population size estimates based on genomic and pedigree information. This point is developed in Chapter 1 and was published in the Journal of Dairy Science 2020, 103:5215–5226.
- Analyse the inclusion of genomic information in genetic evaluations and compare with classic genetic evaluations. This point is developed in Chapter 2 and was published in the Journal of Dairy Science 2020, 103:6346–6353.
- Assess genotyping strategies feasible in the Latxa breeding program. This point is developed in Chapter 3 and was published in the Journal of Dairy Science 2021, in Press.

Moreover, in relation to the studies done in the first point, a technical article was published in the informative journal Albeitar and is included in Annexe II.

3.1. Information included

The data used in the studies making up this thesis manuscript came from the Latxa Breeders' Associations' Confederation (CONFELAC) and comprise pedigree, phenotypic and genomic information available from the beginning of the breeding program until 2017 (except for

section 2.2). The full pedigree and phenotypic data sets are described in Table 2, and a detailed description of the current pedigree and phenotypic data is shown in Table 3.

Genealogical data comprised 263,308 individuals for LCNEUS, 150,185 for LCR and 68,714 for LCNNAF. As previously described, filiations are not systematic, leading to heterogeneous pedigrees with a considerable proportion of individuals with unknown pedigree. As a result, based on the average of the full data set, 37 % of the ewes with milk records have unknown sire and dam as described in Table 2. Even though since 2000 this percentage has decreased to 21 %, while the percentage of individuals with known sire and dam is 45 % for females and 64 % for males as shown in Table 3.

Regarding phenotypic data, we used 120 days standardized milk yield information with 639,517, 392,109 and 183,251 records of 235,360, 133,230 and 61,309 ewes distributed across 456, 325 and 117 flocks for LCNEUS, LCR and LCNNAF, respectively. Currently, the mean number of daughters with lactations per male is 24 among the males with daughters, all the populations has a mean number of lactations per ewe of three (Table 3) and the phenotypic mean and standard deviation is 134 ± 57 for LCNEUS; 148 ± 68 for LCR; 143 ± 57 for LCNNAF.

Table 2: Description of the total number of males, artificial insemination (AI) males and females, number of individuals with known complete pedigree (sire and dam) or with known dam, and number of females with recorded lactations for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF). Based on full data set.

Breed	No. males				No. females			
	Total	AI	Know pedigree (%)	Known dam (%)	Total	Know pedigree (%)	Known dam (%)	With records
LCNEUS	16,916	902	6,042 (36)	6,752 (40)	246,392	53,405 (22)	114,506 (46)	235,360
LCR	10,725	714	5,963 (56)	2,093 (20)	139,460	48,180 (35)	44,255 (32)	133,230
LCNNAF	4,444	604	2,594 (58)	1,052 (24)	64,270	20,872 (32)	21,448 (33)	61,309

The genomic data consisted of AI rams genotyped from 1997 to 2017, obtained from stored DNA samples for old individuals and recent samples for living ones. However, the number of genotyped individuals during the first 10 years was low and variable, which led to gaps across years and weak relationships between genotyped individuals. During the following years more constant genotyping was done and since 2010 all the rams entering the AI center have been genotyped.

Table 3: Description of the total number of males (natural service –NS- or artificial insemination –AI-) and females, number of individuals with known complete pedigree (sire and dam) or with known dam. Total number of males with offspring records, mean offspring and distribution across total, NS and AI rams. Total number of females with records and mean number and distribution of records for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF). Based on data from 2000 to date.

Breed		Males					Females				
		No. (per year)	No. know pedigree (%)	No. known dam (%)	No. with offspring records	Mean offspring (min-max)	No. total (per year)	No. know pedigree (%)	No. known dam (%)	No. with records	Mean no. records (min-max)
LCNEUS	Total	6,719 (373)	3,960 (59)	1,863 (28)	1,021	26 (1-494)	82,987 (4,610)	36,445 (44)	35,045 (42)	81,477	3 (0-12)
	NS	6,102 (339)				7 (1-58)					
	AI	617 (34)				49 (1-494)					
LCR	Total	6,986 (388)	4,628 (66)	1,078 (15)	1,363	24 (1-440)	83,781 (4,654)	39,537 (47)	27,469 (33)	82,031	3 (0-11)
	NS	6,429 (357)				8 (1-57)					
	AI	557 (31)				61 (1-440)					
LCNNAF	Total	2,683 (149)	1,820 (68)	512 (19)	468	23 (1-258)	33,247 (1,956)	14,667 (44)	13,075 (39)	32,857	3 (0-12)
	NS	2,342 (130)				7 (1-43)					
	AI	341 (19)				35 (1-258)					

The genotyping platform was Illumina OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), that contains 54,241 markers. First, a general quality control was done for the three populations, consisting in removing animals with call rate <0.90 and parent-progeny Mendelian conflicts <5 %, and markers with call rate <0.97, minor allele frequency <0.05, monomorphic, or located in sexual chromosomes; resulting in 42,547 markers. A second quality control within breed was done keeping the same criteria and removing markers with heritability estimate of gene content <0.98 (Forneris et al., 2015). After quality control 353 LCNEUS, 427 LCR and 192 LCNNAF rams with 39,159, 38,168 and 39,373 effective SNPs respectively, were considered for the studies. Quality control was performed using PreGSF90 of the BLUPf90 software suite (Misztal et al., 2002). The three populations are historically related, but each population structure is well differentiated as the principal component analysis based on genomic information shows in Figure 5.

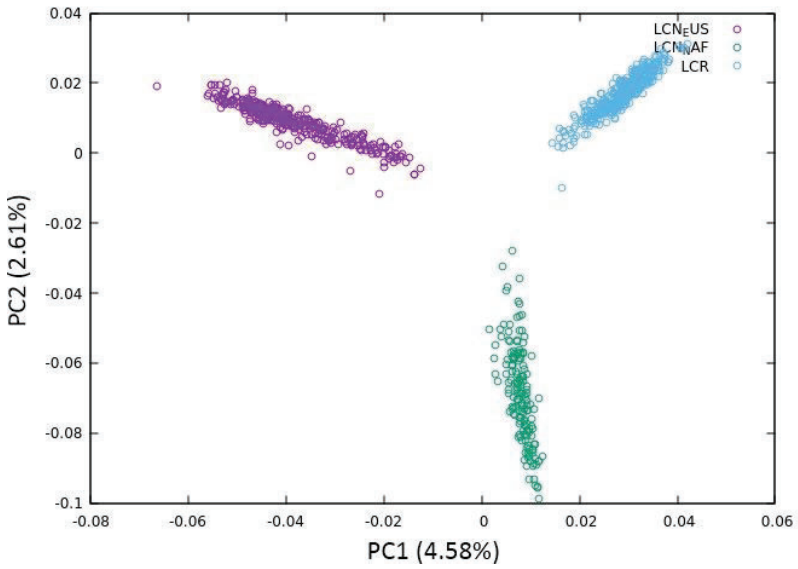

Figure 5: Principal component (PC) analysis for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Moreover, during the last years, alongside the genotyping of AI rams a substantial effort has been done to genotype some natural service (NS) rams and females. For those animals the AxiomTM Ovine Genotyping Array (Thermo Fisher Scientific Inc., Waltham, MA, USA) has been used and between both genotyping platforms, there were 37,121 common SNPs. This new genomic information has only been used in section 2.2, where there is an in-depth description of its characteristics.

4. References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Anim. Sci. 93:743–752.

Alfonso, L., A. Parada, A. Legarra, E. Ugarte, and A. Arana. 2006. The effects of selective breeding against scrapie susceptibility on the genetic variability of the Latxa Black-Faced sheep breed. Genet. Sel. Evol. 38(5):495–511.

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

Baloche, G., J. M. Astruc, P. Boulenc, B. Giral-Viala, P. Guibert, P. Panis, A. Legarra, and F. Barillet. 2014a. Genomic selection experiment in Lacaune dairy sheep: Progeny test results of rams initially selected either on parent average or on genomic prediction. In proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17–22 August. Vancouver, Canada.

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

Barillet, F. and M. Roussely. 1987. Mejora genética del ovino lechero en Francia: balance y perspectivas. ITEA, 72: 3-22.

Bennewitz, J., N. Reinsch, F. Reinhardt, Z. Liu, and E. Kalm. 2004. Top down preselection using marker assisted estimates of breeding values in dairy cattle. J. Anim. Breed. Genet. 121(5):307–318.

Berry, D. P., J. F. Garcia, and D. J. Garrick. 2016. Development and implementation of genomic predictions in beef cattle. Anim. Front. 6(1):32.

Bertolini, F., G. Schiavo, E. Scotti, A. Ribani, P. L. Martelli, R. Casadio, and L. Fontanesi. 2014. High-throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next-generation semiconductor-based sequencing. Anim. Genet. 45(2):304–307.

Boichard, D., S. Fritz, M. N. Rossignol, M. Y. Boscher, A. Malafosse, and J. J. Colleau. 2002. Implementation of marker-assisted selection in French dairy cattle. In proceedings of the 7th World Congress of Genetics Applied to Livestock Production. 19-23 August. Montpellier. France.

Bolormaa, S., J. E. Pryce, K. Kemper, K. Savin, B. J. Hayes, W. Barendse, Y. Zhang, C. M. Reich, B. A. Mason, R. J. Bunch, B. E. Harrison, A. Reverter, R. M. Herd, B. Tier, H.-U. Graser, and M. E. Goddard. 2013. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. J. Anim. Sci. 91(7):3088–3104.

Buisson, D., G. Lagriffoul, G. Baloche, X. Aguerre, P. Boulenc, F. Fidele, G. Frégeat, B. Giral-Viala, P. Guibert, P. Panis, C. Soulas, J. M. Astruc, and F. Barillet. 2014. Genomic breeding schemes in French Lacaune and Manech dairy sheep: Design and expected genetic gain. In proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17–22 August. Vancouver, Canada.

Bulmer, M. G. 1971. The effect of selection on genetic variability. Am. Nat. 105(943):201-211.

Carillier, C., H. Larroque, I. Palhière, V. Clément, R. Rupp, and C. Robert-Granié. 2013. A first step toward genomic selection in the multi-breed French dairy goat population. J. Anim. Sci. 96(11):7294–7305.

Christensen, O. F., and M. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2.

Clark, S. A., J. M. Hickey, H. D. Daetwyler, and J. H. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet. Sel. Evol. 44:4.

Council on Dairy Cattle Breeding. Accessed Nov. 22, 2020. https://queries.uscdcb.com/Genotype/cur_density.html

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124(6):369–376.

Daetwyler, H., A. Swan, J. van der Werf, and B. Hayes. 2012. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet. Sel. Evol. 44:33.

Dalloul, R. A., J. A. Long, A. V. Zimin, L. Aslam, K. Beal, L. A. Blomberg, P. Bouffard, D. W Burt, O. Crasta, R. P M A Crooijmans, K. Cooper, R. A. Coulombe, S. De, M. E. Delany, J. B. Dodgson, J. J. Dong, C. Evans, K. M. Frederickson, P. Flicek, L. Florea, O. Folkerts, M. A. M. Groenen, T. T. Harkins, J. Herrero, S. Hoffmann, H.-J. Megens, A. Jiang, P. de Jong, P. Kaiser, H. Kim, K.-W. Kim, S. Kim, D. Langenberger, M.-K. Lee, T. Lee, S. Mane, G. Marcais, M. Marz, A. P. McElroy, T. Modise, M. Nefedov, C. Notredame, I. R. Paton, W. S. Payne, G. Pertea, D. Prickett, D. Puiu, D. Qioa, E. Raineri, M. Ruffier, S. L. Salzberg, M. C. Schatz, C. Scheuring, C. J. Schmidt, S. Schroeder, S. M. J. Searle, E. J. Smith, J. Smith, T. S. Sonstegard, P. F. Stadler, H. Tafer, Z. J. Tu, C. P. Van Tassell, A. J. Vilella, K. P. Williams, J. A. Yorke, L. Zhang, H.-B. Zhang, X. Zhang, Y. Zhang, K. M. Reed. 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): Genome assembly and analysis. PLoS Biol. 8(9):e1000475.

Davis, G. H. 2005. Major genes affecting ovulation rate in sheep. Genet. Sel. Evol. 37(1):S11–23.

De la Fuente, L. F., G. Fernández, and F. San Primitivo, 1996. A linear evaluation system for udder traits of dairy ewes. Livest. Prod. Sci. 45:171–178.

Dekkers, J. C. 2004. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82:313–328.

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(5):2723–2733.

Elsen, J. M., Y. Amigues, F. Schelcher, V. Ducrocq, O. Andreoletti, F. Eychenne, J. V. Khang, J. P. Poivey, F. Lantier, and J. L. Laplanche. 1999. Genetic susceptibility and transmission factors in scrapie: detailed analysis of an epidemic in a closed flock of Romanov. Arch. Virol. 144(3):431–45.

Esfandyari, H., A. C. Sørensen, and P. Bijma. 2015. A crossbred reference population can improve the response to genomic selection for crossbred performance. Genet. Sel. Evol. 47(1):76.

Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21(4):467–477.

Fisher, R. A. 1918. The correlation between relatives on the supposition of mendelian inheritance. Trans. Roy. Soc. Edinb. 53:399–433.

Forneris, N. S., A. Legarra, Z. G. Vitezica, S. Tsuruta, I. Aguilar, I. Misztal, and R. J. C. Cantet. 2015. Quality control of genotypes using heritability estimates of gene content at the marker. Genetics 199:675–681.

Forutan, M., S. A. Mahyari, C. Baes, N. Melzer, F. S. Schenkel, and M. Sargolzaei. 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics 19:98.

Gabiña, D., E. Urarte, and J. Arranz. 1986. Métodos de simplificación del control lechero cuantitativo. Aplicación a las razas ovinas del País Vasco. Investigaciones agrarias: Producción y sanidad animales. 1: 259-270.

Gabiña, D., F. Arrese, J. Arranz, E. Urarte, I. Beltrán de Heredia, and J. M. López de Munáin. 1989. Heredabilidad y correlaciones genéticas de diversos criterios de estimación de la producción lechera en la raza Latxa. ITEA, Vol. Extra 9:367–369.

Gabiña, D., F. Arrese, J. Arranz, I. Beltrán de Heredia, and J. M. López de Munain. 1991. El programa de control lechero y selección en las razas Latxa y Carranzana de la Comunidad Autónoma Vasca. Informe Técnico Nº 38. Departamento de Agricultura y Pesca. Gobierno Vasco. Vitoria-Gasteiz. Spain.

Gabiña, D., E. Ugarte, and P. Santamaría. 2000. The definition of breeding objectives in the Latxa dairy sheep breed. Opt. Médit. 43:19–26.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 177(4):2389–2397.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5.

Haley, C. S., and P. M. Vischer. 1998. Strategies to utilize marker-quantitative trait loci associations. J. Dairy Sci. 81(2):85–97.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009a. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92(2):433–43.

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009b. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. 91:47–60.

Hazel, L. N. 1943. The genetic basis for constructing selection indexes. Genetics. 28(6):476–90.

Heidaritabar, M., A. Vereijken, W. M. Muir, T. Meuwissen, H. Cheng, H. J. Megens, M. A. M. Groenen, and J. W. M. Bastiaansen. 2014. Systematic differences in the response of genetic variation to pedigree and genome-based selection methods. Heredity. 113(6):503–513.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics. 31:423–447.

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in predicting of breeding values. Biometrics. 32:69–83.

Hsu, W. L., R. L. Fernando, J. C. M. Dekkers, J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, and A. Wolc. 2015. A simulation study on the effect of nested vs factorial mating on response to pedigree and genomic selection. In proceding of the Midwest Meeting of ADSA and ASAS. 16-18 March. Des Moines, Iowa, USA.

Hu, Z.-L., C. A. Park, X.-L. Wu, and J. M. Reecy. 2013. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res. 41:D871–D879.

Hurtado, A., A. L. García-Pérez, I. Beltrán de Heredia, J. Barandika, A. Sanz-Parra, E. Berriatua, and R. A. Juste. 2002. Genetic susceptibility to scrapie in a population of Latxa breed sheep in the Basque Country, Spain. Small Ruminant Res. 45(3):255–259.

Ibañez-Escriche, N. and Gonzalez-Recio, O. 2011. Review. Promises, pitfalls and challenges of genomic selection in breeding programs. Span. J. Agric. Res. 9(2):404–413.

Ibáñez-Escriche, N., S. Forni, J. L. Noguera, and L. Varona. 2014. Genomic information in pig breeding: Science meets industry needs. Livest. Sci. 166(1):94–100.

Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). 2020a. Informe sobre la evaluación genómica en razas ovinas de producción lechera. Accessed Feb. 24, 2021.

https://www.mapa.gob.es/en/ganaderia/temas/zootecnia/informeseleccciongenomicarazasovinodeleche_finaltridion_tcm38-542614.pdf

Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). 2020b. Informe sobre la constitución de una metapoblación de razas ovinas de producción lechera para el desarrollo de las evaluaciones genómicas. Unpublished report.

International Committee for Animal Recording (ICAR). 2018. International agreement of recording practices. Accessed Dec. 17, 2020. https://www.icar.org/Guidelines/16-Dairy-Sheep-and-Goats.pdf

International Sheep Genomic Consortium (ISGC). 2002. Accessed Dec. 17, 2020. https://www.sheephapmap.org/

Jiménez-Montero, J. A., O. González-Recio, and R. Alenda. 2012. Genotyping strategies for genomic selection in small dairy cattle populations. Animal. 6(8):1216–1224.

Jonas, E., and D. J. de Koning. 2015. Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front. Genet. 6:49.

Kemp, R. A. 1985. The effects of positive assortative mating and preferential treatment of progeny on the estimation of breeding values. Unpublished Doctoral thesis. University of Guelph. Guelph, Canada.

Kjetså, M. H., J. Ødegård, and T. H. E. Meuwissen. 2020. Accuracy of genomic prediction of host resistance to salmon lice in Atlantic salmon (*Salmo salar*) using imputed high-density genotypes. Aquaculture. 526:735415.

Lande, R., and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics. 124(3):743-56.

Larroque, H., F. Barillet, G. Baloche, J. M. Astruc, D. Buisson, F. Shumbusho, V. Clément, G. Lagriffoul, I. Palhiere, R. Rupp, C. Carillier, C. Robert-Granié, and A. Legarra. 2014. Toward genomic breeding programs in French dairy sheep and goats. In proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17–22 August. Vancouver, Canada.

Larzul, C., E. Manfredi, and J. M. Elsen. 1997. Potential gain from including major gene information in breeding value estimation. Genet. Sel. Evol. 29(2):161–184.

Legarra, A., J. Arranz, I. Beltrán de Heredia, and E. Ugarte. 1999. Sistema de calificación de morfología mamaria en ovejas de raza Latxa: Resultados preliminares. ITEA, Vol. Extra 20: 345–347.

Legarra, A., and E. Ugarte. 2001. Genetic parameters of milk traits in Latxa dairy sheep. Anim. Sci. 73:407–412.

Legarra, A., E. Ugarte, I. Beltrán de Heredia, and J. Arranz. 2001. Parámetros genéticos y respuesta a diferentes índices de selección de caracteres de morfología mamaria en la raza Latxa. ITEA, Vol. Extra 22:24–26.

Legarra, A. 2002. Optimización del esquema de mejora de la raza Latxa: análisis del modelo de valoración e introducción de nuevos caracteres en el objetivo de selección. Doctoral thesis. Departamento de Producción Agraria. Universidad Pública de Navarra. Pages: 165.

Legarra, A., P. López-Romero, and E. Ugarte. 2005. Bayesian model selection of contemporary groups for BLUP genetic evaluation in Latxa dairy sheep. Livest. Prod. Sci. 93:205–212.

Legarra, A., M. Ramón, E. Ugarte, M. D. Pérez-Guzmán, and J. Arranz. 2007. Economic weights of somatic cell score in dairy sheep. Animal. 1(2):205–212.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

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.

Lillehammer, M., T. H. E. Meuwissen, and A. K. Sonesson. 2011. Genomic selection for maternal traits in pigs. J. Anim. Sci. 89(12):3908–3916.

Lourenco, D. A. L., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguila, and I. Misztal. 2020. Single-Step genomic evaluations from theory to practice: Using SNP chips and sequence data in BLUPF90. Genes. 11(7),790:1–32.

Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common

reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43:43.

Lund, M. S., I. van den Berg, P. Ma, R. F. Brøndum, and G. Su. 2016. How to improve genomic predictions in small dairy cattle populations. Animal. 10(6):1042–1049.

Mefford, C. 2018. Angus genetic evaluation reaches historic milestone. Angus Communications. Accessed Mar. 26, 2021. https://www.angus.org/Pub/Newsroom/Releases/061918-agi-500k.html

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819–1829.

Meuwissen, T., B. Hayes, and M. Goddard. 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 6(1):6–14.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). In proceedings of the 7th World Congress on Genetics Applied to Livestock Production. 19-23 August. Montpellier, France.

Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92(9):4648–4655.

Molina, A., E. Muñoz, C. Díaz, A. Menéndez-Buxadera, M. Ramón, M. Sánchez, M. J. Carabaño, and J. M. Serradilla. 2018. Goat genomic selection: Impact of the integration of genomic information in the genetic evaluations of the Spanish Florida goats. Small Ruminant Res. 163:72–75.

Mouresan, E. F. 2016. Evaluation of the efficiency of the procedures of genomic selection in the autochthonous Spanish beef cattle populations. Doctoral thesis. Departamento Producción Animal y Ciencia de los Alimentos. Universidad de Zaragoza. Pages: 234.

Mrode, R., G. M. Tarekegn, J. M. Mwacharo, and A. Djikeng. 2018. Invited review: Genomic selection for small ruminants in developed countries: How applicable for the rest of the world? Animal. 12:1333–1340.

Obšteter, J., J. Jenko, J. M. Hickey, and G. Gorjanc. 2019. Efficient use of genomic information for sustainable genetic improvement in small cattle populations. J. Dairy Sci. 102(11):9971–9982.

Ogorevc, J., T. Kunej, A. Razpet, and P. Dovc. 2009. Database of cattle candidate genes and genetic markers for milk production and mastitis. Anim. Genet. 40(6):832–851.

Oliehoek, P. A., and P. Bijma. 2009. Effects of pedigree errors on the efficiency of conservation decisions. Genet. Sel. Evol. 41:9.

Palti, Y., G. Gao, S. Liu, M. P. Kent, S. Lien, M. R. Miller, C. E. Rexroad, and T. Moen. 2015. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. Mol. Ecol. Resour. 15(3):662–672.

Perez, B. C., J. C. C. Balieiro, R. Carvalheiro, F. Tirelo, G. A. Oliveira Junior, J. M. Dementshuk, J. P. Eler, J. B. S. Ferraz, and R. V. Ventura. 2019. Accounting for population structure in selective cow genotyping strategies. J. Anim. Breed. Genet. 136(1):23–39.

Quaas, R. L. 1976. Computing the diagonal elements of a large numerator relationship matrix. Biometrics. 32:949–953.

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.

Rothschild, M., C. Jacobson, D. Vaske, C. Tuggle, L. Wang, T. Short, G. Eckardt, S. Sasaki, A. Vincent, D. McLaren, O. Southwood, H. van der Steen, A. Mileham, and G. Plastow. 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pigs. Proceedings of the National Academy of Sciences of the United States of America. 93(1):201–5.

Rupp, R., S. Mucha, H. Larroque, J. McEwan, and J. Conington. 2016. Genomic application in sheep and goat breeding. Anim. Front. 6:39–44.

Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, and D. Altshuler. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 409:928–933.

Salaris, S., M. G. Usai, S. Casu, T. Sechi, A. Manunta, M. Bitti, S. Grande, and A. Carta. 2018. Perspectives of the selection scheme of the Sarda dairy sheep breed in the era of genomics. In proceedings of the 42nd ICAR Conference. 10–11 February. Auckland, New Zealand.

Samore, A. B., L. Buttazzoni, M. Gallo, V. Russo, and L. Fontanesi. 2015. Genomic selection in a pig population including information from slaughtered full sibs of boars within a sib-testing program. Animal. 9(5):750–759.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123(4):218–223.

Schöpke, K., and Swalve, H. H. 2016. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal. 10(6):1050–1060.

Schork, N. J., D. Fallin, and S. Lanchbury. 2000. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin. Genet. 58:250–264.

Shumbusho, F., J. Raoul, J. M. Astruc, I. Palhiere, and J. M. Elsen. 2013. Potential benefits of genomic selection on genetic gain of small ruminant breeding programs. J. Anim. Sci. 91:3644–3657.

Shumbusho, F. 2014. Designing, technical evaluation and profitability estimation of breeding strategies based on molecular information for small ruminant species. Doctoral thesis. Ecole doctorale Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB). Université de Toulouse. Pages: 160.

Silva, R. M. O., J. P. Evenhuis, R. L. Vallejo, G. Gao, K. E. Martin, T. D. Leeds, Y. Palti, and D. A. L. Lourenco. 2019. Whole-genome mapping of quantitative trait loci and accuracy of genomic predictions for resistance to columnaris disease in two rainbow trout breeding populations. Genet. Sel. Evol. 51(1):42.

Simianer, H., 2009. The potential of genomic selection to improve litter size in pig breeding programs. In proceedings of the 60th Annual meeting of the European Association of Animal Production. 24–27 August. Barcelona, Spain.

Sorensen, D. A., and B. W. Kennedy. 1983. The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. Theor. Appl. Genet. 66:217–220.

Thomasen, J. R., A. C. Sorensen, M. S. Lund, and B. Guldbrandsen. 2014. Adding cows to the reference population makes a small dairy population competitive. J. Dairy Sci. 97(9):5822–5832.

Ugarte, E., E. Urarte, F. Arrese, J. Arranz, L. Silio, and C. Rodríguez. 1996. Genetic parameters and trends for milk production of blond-faced Latxa sheep using Bayesian analysis. J. Dairy Sci. 79(12):2268–2277.

Ugarte, E., E. Urarte, J. Arranz, F. Arrese, C. Rodríguez, and L. Silió, 1997. Introducción de grupos genéticos en la evaluación de ovejas Latxa cara negra. ITEA, Vol. Extra 18:397–399.

Ugarte, E., A. Legarra, I. Beltrán de Heredia, and J. Arranz. 2001. Udder morphology: a new trait to introduce in the Latxa breeding programme. In proceedings of the 52nd Annual Meeting of the European Association of Animal Production. 26-29 August. Budapest, Hungary.

Usai, M. G., S. Salaris, S. Casu, T. Sechi, S. Miari, P. Carta, and A. Carta. 2018. Feasibility of genomic predictions of Sarda breed rams using a female reference population. In proceedings of the 11th World Congress on Genetics Applied to Livestock Production. 11–16 February. Auckland, New Zealand.

van den Berg, I., T. H. E. Meuwissen, I. M. MacLeod, and M. E. Goddard. 2019. Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. J. Dairy Sci. 102(4):3155–3174.

Van Eenennaam, A. L., K. A. Weigel, A. E. Young, M. A. Cleveland, and J. C. M. Dekkers. 2014. Applied animal genomics: Results from the field. Annu. Rev. Anim. Biosci. 2(1):105–139.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11):4414–4423.

VanRaden, P. M. 2020. Symposium review: How to implement genomic selection. J. Dairy Sci. 103(6):5291–5301.

Wang, Q., Y. Yu, J. Yuan, X. Zhang, H. Huang, F. Li, and J. Xiang. 2017. Effects of marker density and population structure on the genomic prediction accuracy for growth trait in Pacific white shrimp *Litopenaeus vannamei*. BMC Genetics. 18(1):45.

Weller, J. I., E. Ezra, and M. Ron. 2017. Invited review: A perspective on the future of genomic selection in dairy cattle. J. Dairy Sci. 100(11):8633–8644.

Wiggans, G.R., P.M. VanRaden, and T.A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, future. J. Dairy Sci. 94(6):3202–3211.

Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard. 2017. Genomic selection in dairy cattle: The USDA experience. Annu. Rev. Anim. Biosci. 5:309-327.

Wolc, A., H. H. Zhao, J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, C. Stricker, D. Habier, R. L. Fernando, D. J. Garrick, S. J. Lamont, and J. C. M. Dekkers. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. Genet. Sel. Evol. 47:59.

OBJECTIVES

OBJECTIVES

The main objective of this thesis is to assess the feasibility of implementing genomic selection in Latxa dairy sheep breeds, conditioned by the intrinsic characteristics of the breeds and the limitation of their small size.

This main objective can be disentangled in the following secondary goals:

1. Assess and compare different inbreeding and coancestry estimation methods to analyze the genetic diversity of Latxa dairy sheep and to study the consequences of genetic selection on inbreeding. Moreover, analyze the effect of French semen importation into Latxa Cara Rubia scheme by the genetic contribution to the current population and the evolution of coancestry between these breeds.
2. Take a new look into the effect of including genomic information into milk yield genetic evaluations comparing prediction accuracy of pedigree and genomic evaluations estimated by cross-validation, with updated data and better tools.
3. Compare the effect of different strategies to select the individuals to be genotyped, in terms of number and type of animals, into prediction accuracy of genomic evaluations in contrast to pedigree-based evaluations for milk yield.

CHAPTER 1

INBREEDING, EFFECTIVE POPULATION SIZE AND COANCESTRY IN THE LATXA DAIRY SHEEP BREED

Published in the scientific manuscript: Granado-Tajada, I., S. T. Rodríguez-Ramilo, A. Legarra, and E. Ugarte. 2020. Inbreeding, effective population size and coancestry in the Latxa dairy sheep breed. Journal of Dairy Science. 102:1–11. https://doi.org/10.3168/jds.2019-17743

1. INBREEDING, EFFECTIVE POPULATION SIZE AND COANCESTRY IN THE LATXA DAIRY SHEEP BREED

1.1. Abstract

Traditionally, breeding programs have estimated and managed inbreeding based on pedigree information. The availability of genomic marker panels has made possible new alternatives to achieve more precise estimates, for example in case of missing pedigree. The objective of the present study was to assess and compare, different estimation methods (pedigree-based methodologies, single SNP-based approach (homozygosity) and runs of homozygosity-based method) to analyze the evolution of genetic diversity measured as inbreeding or as coancestry of 3 selected populations of Latxa dairy sheep (Latxa Cara Rubia and Latxa Cara Negra from Euskadi and Navarre). Genomic data came from 972 artificial insemination rams genotyped with the Illumina OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) whose genealogy consisted of 4,484 animals. Inbreeding estimates based on molecular data were more similar between them than compared with those based on pedigree information. However, the SNP-based approach estimations of effective population size differed more, reflecting the sensitivity of effective population size to small changes in the evolution of inbreeding. The 2 Latxa Cara Negra populations showed increases of inbreeding rates with time and effective population sizes between 64 and 103 animals, depending on breed and methodology used. The Latxa Cara Rubia population did not show an increase in inbreeding rate, mainly due to semen importation from the related French population of Manech Tête Rousse. The effective size estimates based on coancestry increase show a higher variability and they are more sensitive to the source of information and the data structure considered. Realized effective population size based on individual increase in inbreeding were in agreement with the previous estimates. Coancestry evolution analysis based on DNA information showed an increase on coancestry during the last 10 yr in all breeds, as a consequence of the selection process. Moreover, the increase on coancestry between Latxa Cara Rubia and Manech Tête Rousse was more noticeable between than within each of those breeds.

1.2. Introduction

The Latxa is a dairy sheep breed autochthonous from the Basque Country and Navarre. Three strains are distinguished according to head color: Latxa Cara Rubia (**LCR**), Latxa Cara Negra

from Euskadi (**LCNEUS**), and Latxa Cara Negra from Navarre (**LCNNAF**). The breeding program started in 1984 and currently shows consolidated results with an annual genetic gain for milk yield between 3 and 4 %, depending on the strain (Ugarte and Legarra, 2003). The breeding program has 25,717, 34,077, and 12,087 adult ewes alive in 2019 for LCNEUS, LCR, and LCNNAF, respectively. Inbreeding is controlled by avoiding matings of relatives (cousins, sibs, and so on) and trying for new males entering the AI center to come from a diversity of paternal origins (Ugarte, 2007). However, a profound analysis of the evolution of inbreeding has not been done in the Latxa breed.

The same breed is located on the other side of the border in France. This breed, named Manech, is also subdivided into 2 strains according to head color: Manech Tête Rousse (**MTR**) and Manech Tête Noire (**MTN**), and both breeding programs started in the 1960s (Larroque et al., 2014). Due to geographical proximity and common use of pastures of the Pyrenees Mountains, animal exchanges between both sides of the borderline are customary. Moreover, in the case of LCR, importation of semen from MTR to the Spanish scheme has been systematic since 2010 to take advantage of the breeding program that was started earlier and is currently successful (Larroque et al., 2014). This connection and differences between breeds have been studied before (Ugarte et al., 1996; Legarra et al., 2014) and there is a consensus to consider that LCR and MTR are the same breed but separated geographically and with separate breeding schemes. The importation of semen from MTR could have affected the inbreeding in LCR and modified the coancestry relationship between both breeds.

Traditionally, the Latxa breeding program has estimated and managed inbreeding based on pedigree information, mainly by mating control (avoiding mating between close relatives) and by selecting lambs for reproduction center (choosing animals from different families and avoiding siblings). The inbreeding coefficient (**F**) of an individual is defined as the probability that 2 randomly chosen alleles at a homologous locus within an individual are identical by descent (**IBD**) where all the alleles in the base population are assumed to be different, meaning the alleles are identical because they are passed down from the same copy of a common ancestor (Malécot, 1948). Homozygosity caused by 2 IBD genomic segments is termed autozygosity, as opposed to allozygosity, which is homozygosity produced by alleles that are identical by state (**IBS**). The F is therefore an estimate of genome-wide autozygosity (Keller et al., 2011). The relationship between allozygosity and autozygosity is, for a biallelic locus, $IBD = \frac{(IBS - p^2 - q^2)}{1 - p^2 - q^2}$ (Li and Horvitz, 1953; Toro et al., 2011), where p and q (p + q = 1) are the allele frequencies on the base population.

However, the pedigree-based inbreeding coefficient ($\boldsymbol{F_{PED}}$) may not be an accurate measure of inbreeding in sheep breeding programs because they have heterogeneous pedigrees with a considerable proportion of unknown pedigree due to uncontrolled natural mating. In the case of Latxa breed, even though in recent years only 10 % of the animals born have both unknown parents, in the complete pedigree an average of 30 % of the animals have both parents unknown. Therefore, taking into account lack of full knowledge of pedigree, as described above, the estimation of inbreeding coefficients based on pedigree will be underestimated. This problem is not new and VanRaden (1992) suggested methods to compensate for missing pedigrees. In addition, pedigree estimates are not perfect due to pedigree errors (Oliehoek and Bijma, 2009), the lack of consideration of Mendelian sampling variation in IBD estimates (Hill and Weir, 2011), as well as due to linkage disequilibrium caused by selection (Forutan et al., 2018). The availability of genomic marker panels has made possible new alternatives to achieve more precise estimates of coancestry, mate assignment, and inbreeding coefficients than using pedigree records (VanRaden, 2008). There are several advantages of genomic *F* estimates over F_{PED}. First, genomic *F* can potentially estimate the proportion of genome autozygosity directly measuring homozygosity. Furthermore, genomic *F* can be estimated in wild populations in which pedigree data are hardly available, as well as bringing the possibility to analyze inbreeding across the genome by chromosome or gene regions (Keller et al., 2011).

To take advantage of these benefits, different methods have been developed to assess the effect of an active breeding program on inbreeding through SNP markers (VanRaden, 2008; Curik et al., 2014; Howard et al., 2017). One of the approaches is to estimate inbreeding SNP by SNP ($\boldsymbol{F_{SNP}}$), which is based on Malécot's (1948) definition: the probability that the 2 alleles of one individual are IBS instead of IBD (Caballero and Toro, 2002). Other useful predictors of inbreeding are the runs of homozygosity (**ROH**), which are continuous stretches of homozygous loci frequently present at genomic regions subjected to high selective pressures. The ROH-based inbreeding ($\boldsymbol{F_{ROH}}$) has been studied in human populations (McQuillan et al., 2008) and in livestock species (Curik et al., 2014; Peripolli et al., 2017), and several studies have reported their high reliability to estimate inbreeding (Ferenčaković et al., 2013; Hammerly et al., 2016; Forutan et al., 2018).

The availability of molecular information has also brought new insights into the estimation of coancestry (half the additive genetic relationship) between and within populations (Álvarez et al., 2005). Coancestry within populations is another parameter that classically has been estimated based on pedigree information. The coefficient of coancestry is useful mainly to plan

matings with the least inbreeding, considering that the inbreeding of an individual depends on the coancestry between their parents (Falconer and Mackay, 1996).

Therefore, the objective of the present study was to assess and compare, for the first time in Latxa breeds, different inbreeding and coancestry estimation methods (pedigree-based methodologies, single SNP-based approach, and ROH-based method) to analyze the genetic diversity of 3 populations of Latxa dairy sheep and to study the consequences of genetic selection on inbreeding. Moreover, the effect of French semen importation into LCR scheme was analyzed by the genetic contribution to the current population and the evolution of coancestry in these breeds.

1.3. Materials and methods

1.3.1. Molecular and Pedigree Data

The molecular information available was from AI rams, genotyped with Illumina OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), which were born between 1996 and 2016. As quality control, animals with call rate <0.90 and parent-progeny Mendelian conflicts (<5 %) were eliminated. For SNP, markers with heritability estimate of gene content <0.98 were removed (Forneris et al., 2015), as well as those with call rate <0.97, minor allele frequency <0.05, monomorphic or located in sexual chromosomes. Starting from the full pedigree, a sub-pedigree was generated with all known ancestors of the genotyped animals in each breed. Details about each population, the characterization of the pedigree (calculated with the software PEDIG by Boichard, 2002), and the evolution of knowledge of pedigree are shown in Table 1 and Figure 1, respectively. Regarding MTR introgressed animals, the identification of animals and their genealogical information is included into the LCR pedigree. The genomic information of the genotyped MTR rams present in the LCR pedigree was also considered.

Table 1: Number of genotyped individuals, number of markers, individuals in full and sub-pedigree, equivalent number of complete generations of full genealogy and generation interval of full genealogy for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR) and Latxa Cara Negra from Navarre (LCNNAF).

Breed	No. Genotyped rams	No. markers	No. full pedigree	No. sub-pedigree	Equivalent number of complete generations	Generation interval
LCNEUS	353	39,159	263,306	1,742	5.61	4.41
LCR	427	38,168	150,183	1,901	4.24	4.26
LCNNAF	192	39,373	68,714	841	4.65	4.65

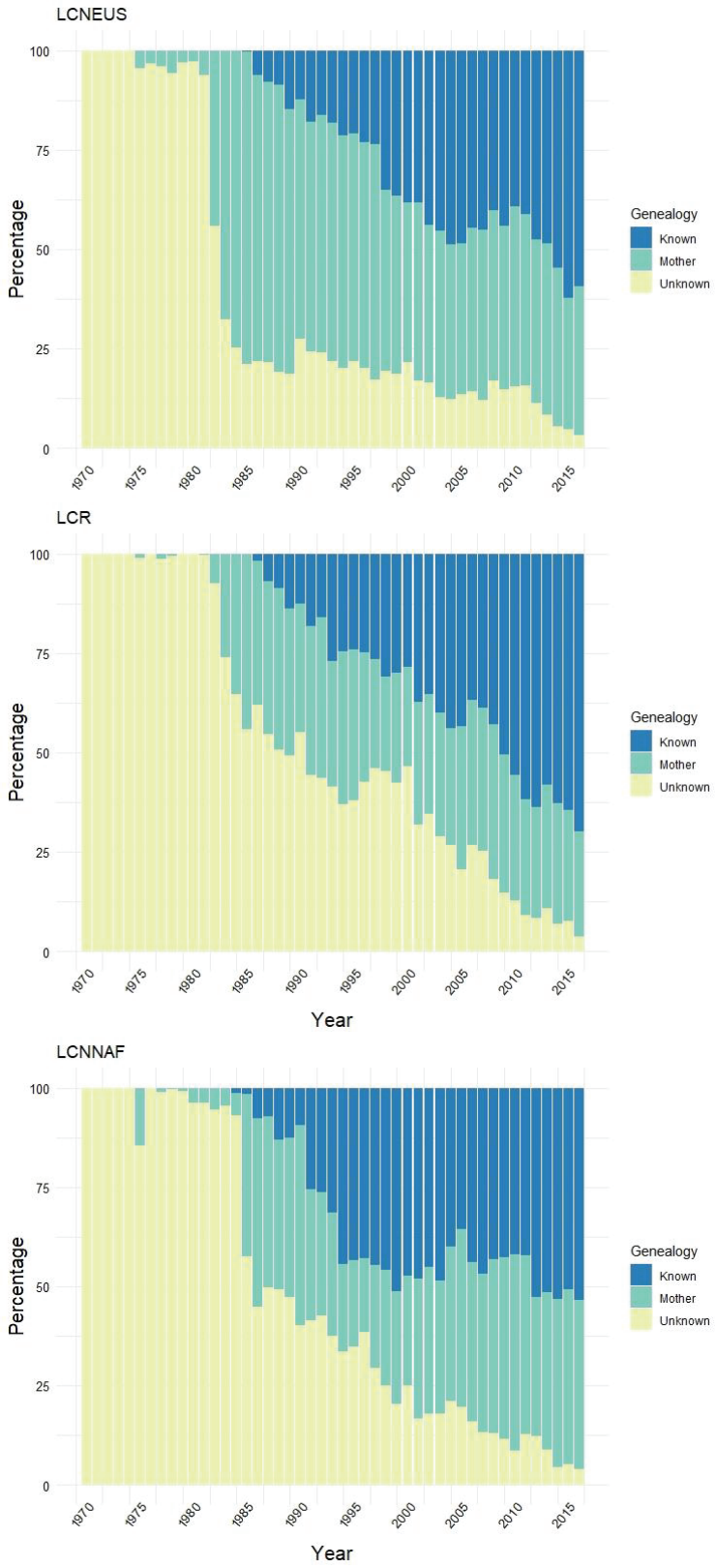

Figure 1: Evolution of known genealogy percentage by individual's year of birth, distinguishing when both parents are known (Known), only mother is known (Mother) and both parents are unknown (Unknown) for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR) and Latxa Cara Negra from Navarre (LCNNAF).

1.3.2. Inbreeding, Coancestry and Effective Population Size Estimates

The F_{PED} of genotyped rams were obtained with the program INBUPGF90 from the BLUPF90 family programs (Misztal et al., 2002), using the method published by Aguilar and Misztal (2008) based on the methodology proposed by VanRaden (1992), to calculate inbreeding accounting for missing parents. This method allows estimating inbreeding coefficients in populations with missing genealogy, assuming that the inbreeding coefficients of animals with missing parents are equal to the mean of the inbreeding coefficients of those animals with known parents and born during the same year. Ten iterations were used. Furthermore, pedigree-based inbreeding estimates of the generated sub-pedigree were used to analyze the evolution of mean inbreeding of the Latxa breeds during the last decades.

The inbreeding estimates of genotyped rams based on molecular information were estimated on a SNP by SNP basis (observed homozygosity, or IBS) and from ROH. The inbreeding estimate based on individual SNP (F_{SNP}) is the probability that 2 alleles are IBS measured across the genome and was calculated as the proportion of homozygous loci present among the typed genetic markers for each individual (Silió et al., 2013).

The inbreeding estimate based on ROH (F_{ROH}), was measured as the proportion of the genome that is contained in the ROH segments (McQuillan et al., 2008), calculated as

$$F_{ROH_i} = \frac{\sum_{k=1}^{n_{ROH_i}} l_{ROH_{ik}}}{l_g}$$

where n_{ROHi} is the total number of ROH in individual *i*, l_{ROHik} is the length of the *k* ROH in individual *i* in base pairs, and l_g is the total length of the genome in base pairs. The criteria established to define a ROH were as follows: the minimum length that constituted a ROH was 4 Mb, the minimum number of SNP was 30, the minimum density was 1 SNP per 100 Kb, the maximum distance allowed between 2 consecutive homozygous SNP in a run was 1 Mb, and a maximum of 2 missing genotypes and 1 heterozygous genotype within a particular ROH were permitted. The same criteria were used by Rodríguez-Ramilo et al. (2019) in the study done on the genetically close Manech breed.

The expected heterozygosity (Nei, 1973) is equivalent to the classical Malécot (1948) coefficient of coancestry (Caballero and Toro, 2002) and can be applied either to genealogical coancestry coming from pedigree information or to molecular coancestry measured from markers.

Pedigree-based coancestry (f_{PED}) coefficients for the genotyped individuals were calculated based on subpedigree information. Estimates were obtained using the software PEDIG (Boichard, 2002) with the option that implements the algorithm of Meuwissen and Luo (1992).

With molecular markers, the molecular coancestry between individuals *i* and *j* ($f_{SNP_{ij}}$) was estimated using Malécot's definition as the probability that 2 alleles drawn at random from each individual are IBS:

$$f_{SNP_{ij}} = \sum_{k=1}^{n_{SNP}} (I_{11,k} + I_{12,k} + I_{21,k} + I_{22,k})/4n_{SNP}$$

where n_{snp} is the total number of SNPs, $I_{xy,k}$ is an indicator variable that is 1 when both alleles of an individual in the SNP *k* are equal and 0 if they are not equal. It can be noted that $F_{SNP_i} = 2f_{SNPii} - 1$ (Caballero and Toro, 2002).

Rates of inbreeding per year of pedigree and genomic-based estimations were calculated as the regression coefficient of the inbreeding on the year of birth, considering in all cases only the inbreeding estimates of genotyped individuals. Then, the effective population size ($\boldsymbol{N_e}$) was calculated by $Ne = 1/2\Delta F$ (Falconer and Mackay, 1996) where ΔF represents the rate of inbreeding per year multiplied by the generation interval. Approximated estimations of confidence intervals for N_e were obtained from the standard error of the rates of inbreeding per year. The same procedure was applied to calculate N_e based on pedigree and molecular coancestry estimates.

In addition, realized effective population size ($\overline{N_e}$) estimates were calculated using ENDOG software (version v4.8; Gutiérrez and Goyache, 2005). These estimates were based on individual increase in inbreeding (ΔF_i) and its equivalent individual increase in coancestry (Δf_i), obtained from pedigree information (Cervantes et al., 2008, 2011; Gutiérrez et al., 2009).

1.3.3. Contribution of Founders and Evolution of Coancestry

To analyze the effect of historical importation of semen from French MTR into Spanish LCR, and to assess how that importation has affected the current population, the contribution of MTR to LCR population and its evolution were studied.

Considering pedigree information of LCR, which includes introgressed MTR individuals and their genealogy, the contribution of founders of each breed to the current population was estimated with the program ENDOG (version v4.8; Gutiérrez and Goyache, 2005), based on the

average relatedness coefficient of each individual. In this case, there is no correction for missing pedigree, so that there is an underestimation of average coancestry.

Additionally, another estimate of the introgression of MTR into LCR was obtained by analyzing the population structure based on genomic information. The software STRUCTURE (version 2.3.4; Pritchard et al., 2000) was used, with a model-based clustering algorithm and considering a subset of 19,084 SNP, by selecting every second marker of 427 LCR and 106 MTR individuals. The admixture model was employed, considering allele frequencies to be correlated between populations (Falush et al., 2003). A burn-in period of 10,000, 50,000 Markov chain Monte Carlo repeats, and 2 genetic clusters were used.

Moreover, taking into account the molecular coancestry (f_{SNP}) estimates of genotyped rams (calculated as previously described), the coancestry within each population and between LCR and MTR was analyzed. In addition, the evolution of average coancestry within and between these breeds during the last decade was analyzed comparing old and current rams, selecting the 25 % of the oldest and the 25 % of the youngest genotyped rams as described in Table 2.

Table 2: Distribution of genotyped rams across old and current categories and year of birth for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), Latxa Cara Negra from Navarre (LCNNAF), and Manech Tête Rousse (MTR).

Breed	No. genotyped rams	Old rams		Current rams	
		No.	Year of birth	No.	Year of birth
LCNEUS	353	98	1997-2005	96	2014-2016
LCR	427	96	1998-2005	97	2014-2016
LCNNAF	192	29	1996-2004	33	2013-2016
MTR	106	21	1999-2001	24	2009-2013

1.4. Results and Discussion

1.4.1. Inbreeding, Coancestry and Effective Population Size Estimates

Average inbreeding values for the genotyped AI rams of the studied breeds are shown in Table 3. The SNP-by-SNP based inbreeding estimates (F_{SNP}) were higher (0.6110 ± 0.0005 for LCNEUS) than those obtained with other methodologies because F_{SNP} does not distinguish between IBS and IBD. The difference is a constant and therefore irrelevant when the rate of increase of inbreeding per generation is calculated (Toro et al., 2011), which is actually a more relevant parameter (e.g., Howard et al., 2017). Pedigree and ROH-based inbreeding estimates of the genotyped animals showed average inbreeding values in the more usual scale of IBD

(F_{PED} 0.0183 ± 0.0012 and F_{ROH} 0.0333 ± 0.0012 for LCNEUS), with ROH-based estimates (F_{ROH}) being higher in all breeds. Mean inbreeding estimates considering all the individuals in pedigree are lower (0.0037 ± 0.0001 for LCNEUS) than current results (Granado-Tajada et al., 2019), as would be expected due to missing genealogy.

Table 3: Mean inbreeding (*F*) ± standard error (SE) and rate of inbreeding per generation (*ΔF*) ± SE; calculated with sub-pedigree information (PED), SNP by SNP (SNP) and runs of homozygosity (ROH) for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Parameter	Method	Breed		
		LCNEUS	LCR	LCNNAF
F mean ± SE	F_{PED}	0.0183 ± 0.0012	0.0159 ± 0.0004	0.0184 ± 0.0012
	F_{SNP}	0.6110 ± 0.0005	0.6076 ± 0.0004	0.6165 ± 0.0008
	F_{ROH}	0.0333 ± 0.0012	0.0268 ± 0.0009	0.0335 ± 0.0017
ΔF ± SE	ΔF_{PED}	0.0079 ± 0.0009	0.0022 ± 0.0004	0.0073± 0.0012
	ΔF_{SNP}	0.0012 ± 0.0004	-0.0001 ± 0.0004	0.0022 ± 0.0009
	ΔF_{ROH}	0.0054 ± 0.0010	0.0007 ± 0.0009	0.0049 ± 0.0018

The correlation coefficients between inbreeding estimates of genotyped rams are shown in Table 4. For the 3 breeds, the correlations between estimates based on molecular information were high (0.84, 0.84, and 0.87 for LCNEUS, LCR, and LCNNAF, respectively) and moderate-low between estimations based on pedigree and molecular data (0.40 and 0.28, 0.27 and 0.25, and 0.48 and 0.46 for ROH and SNP-based estimations for LCNEUS, LCR, and LCNNAF, respectively).

Table 4: Correlation coefficients of inbreeding estimates of genotyped rams, calculated based on sub-pedigree information (F_{PED}), SNP by SNP (F_{SNP}) and runs of homozygosity (F_{ROH}) for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

	Breed		
	LCNEUS	LCR	LCNNAF
F_{PED}-F_{SNP}	0.28	0.25	0.46
F_{PED}-F_{ROH}	0.40	0.27	0.48
F_{SNP}-F_{ROH}	0.84	0.84	0.87

Inbreeding estimates in the Manech Tête Noire and Tête Rousse (LCN and LCR equivalent breeds at France) were done with the same methodologies and reported slightly higher means (e.g., mean F_{PED} 0.0298 and 0.0239 for MTN and MTR, respectively), possibly linked with the fact that they have deeper and more complete genealogy (Rodríguez-Ramilo et al., 2019). Also identical criteria were used to define a ROH and estimate inbreeding in the Spanish Holstein

population, which showed a much higher mean F_{ROH} of 0.0770 (Rodríguez-Ramilo et al., 2015), as could be expected in dairy cattle due to stronger selection pressure.

Several studies have shown that characterizing inbreeding based on ROH provides a better measure of individual inbreeding than using pedigree information (e.g., Ferenčaković et al., 2013 or Forutan et al., 2018) and F_{ROH} have been widely studied in different livestock species (Peripolli et al., 2017). However, there is not an established, consistent, and reproducible criterion, which makes comparisons between studies challenging (Peripolli et al., 2017). Defining different minimum length of ROH is analogous to changing the depth of pedigree; shorter ROH display more ancient inbreeding due to recombination events as a function of the number of generations. However, longer ROH show more recent inbreeding because the probability of breaking up IBD segments from recombination is reduced (Curik et al., 2014). In the same way, all the criteria needed to define a ROH will affect the estimated ROH-based inbreeding (F_{ROH}). On this matter, Rodríguez-Ramilo et al. (2019) have analyzed the effect of 6 parameters to establish suitable criteria to define a ROH, based on Manech sheep data. They found that the minimum ROH length, the minimum number of SNP that constituted a ROH, as well as the minimum density and the maximum distance between 2 homozygous SNP, are ROH-defining factors with important implications in the estimation of the rate of inbreeding. However, inbreeding estimates do not change much unless extreme values are considered.

The rate of inbreeding per generation (ΔF) of genotyped rams followed a similar pattern as the mean inbreeding (Table 3), with pedigree-based estimates close to ROH-based ones (0.0079 ± 0.0009 and 0.0054 ± 0.0010 respectively, for LCNEUS) and higher than SNP-based estimates (0.0012 ± 0.0004 for LCNEUS), which were lower in all breeds. These ΔF are in agreement with those considered as admissible by Sonesson et al. (2012) when they analyzed the balance between genetic gain and inbreeding increase, which may indicate that there has been a suitable mating control and selection of AI rams. These results were consistent for LCN populations, but LCR showed low or almost zero rates (ΔF_{SNP} −0.0001 ± 0.0004). Figure 2 shows the inbreeding evolution of Latxa populations based on sub-pedigree information. Overall, the trend has been to increase, with differences between black and red populations. Both LCN populations showed a more noticeable increase from 2000 until now, whereas LCR inbreeding was rising since the 1990s to the end of 2000s, when it arrived at a plateau. This could be due to historical importation of semen and the systematic importation during the last 10 yr from the French MTR.

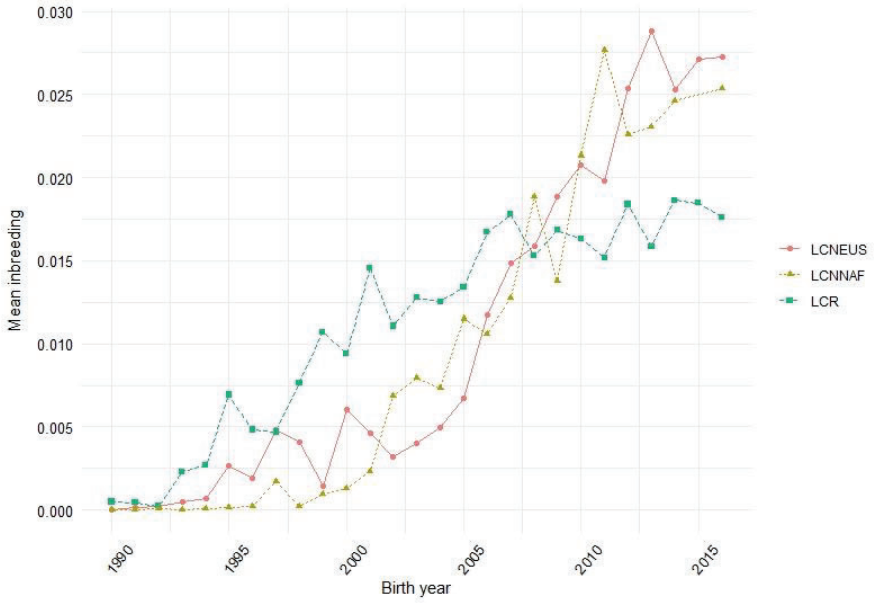

Figure 2: Evolution of inbreeding based on sub-pedigree data for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Effective population sizes are shown in Table 5 (in LCR breed, as the *F* increase per generation based on molecular data was almost zero, no estimates of N_e could be made). In the case of LCN breeds, N_e estimates were lower when pedigree data were used (64 and 69 for LCNEUS and LCNNAF, respectively) and slightly higher in ROH-based ones (93 and 103 for LCNEUS and LCNNAF, respectively). The N_e calculated with pedigree-based inbreeding coefficients of all the individuals in pedigree brings the highest estimates (288, 254, and 213 for LCNEUS, LCR, and LCNNAF, respectively; Granado-Tajada et al., 2019), but we have to keep in mind the effect of the missing genealogy that is only partially corrected using the method of VanRaden (1992). The SNP-based inbreeding estimates also showed high N_e (420 and 227 for LCNEUS and LCNNAF, respectively) but with large confidence intervals. Comparing methodologies, Rodríguez-Ramilo et al. (2019) reported similar results in related breeds, with pedigree and ROH-based N_e being smaller than SNP-based results. Rodríguez-Ramilo et al. (2019) found that the difference depends on the criteria used to define a ROH, and as soon as more genomic information is considered, the N_e estimates are more homogeneous and confidence intervals become more adjusted.

The estimated N_e based on inbreeding showed that, in spite of different population sizes, LCNEUS and LCNNAF have similar effective population sizes, reflecting the importance of the management done at the AI center to keep genetic variability. Nevertheless, it must be considered that our analysis was based on the available molecular information, which comes only from AI rams, which do not represent the whole population of rams contributing to the

population. Hence, it would be necessary to continue completing the information to have a more representative sample of the real population and more accurate effective population size estimates.

Table 5: Effective population size (N_e) estimates with 95 % confidence interval for the genotyped rams and realized effective size ($\overline{\boldsymbol{N_e}}$) ± standard error (SE); calculated with sub-pedigree information (PED), SNP by SNP (SNP) and runs of homozygosity (ROH); based on inbreeding rate per generation (*ΔF*), coancestry rate per generation (*Δf*), individual increase in inbreeding ($\overline{\Delta \boldsymbol{F_i}}$) and individual increase in coancestry ($\overline{\Delta \boldsymbol{f_i}}$) for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF). n/a= not available.

Information source	Method	Parameter	Breed		
			LCNEUS	LCR	LCNNAF
PED	*ΔF*	N_e (95 % CI)	64 (52 - 81)	233 (175 - 347)	69 (52 - 100)
SNP	*ΔF*		420 (246 - 1438)	n/a	227 (129 - 965)
ROH	*ΔF*		93 (69 - 143)	n/a	103 (59 - 389)
PED	*Δf*		202 (161-271)	493 (380-700)	n/a
SNP	*Δf*		253 (201-339)	414 (339-532)	n/a
PED	$\overline{\Delta F_i}$	$\overline{N_e}$ ± SE	83 ± 12	187 ± 24	81 ± 11
PED	$\overline{\Delta f_i}$		36 ± 1	86 ± 2	44 ± 2

The results are in agreement with the estimated population sizes for other sheep breeds with similar characteristics such as Manech Tête Noire (for which N_e was 53, 179, and 81 based on F_{PED}, F_{SNP}, and F_{ROH}, respectively) or Basco-Béarnaise (estimated Ne of 51, 114, and 59 based on F_{PED}, F_{SNP}, and F_{ROH}, respectively) defined with the same criteria as the present study (Rodríguez-Ramilo et al., 2019). For the Churra dairy sheep breed similar results were also found (N_e = 83) by Chitneedi et al. (2017). Other studies have reported effective population sizes between 88 and 1,317 in different milk and meat sheep breeds, which were calculated using linkage disequilibrium estimates (Kijas et al., 2012; Al-Mamun et al., 2015; Beynon et al., 2015) so they are not fully comparable.

Differences between results depending on the applied methodology are expected, given that each inbreeding coefficient is defined in relation to a different base population. In this study, F_{PED} is relative to the pedigree-depth; F_{SNP} is dependent on Hardy-Weinberg equilibrium and the set of loci included in the SNP chip, and F_{ROH} is relative to the ROH length (Rodríguez-Ramilo et al., 2019). Even though pedigree information has been traditionally used to manage

inbreeding (Sonesson et al., 2012), different studies have reported that IBD proportion is better predicted by a large number of genetic markers (Li et al., 2011; Kardos et al., 2015).

Calculations of the effective population size based on coancestry coefficients from sub-pedigree information as well as SNP-by-SNP were also performed (Table 5). The obtained results indicate agreement between N_e estimates using f_{PED} and f_{SNP}. However, for LCNNAF the N_e estimations were negative due to no increase in coancestry (Gutiérrez et al., 2008). Possibly, this is due to the unbalanced number of genotyped animals per year (*e.g.* 9 animals born in 2005 and 67 animals born in 2010). In some cases, to estimate the coancestry between animals of the same year of birth there are few animals and not many coancestry estimates. Results from pedigree indicate that estimates of effective population size obtained from coancestry were higher than estimates obtained from inbreeding, suggesting that matings between relatives are being performed. This is probably due to assortative mating of elite sires and dams to generate elite males, even if the mating of animals with any common grandparent is avoided. However, results obtained from SNP-by-SNP do not support this statement.

The estimates of the realised effective population size based on individual increase in inbreeding and in coancestry from sub-pedigree information are also shown in Table 5. The $\overline{N_e}$ results obtained from $\overline{\Delta F_i}$ were in agreement with the estimates based on ΔF_{PED} and ΔF_{ROH}. The $\overline{N_e}$ results obtained from $\overline{\Delta f_i}$ were different and the lowest among all the computed estimates. Other studies also showed different $\overline{N_e}$ estimates according to the applied methodology [*e.g.* the realized effective population size for Corse dairy sheep breed was 675 based on $\overline{\Delta F_i}$ and 220 based on $\overline{\Delta f_i}$ (Leroy et al., 2013)].

It should be considered that the obtained N_e and $\overline{N_e}$ estimates are likely to be conditioned by the small reference population size and the unbalanced number of genotyped individuals per year, which limits the amount of information taken into account.

1.4.2. Contribution of Founders and Evolution of Coancestry

After 10 yr of systematic semen importation, the coancestry study based on pedigree information reported that MTR individuals represent one-third (0.3266 ± 0.0021) of LCR population founder individuals, whereas LCR individuals are the principal contributors of the current population (0.6734 ± 0.0001). So the effect of French males used historically has been noteworthy and is reflected into the genetics of the current population. However, it must be considered that pedigree information reflects the vast majority of MTR rams used because current importation is done by AI, whereas the contribution of Latxa natural mating rams gets

lost in most cases due to it not being registered and non-systematic paternal filiations being done afterward. Therefore, the contribution of Latxa rams to the current population is likely underestimated.

Regarding the admixture proportion estimates based on genomic information, similar results were found. The LCR individuals showed a small proportion of MTR ancestry (0.3989 ± 0.0108), whereas the main ancestry corresponded to the LCR genetic cluster (0.6011 ± 0.0108). The results show a slightly higher effect of MTR into LCR genetics than based on pedigree information, although as a different source of information was considered, it is not directly comparable.

Additionally, the available molecular information allowed us to calculate the coancestry within breeds and between MTR and LCR breeds. Figure 3A shows that average coancestry across MTR males used in LCR was higher (0.6177 ± 0.0003) than across LCR males (0.6095 ± 0.0001). The coancestry across the 2 breeds, as we could guess, was lower (0.6078 ± 0.0001). Regarding the black populations, coancestry was higher than within the LCR breed and similar within MTR (0.6174 ± 0.0001 and 0.6199 ± 0.0002 for LCNEUS and LCNNAF, respectively). However, if the evolution of this parameter between old and current genotyped rams is considered (Figure 3B), it can be observed that the increase of coancestry between LCR and MTR after around 10 yr is more marked (0.0084) than the increase within each population (0.0058, 0.0046, and 0.0019 for LCR, MTR, and LCNNAF, respectively). The only population with a higher increase in coancestry was LCNNEUS (0.0108), possibly due to being a closed population with no importation of semen from French breeding programs, in contrast to LCR.

As time goes by, coancestry will continue increasing, so if the systematic importation of semen continues, in the future Latxa and Manech red populations will likely be genetically more similar and closer.

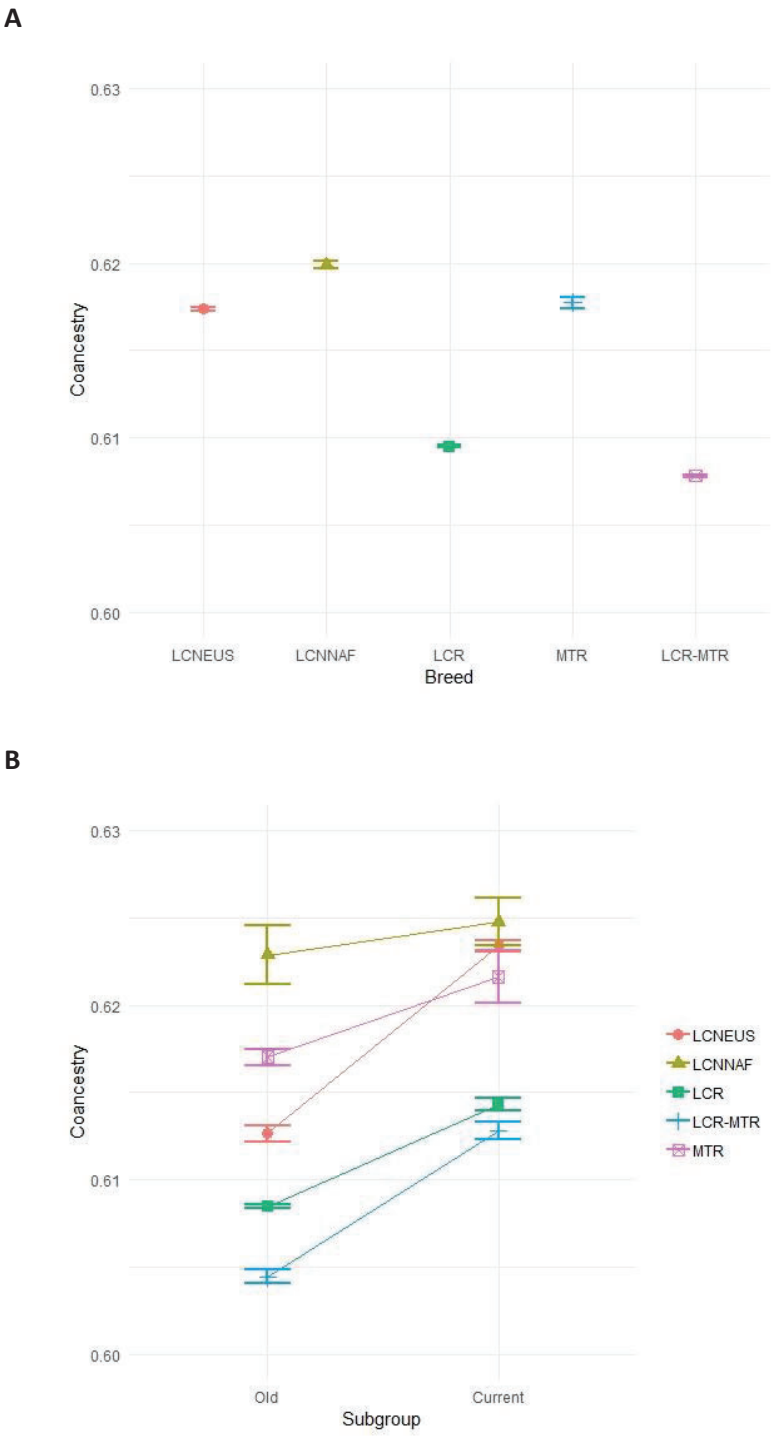

Figure 3: Coancestry within Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Negra from Navarre (LCNNAF), Latxa Cara Rubia (LCR), and Manech Tête Rousse (MTR) breeds and between LCR and MTR breeds. (A) Considering all genotyped rams and (B) considering old and current rams. Bars indicate 95 % confidence interval.

1.5. Conclusion

We have shown the genetic diversity available in the Latxa population. There are appreciable differences between estimates based on genealogy and molecular data, and also the applied methodology. When the proportion of homozygous SNP is used to estimate effective population size, results differ more than with the other methods. This confirms what other authors suggested about the sensitivity of proportion of homozygous SNP methodology when the molecular data are limited. Hence, molecular data are useful to infer inbreeding when pedigree information is not complete or deep enough and makes it possible to obtain more precise inbreeding and N_e estimates. The effective population size estimates based on coancestry increase show a higher variability and are more sensitive to the source of information and the data structure considered. Considering the evolution of the Latxa breeding program, the current inbreeding increase is moderate. Taking into account *F* and N_e estimates based on pedigree and ROH from the available and genotyped AI rams, it seems that there has been a suitable mating control and selection of lambs for AI, even though decisions were taken considering pedigree-based inbreeding estimates. Moreover, if systematic importation of French semen continues, Latxa and Manech red strains will be genetically closer. Therefore, it is necessary to continue controlling inbreeding to avoid reaching undesirable levels, and with this goal, molecular information gives us a new tool that allows a more accurate monitoring of inbreeding and Manech influence.

1.6. References

Aguilar, I. and I. Misztal. 2008. Technical Note: Recursive Algorithm for Inbreeding Coefficients Assuming Nonzero Inbreeding of Unknown Parents. J. Dairy Sci. 91:1669–1672.

Al-Mamun, H.A., S.A. Clark, P. Kwan, and C. Gondro. 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47:90–104.

Álvarez, I., J.P. Gutiérrez, L.J. Royo, I. Fernández, E. Gómez, J.J. Arranz and F. Goyache. 2005. Testing the usefulness of the molecular coancestry information to assess genetic relationships in livestock using a set of Spanish sheep breeds. J. Anim. Sci. 83:737–744.

Beynon, S.E., G.T. Slavov, M. Farre, B. Sunduimijid, K. Waddams, B. Davies, W. Haresign, J. Kijas, I.M. MacLeod, C.J. Newbold, L. Davies, and D. M. Larkin. 2015. Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. BMC Genet. 16:65.

Boichard, D. 2002. Pedig: A fortran package for pedigree analysis suited to large populations. In: 7th World Congress on Genetics Applied to Livestock Production. 19-23 May 2002. Montpellier. France.

Caballero, A., and M.A. Toro. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet. 3:289–299.

Cervantes, I., F. Goyache, A. Molina, M. Valera, and J.P. Gutiérrez. 2008. Application of individual increase in inbreeding to estimate effective sizes from real pedigrees. J. Anim. Breed. Genet. 125:301–310.

Cervantes, I., F. Goyache, A. Molina, M. Valera, and J.P. Gutiérrez. 2011. Estimation of effective population size from the rate of coancestry in pedigreed populations. J. Anim. Breed. Genet. 128:56–63.

Chitneedi, P., J.J. Arranz, A. Suarez-Vega, E. García-Gámez and B. Gutiérrez-Gil. 2017. Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. Anim. Genet. 48:436–446.

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.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to quantitative genetics. Harlow: Longman Scientific & Technical. New York, NY.

Falush D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics. 164:1567–1587.

Ferenčaković, M., J. Sölkner, and I. Curik. 2013. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. Genet. Sel. Evol. 45:42.

Forneris, N.S., A. Legarra, Z.G. Vitezica, S. Tsuruta, I. Aguilar, I. Misztal and R.J.C. Cantet. 2015. Quality control of genotypes using heritability estimates of gene content at the marker. Genetics. 199:675–681.

Forutan, M., S.A. Mahyari, C. Baes, N. Melzer, F.S. Schenkel and M. Sargolzaei. 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics. 19(1):98.

Granado-Tajada, I., S.T. Rodríguez-Ramilo, A. Legarra and E. Ugarte. 2019. Consanguinidad y censo efectivo de población en la raza ovina de leche Latxa Cara Negra. 405-407. AIDA XVIII Jornadas de Producción Animal.7-8 May. Zaragoza. Spain.

Gutiérrez, J.P. and F. Goyache. 2005. A note on ENDOG: a computer program for analysing pedigree information. J. Anim. Breed. Genet. 122:172–176.

Gutiérrez, J.P., I. Cervantes, A. Molina, M. Valera, and F. Goyache. 2008. Individual increase in inbreeding allows estimating realised effective sizes from pedigrees. Genet. Sel. Evol. 40:359–378.

Gutiérrez, J.P., I. Cervantes, and F. Goyache. 2009. Improving the estimation of realised effective population sizes in farm animals. J. Anim. Breed. Genet. 126:327–332.

Hammerly, S.C., D.A. de la Cerda, H. Bailey and J.A. Johnson. 2015. A pedigree gone bad: increased offspring survival after using DNA-based relatedness to minimize inbreeding in a captive population. Anim. Conservation. 19(3):296–303.

Hill, W.G. and B.S. Weir. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. Genet. Res. (Camb.) 93(1):47–64.

Howard, J.T., J.E. Pryce, C. Baes and C. Maltecca. 2017. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100:1–16.

Kardos, M., G. Luikart and F.W. Allendorf. 2015. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity. 115:63–72.

Keller, M.C., P.M. Visscher and M.E. Goddard. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189:237–249.

Kijas, J.W., J.A. Lenstra, B. Hayes, S. Boitard, L.R. Porto Neto, M. San Cristobal, B. Servin, R. McCulloch, V. Whan, K. Gietzen, S. Paiva, W. Barendse, E. Ciani, H. Raadsma, J. McEwan, B. Dalrymple and International Sheep Genomics Consortium. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10:e1001258.

Larroque, H., F. Barillet, G. Baloche, J.M. Astruc, D. Buisson, F. Shumbusho, V. Clément, G. Lagriffoul, I. Palhiere, R. Rupp, C. Carillier, C. Robert-Granié, and A. Legarra. 2014. Toward genomic breeding programs in French dairy sheep and goats. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17-22 August 2014. Vancouver. Canada.

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.

Leroy, G., T. Mary-Huard, E. Verrier, S. Danvy, E. Charvolin, and C. Danchin-Burge. 2013. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. Genet. Sel. Evol. 45(1):1.

Li, C.C. and D.G. Horvitz. 1953. Some methods of estimating the inbreeding coefficient. Am J Hum. Genet. 5:107–117.

Li, M.-H., I. Strandén, T. Tiirikka, M.-L. Sevón-Aimonen and J. Kantanen. 2011. A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. PLoS One. 6(11),e26256.

Malécot, G. 1948. Les Mathématiques de l'Hérédité. Masson & Cie, Paris, France.

McQuillan, R., A.L. Leutenegger, R. Abdel-Rahman, C.S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A.K. MacLeod, S.M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S.H. Wild, M.G. Dunlop, A.F. Wright, H. Campbell and J.F. Wilson. 2008. Runs of Homozygosity in European Populations. Am. J. Hum. Genet. 83:359–372.

Meuwissen, T.H., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24:305–313.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet and D.H. Lee. 2002. BLUPF90 and related programs (BGF90). In proceedings of the 7th World Congress on Genetics Applied to Livestock Production. 19-23 August 2002. Montpellier. France.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA. 70:3321–3323.

Oliehoek, P.A. and P. Bijma. 2009. Effects of pedigree errors on the efficiency of conservation decisions. Genet. Sel. Evol. 41:9.

Peripolli, E., D.P. Munari, M.V.G.B. Silva, A.L.F. Lima, R. Irgang and F. Baldi. 2017. Runs of homozygosity: Current knowledge and applications in livestock. Anim. Genet. 48:255–271.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics. 155: 945–959.

Rodríguez-Ramilo, S.T., J. Fernández, M. A. Toro, D. Hernández and B. Villanueva. 2015. Genome-Wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. PLoS ONE. 10(4),e0124157.

Rodríguez-Ramilo, S.T., J.M. Elsen and A. Legarra. 2019. Inbreeding and effective population size in French dairy sheep: comparison between genomic and pedigree estimates. J. Dairy Sci. 102(5):4227–4237.

Silió, L., M.C. Rodríguez, A. Fernández, C. Barragán, R. Benítez, C. Óvilo and A.I. Fernández. 2013. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. J. Anim. Breed. Genet. 130:349–360.

Sonesson, A.K., J.A. Woolliams and T.H.E. Meuwissen. 2012. Genomic selection requires genomic control of inbreeding. Genet. Sel. Evol. 44(1),27.

Toro, M.A., L.A. García-Cortés and A. Legarra. 2011. A note on the rationale for estimating genealogical coancestry from molecular markers. Genet. Sel. Evol. 43:1–10.

Ugarte, E., E. Urarte, F. Arrese, J. Arranz, L. Silio and C. Rodríguez. 1996. Genetic parameters and trends for milk production of blond-faced Latxa sheep using Bayesian analysis. J. Dairy Sci. 79:2268–2277.

Ugarte E., and Legarra A. 2003. Scientific background of the selection program in the Latxa breed. In : Gabiña D. (ed.), Sanna S. (ed.). Breeding programmes for improving the quality and safety of products. New traits, tools, rules and organization? Zaragoza: CIHEAM, 2003. p. 91-98. (Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 55). Meeting of the Sub-Network on Genetic Resources of the FAO-CIHEAM Inter-Regional Cooperative Research and Development Network on Sheep and Goats, 9-11 May 2002. Sassari. Italy.

Ugarte, E. 2007. The breeding program of Latxa breed. Biotech. In Anim. Husb. 23:97–111.

VanRaden, P.M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. J. Dairy Sci. 75:3136–3144.

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

CHAPTER 2

EXPLORING THE INCLUSION OF GENOMIC INFORMATION AND METAFOUNDERS IN THE LATXA DAIRY SHEEP GENETIC EVALUATIONS

Published in the scientific manuscript: Granado-Tajada, I., A. Legarra, and E. Ugarte. 2020. Exploring the inclusion of genomic information and metafounders in the Latxa dairy sheep genetic evaluations. Journal of Dairy Science. 103:6346–6353.
https://doi.org/10.3168/jds.2019-18033

2. INCLUSION OF GENOMIC INFORMATION IN THE LATXA DAIRY SHEEP GENETIC EVALUATIONS

2.1. EXPLORING THE INCLUSION OF GENOMIC INFORMATION AND METAFOUNDERS IN THE LATXA DAIRY SHEEP GENETIC EVALUATIONS

2.1.1. Abstract

The availability of genomic marker panels has made possible more precise estimates of breeding values. Sheep breeding programs are implementing genomic selection. In Latxa dairy sheep breed, a previous study using pre-corrected data and a small number of genotyped animals did not show a clear advantage of genomic selection. The objective of the present study was to ascertain the possible benefits of GS for the Latxa breed based on more data than before and using better tools, in particular single-step genomic BLUP using metafounders to model missing pedigree. Goodness of prediction of pedigree and genomic evaluations was analyzed by cross-validation comparing predictions of young rams using whole and partial (truncated) data sets. The results showed that with the current available data, genetic and genomic evaluations have the same accuracy. Contrary to the previous study, predictions were nearly unbiased, which shows the advantage of using single-step genomic BLUP. However, genomic information did not yield more precise evaluations. This could be explained by the small number of sibs in the young rams.

2.1.2. Introduction

The Latxa is a dairy sheep breed from the Western Pyrenees. Two strains are distinguished, mainly according to the skin color, the black strain Latxa Cara Negra and the red strain Latxa Cara Rubia (**LCR**). Moreover, the black strain is divided in two geographic areas with slight morphological differences: Euskadi (**LCNEUS**) and Navarre (**LCNNAF**). Three separate breeding programs exist for the three strains. Currently milk yield, milk composition, and udder morphology traits are being selected. Presently, the breeding program is well implemented and shows consolidated results with an annual genetic gain (**AGG**) between 0.19 and 0.23 standard deviations, depending on the strain.

With the development of genomics, SNP are being used to implement genomic selection (**GS**) in several livestock species. In small ruminants, GS is less frequent and generally gains in accuracy introducing GS are small (Mrode et al., 2018). Regarding dairy sheep, empirical studies to analyze accuracies of genomic predictions in the Lacaune breed showed that genomic estimated breeding value (**GEBV**) for milk yield was 0.15 more accurate than (pedigree-based) EBV (Duchemin et al., 2012; Baloche et al., 2014b). An experiment (selecting rams using either GEBV or EBV) confirmed the gain in AGG (Baloche et al., 2014a).

Because generation interval is already short (4 yr), the implementation of GS in dairy sheep depends critically on the ability to select accurately young rams. Modeling genomic versus classic scheme showed that AGG could increase up to 51.7 % with genomic information, with a reference population of at least 1,000 animals (Shumbusho et al., 2013). A performing GS scheme in Lacaune was conceived and implemented (Buisson et al., 2014). This scheme yields extra 10 to 20 % of gain in AGG and 20 to 40 % reduction of AI rams. The latter is important because in progeny-testing dairy sheep schemes using fresh semen, a large number of rams are waiting (and consuming resources) for first crop after their first use on AI, whereas a genomic scheme could avoid this waiting period based on breeding values estimated with a higher precision. Therefore, by modifying the management toward a GS scheme there could be fewer rams in the AI center, reducing costs and with no loss of genetic diversity.

In Western Pyrenees breeds, the implementation of GS was studied for the first time in 2014. Legarra et al. (2014) carried out a study with 6 breeds of dairy sheep, including the 2 strains of French Manech (which is a breed highly related to Latxa), Spanish Latxa with 3 strains, and Basco-Béarnaise. Results showed a gain in accuracy when genomic information was included [0.11 (**MTR**), and 0.06 for Basco-Béarnaise]; therefore, GS was started in 2017. However, the studies of Legarra et al. (2014) did not show a clear advantage of GS in Latxa. Increases in accuracy using GEBV were −0.18 (a deterioration) for LCNEUS, 0.30 for LCR and 0.04 for LCNNAF. Genomic predictions were generally less biased than pedigree predictions. However, some results are difficult to understand. For instance, for LCR EBV had almost 0 accuracy, which does not agree with the observed genetic progress of the breeding program. Predictions in some strains were frankly biased. Authors attributed these problems to poor genotyping (e.g., many “holes” across the years) and perhaps to poor modeling (pseudo-single-step GBLUP was used, using daughter yield deviations). Also, a weak relationship was present between genotyped rams in training and validation groups.

Since 2014, the Latxa breeding program has collected more phenotypic and pedigree data and has genotyped all the new males used in the last years in AI. Also, new developments in evaluation, such as single-step GBLUP (**ssGBLUP**), which enables the use of phenotypes of both animals with and without genotypes, yield better predictions (Legarra et al., 2009; Christensen and Lund, 2010). New methods in validation of results (**LR** method; Legarra and Reverter, 2018) may yield more objective measures of predictive ability. Furthermore, the metafounders (**MF**) theory (Legarra et al., 2015) provides a more coherent framework for genomic evaluation theory (Garcia-Baccino et al., 2017; Meyer et al., 2018; Bradford et al., 2019). In essence, the MF theory models missing pedigrees using populations (such as genetic groups), but these populations are of finite size and related according to observed genomic relationships of their descendants.

Thus, the objective of this work is to take a new look, with more data and better tools, into the effect of including genomic information into the genetic evaluations of LCR and LCNEUS strains. The LCNNAF population was not included in this study due to the small increase in number of genotyped rams.

2.1.3. Materials and Methods

2.1.3.1. Phenotypic and Pedigree Data

All data came from the Latxa Breeders' Associations' Confederation (**CONFELAC**). The full data set included in this study comprised 639,517 lactations and 263,308 animals in the LCNEUS pedigree and 392,109 lactations and 150,185 animals in the LCR pedigree. The evaluated trait was 120-d standardized milk yield whose heritability is 0.18 and 0.22 for LCNEUS and LCR, respectively. More details about phenotypic and pedigree data are shown in Table 1.

Table 1: Description of the pedigree and phenotypic data (milk yield in 120 days of lactation) included in the study for Latxa Cara Negra from Euskadi (LCNEUS) and Latxa Cara Rubia (LCR).

Data (1984-2017)	**Breed**	
	LCNEUS	**LCR**
No. of animals in pedigree	263,308	150,185
% animals with data and complete genealogy	22	35
No. of phenotypic records	639,517	392,109
No. of females in data	235,360	133,230
No. of flocks	456	325
No. of males with progeny in data	1,883	1,837

Because of the frequent use of natural mating with several rams, the breeding program only accepts filiations for AI rams. Also, several dams are unknown, mainly in flocks entering the breeding scheme. As a result, based on the average of the full data set, 37 % of the ewes with milk records have unknown sire and dam, even though since 2000 this percentage has decreased to 21 %. In addition, importation of semen from the French breed MTR to the Spanish LCR scheme has been sporadic since the 1990s and systematic since 2010. The ancestries of those MTR individuals included into LCR pedigree are unknown, and because the French breeding scheme started earlier, they should not be considered genetically equal to Spanish unknown contemporary individuals as was reported by Ugarte et al. (1996).

To palliate for these 2 kinds of missing pedigree, genetic groups (Quaas, 1988) are used for BLUP. However, the use of genetic groups in ssGBLUP is problematic (Misztal et al., 2013) and we decided to use MF (Legarra et al., 2015) to model missing pedigree. The assignment of MF was done looking for a balanced number of individuals in each MF group, considering the year of birth and the Spanish or French origin of the progeny. With this aim, for LCNEUS 11 MF groups were defined, from 1985 to 2015 every 3 yr; and for LCR 14 MF groups, from 1985 to 2015 every 3 yr for Spanish progeny and every 9 yr for French origin.

2.1.3.2. Genotypic Data

Using stored (for old) and recent DNA samples, AI rams were genotyped with the OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA). As quality control, animals with call rate <0.90 and parent-progeny Mendelian conflicts (<5 %) were eliminated. For SNP, markers with heritability estimate of gene content <0.98 were removed (Forneris et al., 2015), as well as those with call rate <0.97, minor allele frequency <0.05, monomorphic, or located in sex chromosomes. After quality control 780 rams (353 LCNEUS and 427 LCR) and 39,159 and 38,168 effective SNP for LCNEUS and LCR, respectively, were considered for genomic evaluations. Regarding the previous study (Legarra et al., 2014), this work includes 178 new genotyped individuals (80 LCNEUS and 98 LCR) corresponding to the rams entering the AI center from 2010 to 2014. Genotyping was irregular during the first years, but from 2011 all the rams entering the AI center have been genotyped, which enables a more balanced and stronger data structure. The distribution of genotyped rams by year of birth is shown in Figure 1.

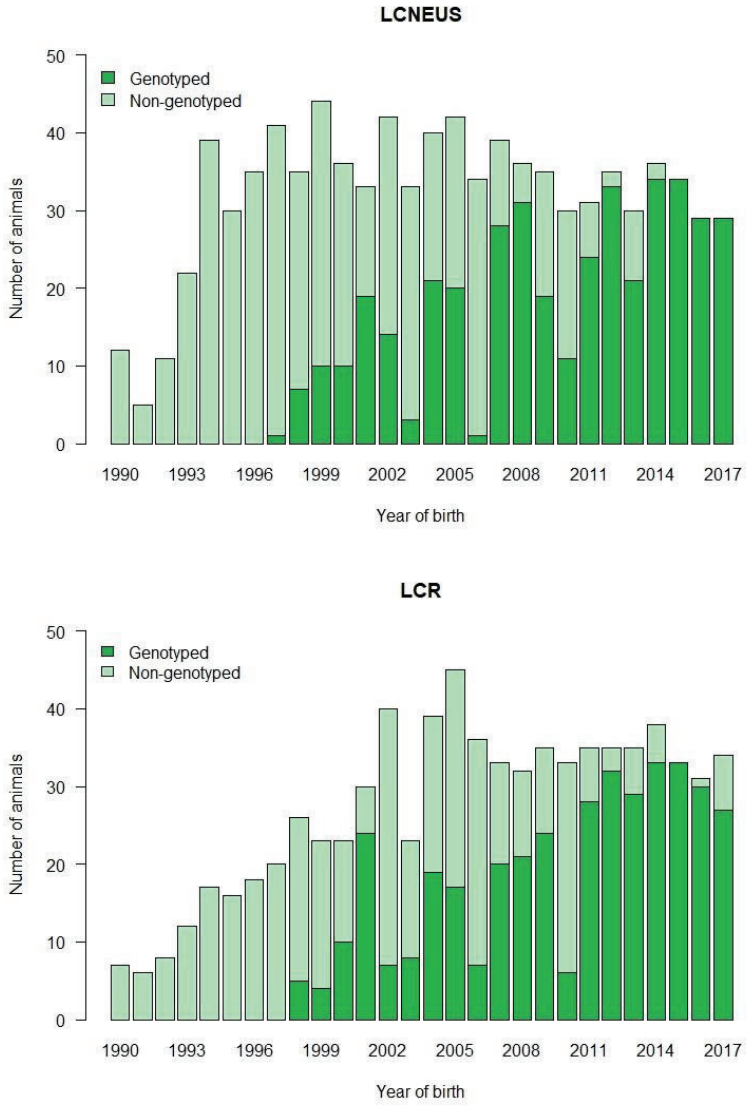

Figure 1: Rams used in artificial insemination by year of birth and the distribution of the genotyped individuals for Latxa Cara Negra from Euskadi (LCNEUS) and Latxa Cara Rubia (LCR).

2.1.3.3. Estimation of (Genomic) Breeding Values

To obtain (G)EBV the Blupf90 family programs (Misztal et al., 2002) were used, using either pedigree- based BLUP or ssGBLUP (Legarra et al., 2009; Christensen and Lund, 2010). In both cases the model for the milk yield included flock-year-season, parity number-age at lambing, number of alive lambs born, and interval between lambing and the first test day as fixed effects, and permanent environmental and additive genetic (Legarra et al., 2005) with MF as random effects. Strains were evaluated separately because they do not interbreed. The MF approach considers relationships within and across base populations. These relationships are summarized in a matrix **Γ**, which was estimated using the generalized least squares method in Garcia-Baccino et al. (2017). Supplemental Tables S1 and S2 (Annexe III) show **Γ** estimates for each breed.

2.1.3.4. Goodness of Prediction

The goodness of prediction of pedigree and genomic evaluations was analyzed by cross-validation using the LR method (Legarra and Reverter, 2018). This method measures biases and accuracies from global changes in (G)EBV of candidates to selection from a "partial" (1984–2014) to a "whole" (1984–2017) data set. We chose a set of validation individuals, considering genotyped AI rams born into the last 3 cohorts of the partial data set (between 2012 and 2014). The details of rams' distribution across validation and training group are described in Table 2. These animals were not progeny tested in the partial data set and have at least 10 daughters with lactations in the whole data set. First, GEBV were estimated in whole and partial data sets using ssGBLUP. Second, EBV were estimated in whole and partial data sets using BLUP. Statistics of bias, slope, and correlation $\rho_{w,p}$for validation individuals were calculated from the linear regression of GEBV estimated with whole data on GEBV estimated with partial data and from the linear regression of EBV estimated with whole data on EBV estimated with partial data.

Table 2: Number of rams distributed across categories, year of birth and average daughter number with lactations in the whole data set per genotyped ram for Latxa Cara Negra from Euskadi (LCNEUS) and Latxa Cara Rubia (LCR).

Breed	Rams in training group, AI progeny tested				Rams in validation group, AI progeny tested		
	No. of genotyped	No. of nongenotyped	Birth years	Average daughters	No. of genotyped	Birth years	Average daughters
LCNEUS	216	1,401	1984-2011	54.04	59	2012-2014	21.32
LCR	202	1,305	1983-2011	67.53	91	2012-2014	30.54

Bias defines the capacity to predict average value of candidates to selection and 0 would mean unbiasedness and is measured as the difference of average (G)EBV across whole and partial; slope could be taken as a measure of the dispersion of (G)EBV of candidates to selection, which should be 1, and it was obtained as the slope of the regression of (G)EBV estimated with whole data on (G)EBV estimated with partial data. Finally, we estimated the correlation ρw,p (G)EBV from partial data with (G)EBV from whole data, which could be understood as the relative increase of accuracy from partial to whole data. For instance, a value of 0.9 means that (G)EBV with partial data (with parent average or genomic information, but not offspring) were almost equal to (G)EBV with whole data (with progeny). In that sense, high values are expected

for genomic prediction, and low values for pedigree-based prediction. Due to the small size and selected character of validation groups, a bootstrap procedure with 1,000 iterations was implemented to sample with replacement the validation individuals and calculate the bootstrap distribution of bias, slope, and $\rho_{w,p}$ and their differences (e.g., the distribution of bias in ssGBLUP minus bias in BLUP).

Table 3: Considering validation rams, number of validation sires or maternal grandsires genotyped or nongenotyped for Latxa Cara Negra from Euskadi (LCNEUS) and Latxa Cara Rubia (LCR).

Breed	No. of rams in validation group	No. of sires in training group		No. of maternal grandsires in training group	
		Genotyped	Nongenotyped	Genotyped	Nongenotyped
LCNEUS	59	31	25	18	36
LCR	91	43	46	16	67

The relationship between and within training and validation group is known to be important for genetic prediction (Pszczola et al., 2012). For both breeds a high proportion of validation individuals have sire and maternal grandsire in the training group. In addition, around half of the validation rams have their sire genotyped, whereas one-third of the maternal grandsires are genotyped. In the case of nongenotyped relatives GS could still work, but with less advantage than with genotyped relatives (Pszczola et al., 2011). The described pedigree relationships (shown in Tables 3 and 4) may not be optimal, but compared with the previous study (Legarra et al., 2014) the reference population is larger and more related to the candidates for selection.

Table 4: Considering validation rams, number of full-sib rams, number of sires, sires with one or more than one ram, and rate of rams per sire for Latxa Cara Negra from Euskadi (LCNEUS) and Latxa Cara Rubia (LCR).

Breed	No. of rams in validation group	No. of full-sib rams	No. of sires			Rams/Sire
			Total	With 1 ram	With > 1 ram	
LCNEUS	59	2	20	8	12	2.95
LCR	91	2	36	22	14	2.53

2.1.4. Results and Discussion

Table 5 shows the main results of the cross-validation using the LR method. The bias of validation individuals was small, not significantly different from 0, and very similar in pedigree and genomic evaluations (the difference of estimated biases is 4.21 ± 2.63 and 4.08 ± 2.55, for LCNEUS and LCR, respectively). Note that these differences are expressed in liters, and the genetic standard deviation is 17.52 and 21.31 for LCNEUS and LCR, respectively. Across both breeds, BLUP had slopes closer to 1 than ssGBLUP, and LCR was significantly different from 1. Although the use of selectively genotyped individuals could affect the results (Vitezica et al., 2011), experience in dairy cattle and other dairy sheep breeds (VanRaden et al., 2009; Baloche et al., 2014b) shows that genotyping selected males does not result in bias because these animals will become average animals in the next generation. Moreover, a possible explanation for LCR lower estimates, which is hard to verify, is that imported MTR rams are a selected population and this was not correctly accounted for. Regarding the estimator of accuracies, $\rho_{w,p}$, predictions based on pedigree information had moderate-high $\rho_{w,p}$, 0.55 and 0.50 for LCNEUS and LCR, respectively. When predictions were based also in genomic information, $\rho_{w,p}$was slightly higher, 0.56 and 0.51 for LCNEUS and LCR, respectively. The comparison of BLUP and ssGBLUP shows that the inclusion of genotypes did not increase or decrease the accuracy of predictions for any of the breeds.

Table 5: Bias, slope and accuracy ± standard errors (by bootstrap) of pedigree and genomic evaluations and their difference (genomic-pedigree) for Latxa Cara Negra from Euskadi (LCNEUS) and Latxa Cara Rubia (LCR).

Breed	Methodology	Bias	Slope	$\rho_{w,p}$
LCNEUS	**Pedigree**	1.41 ± 6.41	0.96 ± 0.20	0.55 ± 0.09
	Genomic	5.62 ± 5.52	0.82 ± 0.17	0.56 ± 0.08
	Genomic-pedigree	4.21 ± 2.63	-0.14 ± 0.08	0.01 ± 0.04
LCR	**Pedigree**	7.66 ± 6.06	0.68 ± 0.12	0.50 ± 0.07
	Genomic	11.74 ± 5.32	0.57 ± 0.10	0.51 ± 0.07
	Genomic-pedigree	4.08 ± 2.55	-0.11 ± 0.05	0.01 ± 0.04

Comparing the current results with the previous ones (Legarra et al., 2014), in general predictions are more consistent, indicating that breeding values are more correctly estimated. Biases are lower and with considerably smaller standard errors, which were remarkably high in the previous study (around 60 for LCNEUS and 100 for LCR). The current slopes of the regression for both methodologies are closer to 1 than previous ones, which were around 0.30

in both LCNEUS and LCR. Accuracy of estimations is in all cases positive (whereas negative values were observed in the previous work) and with small standard errors, they are higher than the previous estimates and more coherent with the results of the breeding program. The change achieved in the pedigree-based EBV for LCR is especially relevant, whose accuracy was −0.05 ± 0.26, and the current one is 0.50 ± 0.07.

These better results may be associated with a stronger data structure (e.g., higher number of genotyped rams, closer relationship between reference and candidates to selection). In addition, the current evaluations were done with ssGBLUP, instead of using pseudo-ssGBLUP, avoiding daughter yield deviations calculation and transmission of inaccuracies to genomic predictions (daughter yield deviations in dairy sheep are not as accurate as in dairy cattle). Moreover, in these evaluations missing pedigree has been modeled using MF (Legarra et al., 2015), which is more consistent with genomic information. We did some analysis using genetic groups instead of MF, and results were not satisfactory because we obtained more biased evaluations and smaller slopes (results not shown).

The lack of increase of accuracy with genomic evaluations is influenced by the characteristics of the genotyped animals. VanRaden (2008) and Hayes et al. (2009) showed that the increase in accuracy from pedigree to genomic predictions can be explained as better relationships. Hayes et al. (2009) considered the case of sibships and showed that extra accuracy is a function of sibship size. In our case, the size of sibships is small (2–3 animals) and thus the expected increase using genomic is small. Moreover, only the rams that went through progeny testing have been genotyped, and these rams were selected based on classical pedigree evaluations (parent average). Selection of siblings is typically avoided to handle genetic diversity. Therefore, the set of genotyped rams and of candidates to selection in the validation population is not a representative sample of all of the population. However, this does not hamper GS because in practical GS the breeding scheme would genotype a larger number of half-sib candidates for selection.

Lacaune and MTR reported increases in accuracy using genomic information of 0.15 and 0.16 for milk yield, respectively. These results were based on bigger reference populations with 1,900 and 1,000 genotyped animals (Baloche et al., 2014b; Legarra et al., 2014). In fact, the correlation between gain in accuracy and the reference population size (and the heritability of the trait) is known to be high (Goddard, 2009; Daetwyler et al., 2012; Auvray et al., 2014). Moreover, the constitution of a big reference population based on AI rams could be difficult due to the general low use of AI and the small size of populations. But the decrease in

genotyping cost, combining with cheaper low-density chips and imputation methodologies, has brought the possibility of carrying on other strategies to maximize the benefits of genomic information (Rupp et al., 2016) by genotyping natural service rams or genotyping females, for instance. Genotyping other types of animals could be a good strategy and a simulation study is planned in Latxa.

In addition, we performed an extra analysis regarding the effect that genomic evaluations may have into lamb selection for AI center. Currently around 120 lambs of each breed are considered based on their EBV for progeny testing. These lambs come from the 5 % genetically best ewes, which have been inseminated by progeny-tested males (CONFELAC, 2016). Based on this working procedure, the selected lambs based on EBV or GEBV were compared and no dissimilarities were found; the same animals would be selected by both methodologies. The absence of sibships into the genotyped rams and the limited number of genotyped rams each year is hampering the benefit from the use of genomic information. A bigger genotyping population to preselect AI rams based on genomic values will increase the advantages of the methodology (Aguilar et al., 2010). So, a change in the breeding scheme organization would be necessary, like that proposed by Buisson et al. (2014), which provided a 15 % increase in AGG in Lacaune breed by genotyping 3 times the number of lambs required to fulfill the needs of AI center. Therefore, the implementation of GS needs to be tailored to the specific requirements of the breeding programs (Jonas and de Koning, 2015; Mrode et al., 2018).

2.1.5. Conclusion

No clear advantages have been found for GS of Latxa breeds. The characteristics and structure of each breed and breeding scheme are key factors that determine the possible beneficial effect of genomic information in terms of evaluation accuracy and AGG. We perceive that the number of genotyped animals (all of them AI males) is low and the link between genotyped individuals with phenotypic data and candidates to selection is not strong enough. Therefore, we have to solve these 2 handicaps before moving toward GS, and to have reliable information about how this change would affect the breeding scheme, further studies are planned in Latxa breeds.

2.2. RESULTS UPDATE

Since the work developed in the previous section 2.1 was published, the Latxa breeding program has continued recording data. During recent years, in addition to phenotypic and pedigree data and genomic information of all the new males used in artificial insemination (**AI**), a substantial effort has been made to genotype some natural service (**NS**) rams and females. Therefore, a considerable increase in the genotyped population size has been achieved.

Thus, given the current size and new characteristics of the genotyped population, an update of the latest results was found to be necessary to continue analysing the effect of including genomic information in the genetic evaluations of Latxa breeds taking into account the new information. Moreover, the considerable increase of genotyped individuals in the LCNNAF breeding program has allowed it to be included it in this study.

2.2.1. Materials and Methods

2.2.1.1. Phenotypic and Pedigree Data

As in the work shown in section 2.1, all data came from the Latxa Breeders' Associations' Confederation (CONFELAC). In this case, the full data set included was made up of 664,878 lactations and 272,049 animals in the LCNEUS pedigree, 431,692 lactations and 163,737 animals in the LCR pedigree and 197,081 lactations and 73033 animals in the LCNNAF pedigree. The evaluated trait was 120-d standardized milk yield whose heritability is 0.18, 0.22 and 0.25 for LCNEUS, LCR and LCNNAF, respectively. More details about phenotypic and pedigree data are shown in Table 6.

Table 6: Description of the phenotypic and pedigree data (milk yield in 120 days of lactation) included in the study for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR) and Latxa Cara Negra from Navarre (LCNNAF).

Data (1984-2019)	Breed		
	LCNEUS	LCR	LCNNAF
No. of animals in pedigree	272,049	163,737	73,033
% animals with data and complete genealogy	23	37	34
No. of phenotypic records	664,878	431,692	197,081
No. of females in data	242,848	144,993	65,060
No. of flocks	463	336	121
No. of males with progeny in data	2,046	2,054	810

As previously described, the breeding program only accepts filiations for AI rams. In this data set 37 % of the ewes with milk records have unknown sire and dam, even though since 2000 this percentage has decreased to 21 %. In comparison with the data used in the previous study, there are 26,600 new animals in pedigree and 78,800 newly recorded lactations, which represented a data increase of 3.65, 9.56 and 6.93 % for LCNEUS, LCR and LCNNAF, respectively.

2.2.1.2. Genotypic Data

Previously available genomic information came from AI rams genotyped with the OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA). During recent years the AxiomTM Ovine Genotyping Array (Thermo Fisher Scientific Inc., Waltham, MA, USA) has been used. Between both genotyping platforms, there were 37121 common SNPs. As quality control, the same parameters as in section 2.1 were applied: animals with call rate <0.90 and parent-progeny Mendelian conflicts (<5 %) were eliminated. For SNP, markers with heritability estimate of gene content <0.98 were removed (Forneris et al., 2015), as well as those with call rate <0.97, minor allele frequency <0.05, monomorphic, or located in sex chromosomes. After quality control 2267 individuals (917 LCNEUS, 923 LCR and 427 LCNNAF) and 32,077, 32,445 and 31,514 effective SNPs for LCNEUS, LCR and LCNNAF, respectively, were considered for genomic evaluations.

Table 7 shows a description by sex and characteristics of genotyped animals in each breed. Comparing with the previous study, there are 564, 496 and 235 new genotyped animals for LCNEUS, LCR and LCNNAF, respectively. The new genotyped individuals correspond to the new AI rams (11 %), but also NS rams (43 %) and females (46 %). Since 2011 all the rams entering the AI center have been genotyped, which enables a balanced and strong data structure. To complement AI rams, NS rams were genotyped because, even if they have fewer recorded daughters, they are connected with AI rams and their genomic information brought highly interesting information due to the low use of AI (71 % of the flocks, but 26 % of the adult ewes) and its low fertility rates (50 %). Regarding the genotyped females, they were mainly dams of AI rams born during the last decade. Figure 2 shows a graphical description of the distribution of genotyped animals by year and sex.

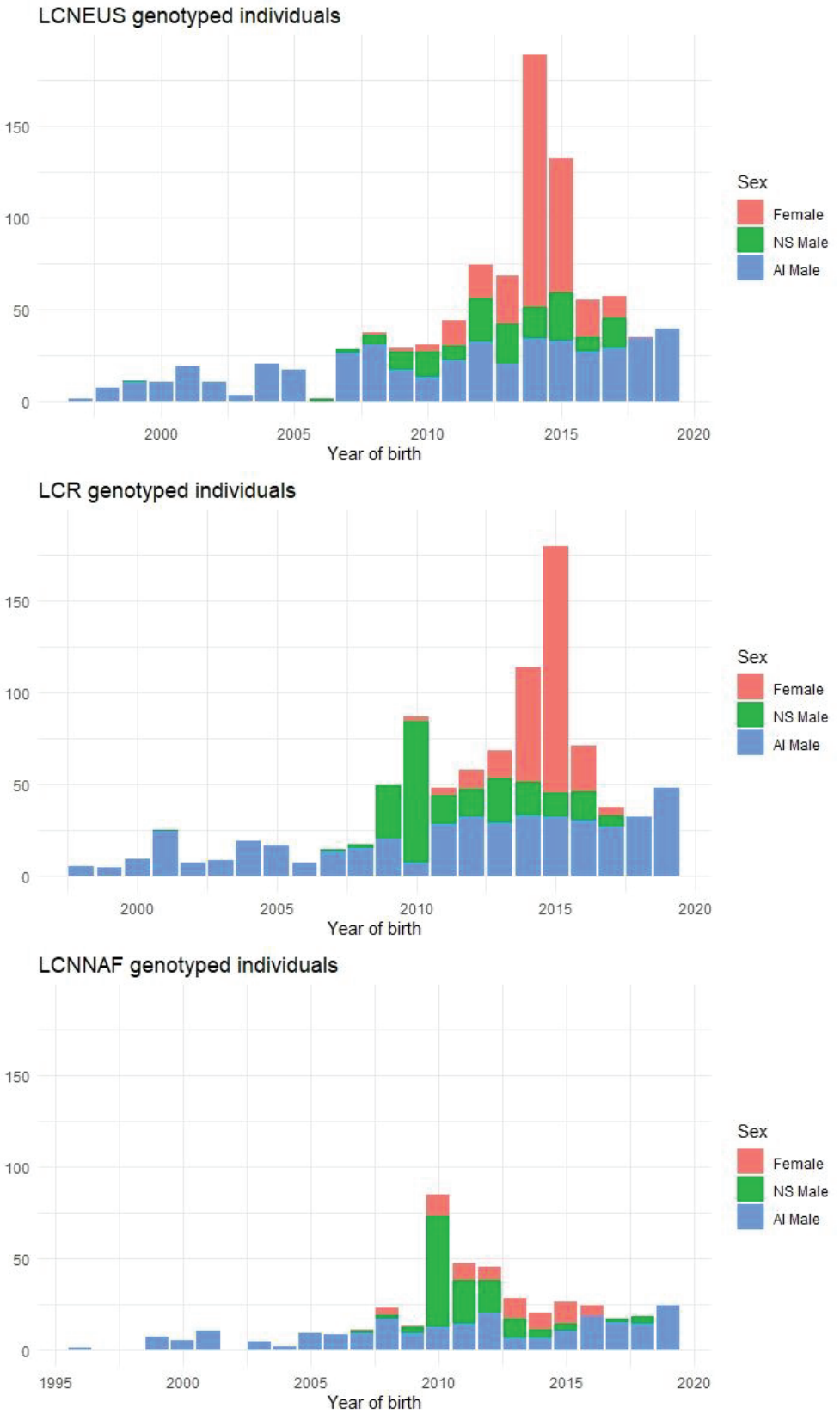

Figure 2: Distribution by year of birth of genotyped natural service (NS) rams, artificial insemination (AI) rams and females for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Table 7: Description of the genotyped individuals included, by natural service (NS) rams, artificial insemination (AI) rams and females, per year of birth and their average daughters or mean records for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Breed		Males			Females		
		No. of genotyped	Birth years	Average daughters	No. of genotyped	Birth years	Mean records
LCNEUS	**NS**	154	2006-2017	6.53	309	2008-2018	1.38
	AI	454	1997-2019	55.56			
LCR	**NS**	218	2007-2017	10.86	260	2010-2017	1.14
	AI	445	1998-2019	79.54			
LCNNAF	**NS**	135	2007-2018	11.02	72	2007-2016	2.35
	AI	220	1996-2019	76.88			

2.2.1.3. Estimation of (Genomic) Breeding Values

To obtain (G)EBV the BLUPf90 software suit (Misztal et al., 2002) was used, using either pedigree-based BLUP or ssGBLUP (Legarra et al., 2009; Christensen and Lund, 2010). In both cases the model for the milk yield was the same as that used in the previous section 2.1:

$$Y_{ijklm} = FYS_i + A_j + L_k + I_l + u_m + p_m + e_{ijklm}$$

Where flock-year-season (***FYS***), age-parity number (***A***), number of lambs born alive (***L***) and interval from lambing to first milk recording (***I***), were included as fixed effects (Legarra et al., 2005); additive genetic (***u***) with UPG (Misztal et al., 2013) and permanent environmental (***p***) were considered as random effects.

To palliate the missing pedigree genetic groups were used (Quaas, 1988) as defined in routine evaluations, instead of the metafounders used in the previous study. Latxa and Manech contemporary unknown ancestries were distinguished (Ugarte et al., 1996), because the French breeding scheme started earlier. Genetic groups or unknown parent groups (**UPG**) were assigned considering the year of birth and the Spanish or French origin of the progeny. In this way groups were defined from 1970 to 2019 every 3 years: 33 UPG for LCNEUS, 30 UPG for LCNNAF and 41 UPG for LCR, where 31 groups for Spanish progeny and 10 groups for French origin were differentiated.

Table 8: Number of rams and females distributed across categories, year of birth, and average daughter number with lactations or mean number of lactations recorded in the whole data set per genotyped individual for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Breed	Rams in training group, AI progeny tested				Ewes in training group, lactations recorded			Rams in validation group, AI progeny tested		
	No. of genotyped	No. of nongenotyped	Birth years	Average daughters	No. of genotyped	Birth years	Mean records	No. of genotyped	Birth years	Average daughters
LCNEUS	323	1,487	1984-2013	40.95	220	2008-2015	1.94	65	2014-2016	17.26
LCR	309	1,441	1983-2013	43.83	189	2010-2015	1.57	95	2014-2016	33.93
LCNNAF	185	537	1984-2013	56.07	55	2007-2015	3.07	27	2014-2016	32.71

Table 9: Considering validation rams, number of validation sires or maternal grandsires genotyped or nongenotyped for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Breed	No. of rams in validation group	No. of sires in training group		No. of dams in training group		No. of maternal grandsires in training group	
		Genotyped	Nongenotyped	Genotyped	Nongenotyped	Genotyped	Nongenotyped
LCNEUS	65	52	12	1	64	29	32
LCR	95	50	32	3	91	20	66
LCNNAF	27	21	6	25	2	10	13

2.2.1.4. Goodness of Prediction

The prediction accuracy of pedigree and genomic evaluations was analysed by cross-validation using the LR method (Legarra and Reverter, 2018) following the same criteria as described in section 2.1. Statistics of bias, dispersion of breeding values or slope, and correlation $\rho_{w,p}$ for validation individuals were calculated from changes in (G)EBV from a "partial" (1984–2016) to "whole" (1984–2019) data set. To keep the same structure as the previous section, validation individuals were AI rams born in the last 3 cohorts of the partial data set (between 2014 and 2016), which were not progeny tested in the partial data set and have at least 10 daughters with lactations in the whole data set. A bootstrap procedure with 1,000 iterations was also implemented to deal with the small size and selected character of validation groups. The details of distribution across validation and training group of genotyped individuals are described in Table 8.

The relationship between the reference population and the predicted animals (training and validation group) is a determining factors for the accuracy of GS (Pszczola et al., 2012; Wientjes et al., 2013), which is especially important when the reference population is small. Almost all the sires and maternal grandsires of validation individuals were found to be in the training group for the three breeds. Moreover, 70 % of the sires were genotyped, as were 34 % of the maternal grandsires. The black populations were shown to have a better relationship between training and validation individuals, whereas for LCR this was improvable. Furthermore, dams of predicted rams were also in training group, but were only genotyped for LCNNAF. The genotyped females of LCNEUS and LCR were younger, so only a few females were dams of validation individuals. The pedigree relationships described are shown in Table 9. Regarding the relationship within training and validation group, full-sib rams were unusual and more than two thirds were paternal siblings, the mean rate of rams being 2.38 per sire. Comparing the current relationship between and within training and validation groups with the previous study, a stronger relationship has been achieved.

2.2.2. Results and Discussion

Table 10 shows the bias, dispersion and accuracy results of the cross-validation using the LR method. Bias was very similar in pedigree and genomic predictions, and was closer to zero in black populations and moderately higher in LCR. Estimated slopes were also similar for BLUP and SSGBLUP evaluations, closer to one for LCNNAF and LCNEUS and lower for LCR. Worse estimates for LCR population were also described in the previous section 2.1, which was

attributed to the use of imported MTR rams coming from a selected population and may not have been correctly accounted for.

Table 10: Bias, slope, and accuracy ± SE (by bootstrap) of pedigree and genomic evaluations and their difference (genomic-pedigree) for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

Breed	Methodology	Bias	Slope	$\rho_{w,p}$
LCNEUS	Pedigree	4.20 ± 7.75	0.88 ± 0.16	0.57 ± 0.08
	Genomic	7.86 ± 5.43	0.79 ± 0.11	0.65 ± 0.07
	Genomic-pedigree	3.66 ± 4.91	-0.09 ± 0.11	0.08 ± 0.05
LCR	Pedigree	21.85 ± 9.12	0.59 ± 0.13	0.48 ± 0.09
	Genomic	23.37 ± 6.43	0.56 ± 0.08	0.55 ± 0.07
	Genomic-pedigree	1.51 ± 5.77	-0.03 ± 0.09	0.07 ± 0.05
LCNNAF	Pedigree	7.76 ± 6.59	1.02 ± 0.18	0.63 ± 0.12
	Genomic	8.97 ± 8.10	1.00 ± 0.20	0.66 ± 0.12
	Genomic-pedigree	1.21 ± 6.98	-0.02 ± 0.17	0.03 ± 0.08

In comparison with the results in section 2.1, current evaluations followed a similar trend. Nevertheless, slightly higher bias and lower slopes farther from one (except for LCNNAF) were found, which could be due to the use of UPG instead of metafounders (Legarra et al., 2015). UPG do not take into account the existing relationship between unknown groups of individuals as metafounders (**MF**) do, which leads to more biased predictions as reported in cattle and sheep (Bradford et al., 2019; Macedo et al., 2020; Junqueira et al., 2020). However, to use UPG in ssGBLUP evaluations to model missing pedigree allows one to avoid more biased results (Bradford et al., 2019; Macedo et al., 2020).

Regarding the accuracy of estimations, predictions were positive in all cases and the comparison of BLUP and ssGBLUP showed that the inclusion of genomic information increased the accuracy of predictions for the three breeds. Black populations were found to have the higher accuracies (around 0.61 for LCNEUS and 0.64 for LCNNAF). However, the bigger differences between genetic and genomic prediction accuracies were found in LCNEUS and LCR populations, with ssGBLUP evaluation 14 % more accurate than BLUP evaluations. These populations had a genotyped population size near to 1,000 individuals, which was close to the minimum set by Shumusho et al. (2013) of 1,000 genotyped individuals to achieve benefits from genomic selection over classical selection in dairy sheep. In the case of LCNNAF, the smaller genotyped population size (half of other breeds) was probably a limitation to have bigger gains in genomic prediction accuracy, which were found to be 5 % more accurate than pedigree-based evaluations.

The benefit of including genomic information in Latxa evaluations found is likely to be due to the increased number of genotyped individuals, together with the inclusion of natural service rams and females. Gains in prediction accuracy and genotyped population sizes are strongly related (Daetwyler et al., 2012; Auvray et al., 2014; Schöpke and Swalve, 2016) and the increase in the genotyped population size has been substantial. The current study was based on a genotyped reference population 2.5 times bigger than the work presented in section 2.1 and 3.6 times bigger than the data included in the study by Legarra et al (2014). The increase in genotyped population size was especially relevant for LCNNAF, because this population stopped genotyping for some years and it was not possible for it to be analysed in the cross-validation study of the previous section. Moreover, the increase in size kept improving the relationship between the reference population and predicted animals, a determining factor especially for the accuracy of GS in small reference populations (Pszczola et al., 2012; Wientjes et al., 2013). The relationship between predicted animals and genotyped sires and maternal grandsires were 0.4 better than in the previous section 2.1; 3.5 and 2 times better than in Legarra et al (2014). Therefore, compared with the previous studies, the current data showed a stronger relationship between training and validation groups.

Regarding the composition of the genotyped population, 46 % of the new data set were females (28 % over the total). This was the first time that female genomic information was included in Latxa evaluations, which could have been a key to obtaining a gain in accuracy of genomic predictions. The benefits of genotyped populations that include female genomic information on genomic prediction accuracy has been thoroughly studied (Koivula et al., 2016; Uemoto et al., 2017; Perez et al., 2019) and in species where the impact of female paths on genetic progress is strong has been described to be advantageous (Lourenco et al., 2015). Therefore, impact of genotype data on accuracy depends on the existing population structure, and the implementation of GS needs to be tailored to the specific characteristics and limitations of the breeding program (Jonas and de Koning, 2015; Mrode et al., 2018).

In dairy sheep breeds, and in general in small populations, building up a big reference population based on AI rams could be difficult due to the general low use of AI and the small size of populations, so genotyping females could be interesting to benefit from GS. The Italian Sarda breed made a female reference population of approximately 3,700 ewes that brought 0.13 more accurate genomic predictions for milk yield (Usai et al. 2018), whereas French dairy sheep breeds have an extensive use of AI and consequently better pedigree knowledge, so their genotyped populations were composed of AI rams. The inclusion of molecular

information was described to increase prediction accuracies of milk yield evaluations 0.15 for Lacaune breed and 0.16 for MTR, whose genotyped population sizes were of 1,900 and 1,000 individuals respectively (Baloche et al., 2014b; Legarra et al., 2014). These increases in accuracy were in agreement with the results found in Latxa breeds, given the smaller genotyped population sizes. So, although it is necessary to keep increasing the reference population, the implementation of a genotyped population composed of both sexes would be worthwhile for the Latxa breeding program.

2.2.3. Conclusion

The increase in number of genotyped individuals, together with the inclusion of natural service rams and females made possible to have a stronger relationship between genotyped animal and were essential to improve genomic predictions. The inclusion of genomic information in genetic evaluations brought little biased predictions for young rams, with small dispersion of EBVs and prediction accuracies were higher than pedigree based evaluations. LCNEUS and LCR showed bigger gains in accuracy of genomic predictions, but the inclusion of genomic information was also beneficial for LCNNAF. Including genotyped females in the reference population increased the accuracy of genomic predictions. So the implementation of a genotyped population composed of both sexes would be worthwhile.

2.3. References

Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

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

Baloche, G., J.M. Astruc, P. Boulenc, B. Giral-Viala, P. Guibert, P. Panis, A. Legarra, and F. Barillet. 2014a. Genomic selection experiment in Lacaune dairy sheep: progeny test results of rams initially selected either on parent average or on genomic prediction. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17-22 August 2014. Vancouver, Canada.

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

Bradford, H.L., Y. Masuda, P.M. VanRaden, A. Legarra, and I. Misztal. 2019. Modeling missing pedigree in single-step genomic BLUP. J. Dairy Sci. 102:1–11.

Buisson, D., G. Lagriffoul, G. Baloche, X. Aguerre, P.Boulenc, F. Fidele, G. Frégeat, B. Giral-Viala, P. Guibert, P. Panis, C. Soulas, J.M. Astruc, and F. Barillet. 2014. Genomic breeding schemes in French Lacaune and Manech dairy sheep: design and expected genetic gain. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17-22 August 2014. Vancouver, Canada.

Christensen, O.F., and M. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2.

CONFELAC. 2016. Programa de mejora de las razas ovinas Latxa y Carranzana. NEIKER-Tecnalia. Granja Modelo de Arkaute. Álava. Spain.

Daetwyler, H., A. Swan, J. van der Werf, and B. Hayes. 2012. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet. Sel. Evol. 44:33.

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.

Forneris, N.S., A. Legarra, Z.G. Vitezica, S. Tsuruta, I. Aguilar, I. Misztal, and R.J.C. Cantet. 2015. Quality control of genotypes using heritability estimates of gene content at the marker. Genetics. 199: 675–681.

Garcia-Baccino, C.A., A. Legarra, O.F. Christensen, I. Misztal, I. Pocrnic, Z.G. Vitezica, and R.J.C. Cantet. 2017. Metafounders are related to F_{st} fixation indices and reduce bias in single-step genomic evaluations. Genet. Sel. Evol. 49:34.

Goddard, M.E. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica (The Hague). 136:245–257.

Hayes, B.J., P.M. Visscher, and M.E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res., Camb. 91:47–60.

Jonas, E., and D.J. de Koning. 2015. Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front. Genet. 6:49.

Junqueira, V. S., P. S. Lopes, D. Lourenco, F. F. E. Silva, and F. F. Cardoso. 2020. Applying the metafounders approach for genomic evaluation in a multibreed beef cattle population. Frontiers in Genetics. 11,556399.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal. 10(6):1061–1066.

Legarra, A. P. López-Romero, and E. Ugarte 2005. Bayesian model selection of contemporary groups for BLUP genetic evaluation in Latxa dairy sheep. Livest. Prod. Sci. 93:205–212.

Legarra, A., Aguilar, I., and Misztal, I. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

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.

Legarra, A., O.F. Christensen, Z.G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. Genetics. 200:455–68.

Legarra, A., and A. Reverter. 2018. Semi-parametric estimates of population accuracy and bias of predictions of breeding values and future phenotypes using the LR method. Genet. Sel. Evol. 50:53.

Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, A. Legarra, and I. Misztal. 2015. Accuracy of estimated breeding values with genomic information on males, females, or both: An example in broiler chicken. Genet. Sel. Evol. 47(1):56.

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. Genetics Selection Evolution. 52(1):47.

Meyer, K., B. Tier, and A. Swan. 2018. Estimates of genetic trend for single-step genomic evaluations. Genet. Sel. Evol. 50:39.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D.H. Lee. 2002. BLUPF90 and related programs (BGF90). In: Proceedings 7th World Congress on Genetics Applied to Livestock Production. 19-23 August 2002. Montpellier, France.

Misztal, I., Z.G. Vitezica, A. Legarra, I. Aguilar, and A.A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation. J. Anim. Breed. Genet. 130(4):252–258.

Mrode, R., G.M. Tarekegn, J.M. Mwacharo, and A. Djikeng 2018. Invited review: Genomic selection for small ruminants in developed countries: how applicable for the rest of the world? Animal. 1333–1340.

Perez, B. C., J. C. C. Balieiro, R. Carvalheiro, F. Tirelo, G. A. Oliveira Junior, J. M. Dementshuk, J. P. Eler, J. B. S. Ferraz, and R. V. Ventura. 2019. Accounting for population structure in selective cow genotyping strategies. J. Anim. Breed. Genet. 136(1):23–39.

Pszczola, M., H.A. Mulder, and M.P.L. Calus. 2011. Effect of enlarging the reference population with (un)genotyped animals on the accuracy of genomic selection in dairy cattle. J. Dairy Sci. 94:431–441.

Pszczola, M., T. Strabel, H.A. Mulder, and M.P.L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. J. Dairy Sci. 95:389–400.

Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:1338–1345.

Rupp, R., S. Mucha, H. Larroque, J. McEwan, and J. Conington. 2016. Genomic application in sheep and goat breeding. Anim. Front. 6:39–44.

Schöpke, K., and H. H. Swalve. 2016. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal. 10(6):1050–1060.

Shumbusho, F., J. Raoul, J.M. Astruc, I. Palhiere, and J.M. Elsen. 2013. Potential benefits of genomic selection on genetic gain of small ruminant breeding programs. J. Anim. Sci. 91:3644–3657.

Uemoto, Y., T. Osawa, and J. Saburi. 2017. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. Animal. 11(3):382–393.

Ugarte, E., E. Urarte, F. Arrese, J. Arranz, L. Silio and C. Rodríguez. 1996. Genetic parameters and trends for milk production of blond-faced Latxa sheep using Bayesian analysis. J. Dairy Sci. 79:2268–2277.

Usai, M. G., S. Salaris, S. Casu, T. Sechi, S. Miari, P. Carta, and A. Carta. 2018. Feasibility of genomic predictions of Sarda breed rams using a female reference population. In proceedings of the 11th World Congress on Genetics Applied to Livestock Production. 11–16 February. Auckland, New Zealand

VanRaden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P.P., C.P. Van Tassell, G.R. Wiggans, T.S. Sonstegard, R.D. Schnabel, J.F. Taylor, and F.S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.

Vitezica, Z.G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res., Camb. 93:357–366.

Wientjes, Y. C. J., R. F. Veerkamp, and M. P. L. Calus. 2013. The Effect of Linkage Disequilibrium and Family Relationships on the Reliability of Genomic Prediction. Genetics. 193(2):621-631.

CHAPTER 3

GENOTYPING STRATEGIES FOR MAXIMIZING GENOMIC INFORMATION IN EVALUATIONS OF THE LATXA DAIRY SHEEP BREED

Published in the scientific manuscript: Granado-Tajada, I., Varona, L. and Ugarte, E. 2021. Genotyping strategies for maximizing genomic information in evaluations of the Latxa dairy sheep breed. Journal of Dairy Science. *In Press*. https://doi.org/10.3168/jds.2020-19978

3. GENOTYPING STRATEGIES FOR MAXIMIZING GENOMIC INFORMATION IN EVALUATIONS OF THE LATXA DAIRY SHEEP BREED

3.1. Abstract

Genomic selection has been implemented over the years in several livestock species, due to the achievable higher genetic progress. The use of genomic information in evaluations provides better prediction accuracy than do pedigree-based evaluations, and the makeup of the genotyped population is a decisive point. The aim of this work is to compare the effect of different genotyping strategies (number and type of animals) on the prediction accuracy for dairy sheep Latxa breeds. A simulation study was designed based on the real data structure of each population, and the phenotypic and genotypic data obtained were used in genetic (BLUP) and genomic (single-step genomic BLUP) evaluations of different genotyping strategies. The genotyping of males was beneficial when they were genetically connected individuals and if they had daughters with phenotypic records. Genotyping females with their own lactation records increased prediction accuracy, and the connection level has less relevance. The differences in genotyping females were independent of their estimated breeding value. The combined genotyping of males and females provided intermediate accuracy results regardless of the female selection strategy. Therefore, assuming that genotyping rams is interesting, the incorporation of genotyped females would be beneficial and worthwhile. The benefits of genotyping individuals from various generations were highlighted, although it was also possible to gain prediction accuracy when historic individuals were not considered. Greater genotyped population sizes resulted in more accuracy, even if the increase seems to reach a plateau.

3.2. Introduction

The procedures of genetic evaluation of livestock animals have been revolutionized over the last 2 decades by the introduction of the genomic selection (**GS**; Meuwissen et al., 2001). Its implementation has been shown to be beneficial in simulations (Meuwissen et al., 2001), as well as in several livestock species (VanRaden et al., 2009; Ibañez-Escriche and Simianer, 2016; VanRaden, 2020). There is consensus on the benefits of GS for shortening the generation interval and increasing prediction accuracies of genetic values, especially in those individuals without their own phenotypic data. Therefore, its implementation makes it possible to achieve

higher genetic gains compared with classical selection (Ibañez-Escriche and Gonzalez-Recio, 2011).

The reliability of genomic prediction is influenced by several factors, including the specific features of the trait of interest (heritability, genetic architecture, number and distribution of genes, and linkage disequilibrium between prediction markers and QTL), the characteristics of the population, and the design of the genotyped reference population (size, sex ratio, or the relationship of genotyped individuals within them, with candidates for selection, and with the overall population; Lund et al., 2016; Schöpke and Swalve, 2016; van den Berg et al., 2019).

There is a general agreement that before the application of GS to a breeding program, it is important to have adapted it to the biological, productive, and economic circumstances of the population (Boichard et al., 2016). In small ruminant populations, as is the case of Latxa breeds, the potential gain in accuracy provided by molecular information may be lower than in other species because of the low linkage disequilibrium (Kijas et al., 2014), due to higher effective population size and introgression of other populations (Ibañez-Escriche and Gonzalez-Recio, 2011; Rupp et al., 2016). In addition, the population size is usually small, they have a short generation interval, the use of AI is limited, and there is a considerable percentage of unknown parents in the pedigree.

Nevertheless, there are examples that highlight how the implementation of GS could benefit dairy sheep breeding programs, and in general, small populations. The Italian Sarda breed found that genomic predictions of rams were 0.13 more accurate for milk yield and 0.21 for milk fatty acid composition based on a female reference population (Usai et al., 2018; Cesarani et al., 2019a). French dairy sheep breeds used AI rams in the reference genotyped population, due to their extensive use. In the Lacaune breed, the inclusion of molecular information (compared with traditional evaluations) increased accuracies of predicted breeding values between 0.10 and 0.20, according to the trait (Baloche et al., 2014). Similar trends were described for milk yield evaluations of Manech and Basco-Béarnaise breeds, with those being between 0.06 and 0.16 more accurate than pedigree-based evaluations (Legarra et al., 2014).

The size and characteristics of the genotyped population depends on the species and the breeding program, and its composition plays an important role in prediction accuracy. Gains in prediction accuracy and genotyped population sizes are strongly related (Daetwyler et al., 2012; Auvray et al., 2014; Schöpke and Swalve, 2016). However, breeding programs usually have limited economic resources for genotyping, making the decision to select candidates a

much more cautious one. To solve this common issue, a large amount of research focuses on the design of genotyping strategies to optimize the selection of individuals to be genotyped. In this respect, there are studies based on small dairy cattle data, sheep data, simulations of the effect of selecting individuals genetically related to candidates for selection (Hayes et al., 2009; Habier et al., 2010; Clark et al., 2012), or genotyping females selected randomly, by EBV, EBV accuracy, or phenotypic value (Jiménez-Montero et al., 2012; Gao et al., 2015; Cesarani et al., 2019b). Moreover, in small dairy cattle, increasing the genotyped population size with data from other populations of the same breed, or from different but related breeds, is known to be beneficial (Lund et al., 2016; Schöpke and Swalve, 2016).

The Latxa breed is a dairy sheep breed autochthonous from the Western Pyrenees. Three strains are distinguished according to head color: Latxa Cara Rubia (**LCR**), Latxa Cara Negra from Euskadi (**LCNEUS**), and Latxa Cara Negra from Navarre (**LCNNAF**). Each strain has a separate breeding program. Breeding objectives are milk yield, milk composition, and udder morphology traits. The breeding program started in 1984 and is now well established, showing consolidated results with an annual genetic gain for milk yield between 0.19 and 0.23 standard deviations, depending on the strain (Granado-Tajada et al., 2020).

Although GS is not implemented in Latxa breeding programs, several exploratory studies have been developed. Legarra et al. (2014) found inconsistent results when the accuracy of genomic predictions was evaluated, attributed to the distribution of genotypes across the population and the weak link between genotyped individuals with phenotypic data and candidates for selection. After some years of systematic genotyping directed toward solving these handicaps, Granado-Tajada et al. (2020) looked again into the effect of including genomic information; more coherent results were found, although the inclusion of genotypes did not increase the accuracy of predictions for any of the Latxa breeds. Only AI rams were genotyped, which is a common strategy because males drive the genetic structure of the population and provide high predictive accuracy, due to the information from their daughters, as Jiménez-Montero et al. (2012) stated for dairy cattle.

Therefore, given the importance of the makeup of the genotyped population in the accuracy of genetic evaluations (Clark et al., 2012), a simulation study was designed based on the real data structure of the Latxa breeds. Thus, the aim of this work is to compare several selection strategies to form the genotyped population, in the 3 Latxa ecotypes, LCR, LCNEUS, and LCNNAF, by comparing the prediction accuracy of genomic evaluations against pedigree-based evaluations.

3.3. Materials and Methods

3.3.1. Genealogical, Phenotypic and Genomic Data

The simulation study was based on the available genealogical, phenotypic, and genomic information. Genealogical data comprised 263,308 individuals for LCNEUS, 150,185 for LCR, and 68,714 for LCNNAF. Individuals recorded into genealogy are the ones that remain in the flock (75 % of females and 8 % of males born from AI). We do not have information about the removed animals, which is especially relevant for males, due to the high selection pressure. Moreover, in the full data set, 25 % of the ewes with milk records have known sire and dam, even though in data since 2000, this percentage has increased to 42 %. This is because the breeding program only recognizes as fathers the AI males or natural service males after paternal filiations. Details about the males and females included in pedigree and candidates to be genotyped are shown in Table 1.

Regarding phenotypic data, we used 120 d standardized milk yield information with 639,517; 392,109; and 183,251 records of 235,360; 133,230; and 61,309 ewes for LCNEUS, LCR, and LCNNAF, respectively. The phenotypic means and standard deviations were 134 ± 57 for LCNEUS, 148 ± 68 for LCR, and 143 ± 57 for LCNNAF.

The genomic data consisted of genotypes of 353 LCNEUS, 427 LCR, and 192 LCNNAF AI rams selected for milk yield, and in an effort to keep a bigger genetic diversity, animals from different families were chosen and siblings avoided. The AI rams diffuse their genetics through the population by having many daughters and being the sires of natural service rams; therefore, they are expected to be representative of each population. They were genotyped with the Illumina OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), and as quality control, markers with call rate <0.97 or minor allele frequency <0.05, and markers that were monomorphic or located in sexual chromosomes were removed, resulting in 42,547 markers. The imputation of missing genotypes was conducted with FImpute software (Sargolzaei et al., 2014).

Table 1: Total number (and mean per year) of male and female candidates to be genotyped recorded in pedigree (1987-2017) and number with complete pedigree; males classified as artificial insemination (AI) or natural service (NS) rams, number with offspring records, mean offspring and distribution; and mean lactation records and distribution for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF).

	Males						Females		
	Total (per year)	No. with pedigree (per year)	No. AI (per year)	No. NS (per year)	No. with offspring records	Mean offspring (min-max)	Total (per year)	No. with pedigree (per year)	Mean lactations (min-max)
LCNEUS	12,882 (416)	5,924 (191)	848 (27)	12,034 (388)	1,932	26 (1-681)	167,011 (5,567)	51,951 (1,676)	3 (1-12)
LCR	9,947 (321)	5,926 (191)	668 (22)	9,279 (299)	1,977	24 (1-445)	122,158 (4,072)	47,070 (1,518)	3 (1-12)
LCNNAF	4,419 (143)	2,658 (86)	522 (17)	3,897 (126)	866	25 (1-457)	56,404 (1,880)	20,424 (681)	3 (1-12)

3.3.2. Simulation

The simulation was carried out based on the work done by Mouresan et al. (2017, 2018) with a fortran 90 code, available upon request. A summary of the simulation structure is presented in Figure 1. The genomic information available was used to create the base generation. Based on these genotypes, 3 discrete generations of 1,000 individuals were generated by gene-dropping (MacCluer et al., 1986), with random mating and no sex differences, with the aim of capturing the linkage disequilibrium structure and evolve toward the overall analyzed population. Every individual had a genome of 26 chromosomes, where the 42,547 markers were distributed asymmetrically, as shown in Supplemental Table S1 (Annexe IV). Out of all the markers, a 5 % were randomly selected as causative mutations (QTL) from a Gaussian distribution with mean zero and variance one; a recombination rate of 1 % per Mb was fixed. The last simulated generation was used to generate the genotypes of the founders of the real pedigree of each population. The aim of this procedure was to obtain simulated genotypes with the same genealogical structure as in the real populations, to be as close as possible to real data.

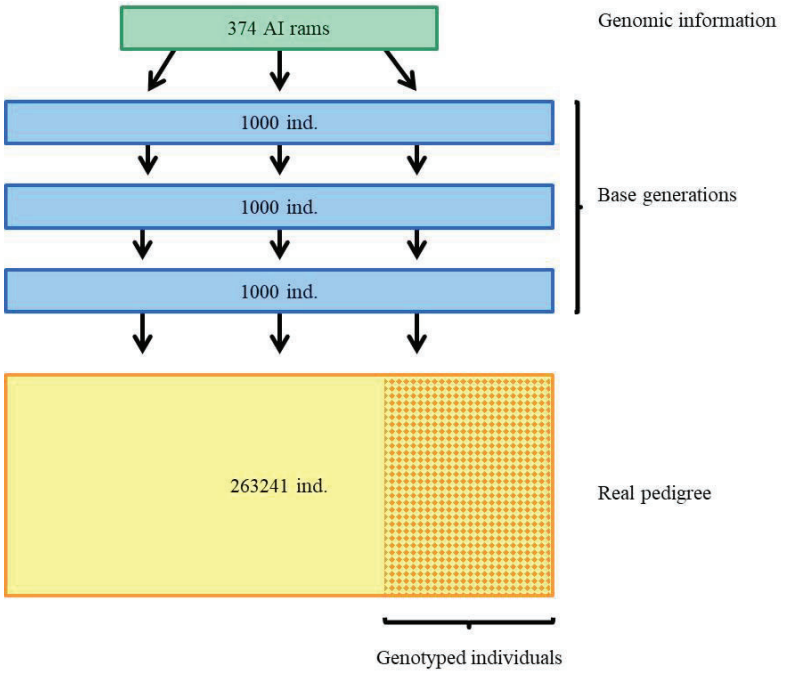

Figure 1: Structure of simulation strategy from genomic information to the simulation of genotypes for individuals (ind.) in Latxa Cara Negra from Euskadi pedigree.

For each individual, the true breeding value (**TBV**) was calculated as the sum of the effects of their genotypes for each QTL. The value was rescaled by its additive standard deviation. The assumed heritability was 0.1847 ± 0.0031 for LCNEUS, 0.2203 ± 0.0042 for LCR, and 0.2225 ± 0.0061 for LCNNAF (Granado-Tajada et al., 2020).

Phenotypic records were simulated for the individuals who already had a phenotype recorded in the real data set. Following the strategy of Mouresan et al. (2018), the simulated data reproduced the actual distribution of records across fixed and random (additive and permanent) effects. The phenotypic values of individuals were simulated by adding together the phenotypic mean, TBV, and an error term drawn from a Gaussian distribution with mean zero and variance one and rescaled by the residual standard deviation.

3.3.3. Simulated Genotyped Population Scenarios

To evaluate the described simulation process and to assess how close to reality the obtained results would be, a control scenario simulated the genotypes of the same individuals included in Granado-Tajada et al. (2020), keeping the same validation group.

For the simulated scenarios to test genotyping strategies, several variables were applied to select the individuals of the genotyped population, whereas the validation group was genotyped in all the scenarios (as described below). The selection variables considered when designing the scenarios were the sex of individuals, the EBV, pedigree knowledge, the starting year of genotyping, and the number of genotyped individuals. The selection variables used are detailed in Table 2. Each variable was defined to understand how different constitutions of the genotyped population contribute to prediction accuracy:

- Sex. Due to limited economic resources for genotyping and the higher selection pressure on males, the standard approach is to genotype males (**M**). Genotyped populations comprised of females (**F**) or both sexes (**M+F**) were considered to analyze the benefits that each sex or the combination could bring. The males selected for genotyping were not required to have daughters with phenotypic data, because in some cases, paternity is not recognized or recorded in pedigree. However, for selected females, it was compulsory to have their own phenotypic record. Selection pressure among females is low (75 % of females born from AI remain in the flock), and almost all the ewes have recorded lactations.
- EBV. The genotyped individuals were selected based on their EBV obtained from a standard BLUP evaluation. The males were always selected by best EBV. For females, 2 variables were tested based on EBV: best ewes (**B**) or extreme ewes (**E**). To select extreme individuals, best and worst females per year were combined in equal proportions. Because females were required to have their own records, as they are

the ones selected to remain in the flock, the selected worst females are not necessarily the worst born females.

- Pedigree knowledge. Genotyping individuals with and without known pedigree (indifferent, **I**), and genotyping restricted to individuals with complete pedigree (**P**) were tested to analyze effects of the connection with the overall population.
- Starting year. Although the breeding programs started in 1984, the start of unrestricted genotyping was fixed at 1987, when inseminations began. This variable allowed having a high number of genotyped animals from different generations, because the small population size could limit the potential benefits of genomic evaluations, and it meant to provide insight about future chances of the breeding program. A start restricted to 2010 (**R**) aimed to reflect the reality of the breeding programs (systematic genotyping started at 2010), in which the gain in prediction accuracy of genomic EBV (**GEBV**) could be reduced by the limited number of genotyped animals.
- Number of genotypes. To simulate scenarios close to reality, the limitations of the improvement programs were considered to set a maximum number of genotyped animals per year (**Max**) of 300 for LCNEUS and LCR breeds and 150 for LCNNAF for each program (CONFELAC Latxa Breeders' Association, personal communication). Moreover, the effect of genotyping fewer males per year was analyzed by testing different percentages of animals below the defined maximum (around 10, 20, 30, 50, and 70 %).

Table 2: Selection variables to makeup the genotyped population.

Selection variable	Acronym	Description	Acronym	Description
Sex	M	Males	F	Females
EBV	B	Best EBV	E	Extreme EBV
Pedigree knowledge	I	Indifferent	P	Known[1]
Starting year	—	Unrestricted	R	Restricted
Number of genotypes	Max	Maximum	%	Percentages

[1]Both parents are known.

By combining these selection variables, several genotyped population scenarios were designed and are detailed in Table 3. The main strategies followed were maximum genotyping and genotyping by percentages, and all the scenarios designed by means of these strategies were also simulated with restricted start of genotyping to 2010 (scenario acronym + **R**).

Table 3: Scenarios designed by the combination of selection variables to makeup the genotyped population, classified into maximum genotyping and genotyping by percentages main strategies.

Scenario[1]	Sex					
	Males[2]			Females		
	EBV	Pedigree knowledge	Percentage	EBV	Pedigree knowledge	Percentage
Maximum Genotyping						
MIMax	Best	Indifferent	Maximum			
MP	Best	Known				
FBIMax				Best	Indifferent	Maximum
FEIMax				Extreme	Indifferent	Maximum
FBP				Best	Known	
FEP				Extreme	Known	
MP+FBI	Best	Known		Best	Indifferent	
MP+FEI	Best	Known		Extreme	Indifferent	
MP+FBP	Best	Known		Best	Known	
MP+FEP	Best	Known		Extreme	Known	
Genotyping by Percentages						
MI	Best	Indifferent	10			
MI	Best	Indifferent	20			
MI	Best	Indifferent	30			
MI	Best	Indifferent	50			
MI	Best	Indifferent	70			
MI+FBI	Best	Indifferent	10	Best	Indifferent	90
MI+FBI	Best	Indifferent	20	Best	Indifferent	80
MI+FBI	Best	Indifferent	30	Best	Indifferent	70
MI+FBI	Best	Indifferent	50	Best	Indifferent	50
MI+FBI	Best	Indifferent	70	Best	Indifferent	30
MI+FEI	Best	Indifferent	10	Extreme	Indifferent	90
MI+FEI	Best	Indifferent	20	Extreme	Indifferent	80
MI+FEI	Best	Indifferent	30	Extreme	Indifferent	70
MI+FEI	Best	Indifferent	50	Extreme	Indifferent	50
MI+FEI	Best	Indifferent	70	Extreme	Indifferent	30

[1]M: male; F: female; I: indifferent pedigree; P: known pedigree; B: genetically best individuals; E: combination of genetically extreme animals; Max: maximum genotyping per year; restricted scenarios follow the same strategy, but the starting year of genotyping was 2010.

[2]Males born in the last year (2017) are genotyped in all scenarios.

3.3.4. Estimated Breeding Values and Validation

The phenotypic and genotypic data obtained were used to calculate EBV, by pedigree-based BLUP; and GEBV, using single-step genomic BLUP (**ssGBLUP**; Aguilar et al., 2010; Christensen and Lund, 2010). In both cases the model for milk yield was as follows:

$$Y_{ijklm} = (FYS)_i + A_j + I_k + L_l + u_m + p_m + e_{ijklm}$$

where flock-year-season (***FYS***), age-parity number (***A***), interval from lambing to first milk recording (***I***), and number of lambs born alive (***L***) were included as fixed effects (Legarra et al., 2005), with 19062, 11, 8 and 3 levels for LCNEUS; 10908, 10, 8 and 3 levels for LCR; and 4800, 9, 8 and 3 levels for LCNNAF, respectively. In addition, additive genetic (***u***) and individual random environmental or permanent (***p***) were considered as random effects. Both analyses were performed using the BLUPf90 software suite (Misztal et al., 2002) with the default quality control values.

The different genotyping strategies were compared in terms of relative difference in accuracy over pedigree evaluations, estimated as the Pearson correlation between TBV and estimated breeding values (EBV and GEBV) of the validation group, made up of the males born in the previous year (2017). These validation individuals were considered the actual candidates to selection; they do not have daughters with records and were genotyped in all the scenarios (298 for LCNEUS; 287 for LCR; 126 for LCNNAF). Each one of the genotyping scenarios described was replicated 20 times, and we present the mean and standard error of the relative accuracy difference.

3.4. Results and Discussion

In a preliminary approach, a control scenario was carried out to mimic the study done with real data (Granado-Tajada et al., 2020). In the control scenario, the accuracy difference between genomic and pedigree-based evaluations was 0.0113 ± 0.0030 for LCNEUS and 0.0017 ± 0.0015 for LCR. The real data study described accuracy differences of 0.0097 ± 0.0423 for LCNEUS and 0.0064 ± 0.0387 for LCR. The LCNNAF was not included in the study mentioned, due to the limited number of genotyped individuals. The results are not directly comparable because the current work compared predicted breeding values with TBV, and the previous study was a cross validation in which the metafounder theory (Legarra et al., 2015) was used; however, similar results were obtained, reinforcing the validity of the simulation approach.

The definition of the genotyped population is a determining factor for prediction accuracy under genomic evaluation. To analyze different ways to select animals for the genotyped population, various combinations of sex of individuals, EBV, pedigree knowledge, starting year of genotyping, and number of genotyped individuals were tested. In all cases, breeding values estimated by ssGBLUP showed higher accuracies (mean 0.5841 for LCNEUS, 0.5981 for LCR, 0.5507 for LCNNAF) than pedigree-based BLUP evaluations (mean 0.4570 for LCNEUS, 0.5190 for LCR, 0.5111 for LCNNAF), so the mean relative accuracy differences were always positive.

3.4.1. Maximum Genotyping

Regarding the scenarios designed by genotyping the maximum number of individuals, results of the relative accuracy difference of genomic evaluations in contrast to pedigree-based evaluations are summarized in Figure 2 and Supplemental Tables S2–S7 (Annexe IV). Broadly speaking, the genotyped population comprised of only females provided the highest relative accuracies. These higher accuracies of genomic predictions were more relevant in unrestricted data sets (Figure 2, left). In restricted data sets (Figure 2, right), genotyped population of only females brought similar accuracies to genotyped populations with both sexes. When only males made up the genotyped population, lower relative accuracies were shown in both unrestricted and restricted data sets. Genotyping exclusively individuals with known pedigree, with the consequent reduction in number, showed a noticeable reduction in accuracy in female-only unrestricted scenarios. These results highlighted that when there is a considerable proportion of unknown pedigree, as in our case (around 45 % for males and 65 % for females), the increase in accuracy of ssGBLUP is strengthened by the incorporation of genotyped females with their own records, rather than genotyping unconnected males.

Regarding unrestricted scenarios (Figure 2, left), when only females were genotyped, the effect of selecting the best ewes or extreme ewes by EBV with indifferent pedigree (FBIMax and FEIMax) was tested. Both alternatives showed quite similar relative accuracy differences (e.g., for LCNEUS 0.5369 and 0.5492, respectively). When only females with complete pedigree were genotyped (FBP and FEP), the relative accuracy difference diminished (0.3657 and 0.3789, respectively, for LCNEUS), even though the number of genotyped females per year was much lower (around 50 % less), and the mean number of lactations per genotyped ewe kept constant around 3. Similar trends were shown for LCR and LCNNAF breeds, with smaller relative accuracy differences (e.g., FBIMax and FBP were 0.3322 and 0.2375 for LCR, and 0.1252 and 0.1447 for LCNNAF, respectively). There is a loss in accuracy of GEBV when the number of genotyped females decreases, which is less noticeable according to genotyped

population size. This is probably due to the fact that although these individuals do not have complete pedigree, their own records, and possibly the data and connections of their offspring, provide useful information for more accurate evaluations.

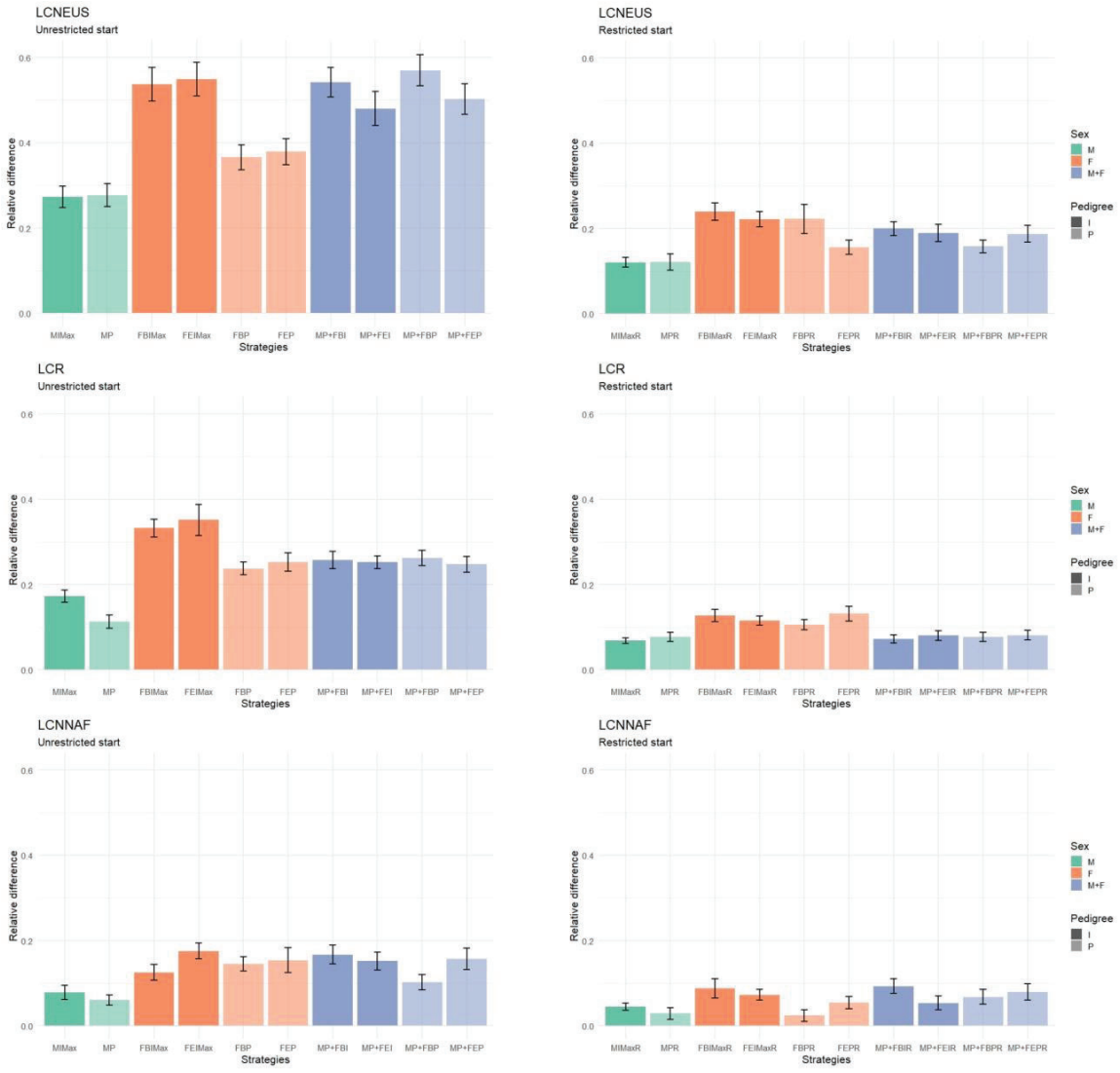

Figure 2: Relative accuracy difference of genomic evaluations in contrast to pedigree-based evaluations, comparing unrestricted (left) and restricted (right) start of genotyping into different genotyping strategies for the three breeds. M: males; F: females; M+F: complementation of males with complete pedigree with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; P: only individuals with complete pedigree; Max: maximum genotyping per year; R: restricted start of genotyping; for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR), and Latxa Cara Negra from Navarre (LCNNAF). Bars indicate standard error.

Similar results found between scenarios genotyping best or extreme females could be due to some degree of selection, because genotyped females were required to have their own records. Therefore, these individuals have already been selected to remain at the flock. The applicability of genotyped populations formed exclusively by females has been investigated in the literature, and some studies describe better accuracies than pedigree index (Ding et al., 2013) or genotyped population of males (Jiménez-Montero et al., 2012).

Genotyped populations of only males have been the standard approach since the beginning of GS (Van-Raden, 2020) due to limited economic resources, because the selection pressure exerted on this sex is higher, and the diffusion of the genetics of rams achieved by AI is larger. When all the efforts focused on genotyping the best males (indifferent pedigree) per year up to the stabilized maximum number of individuals (MIMax) in the unrestricted scenarios (Figure 2, left), a relative accuracy difference of 0.2729 was achieved for LCNEUS. The genotyped males in this breed had a mean number of daughters in data of 3.3. Selecting among those best males the ones with complete pedigree (MP) reduced the number of genotyped males to around 130 animals per year in LCNEUS, but the gain in accuracy was kept at 0.2769, and the mean number of daughters in data increased to 6.7. Similar trends were shown for LCR and LCNNAF breeds, with smaller relative accuracy differences (MIMax and MP were 0.1729 and 0.1131, respectively, for LCR; and 0.0777 and 0.0607 for LCNNAF). By contrast to female genotyping, the relative accuracy difference decrease was slight, whereas the number of genotyped rams was notably reduced (even 50 %), and the average number of daughters in data was increased (around 150 %), which reflects the importance of genotyping animals well connected in the population. These results confirmed that the genotyping of rams in flocks have a limited effect on prediction accuracy if they are isolated in the pedigree and without offspring records (Pszczola et al., 2012; Shabalina et al., 2017; de Oliveira et al., 2019).

Regarding the scenarios composed of both sexes (males with complete pedigree combined with females) in unrestricted data set (Figure 2, left), intermediate results between only female and only male genotyping were found. As in the case of only female genotyped scenarios, a similar relative accuracy difference was found regardless of the strategy used to select the genotyped females. Moreover, females with best EBV (MP+FBI) or with extreme EBV (MP+FEI) showed almost the same relative accuracy difference as females with complete pedigree (MP+FBP and MP+FEP). Mean relative accuracy difference of these scenarios was 0.5234 for LCNEUS, 0.2549 for LCR, and 0.1445 for LCNNAF.

The inclusion of females in the genotyped population was also found to be beneficial in species where the effect of female paths on genetic progress is strong, and the effect of genomic data on accuracy depends on the existing population structure (Lourenco et al., 2015). The effect of genotyped populations including female genomic information on genomic prediction accuracy has been thoroughly studied (Koivula et al., 2016; Uemoto et al., 2017; Perez et al., 2019), and also as a strategy to increase the genotyped population size into small populations (Jiménez-Montero et al., 2012; Lund et al., 2016; Jenko et al., 2017). Different strategies have been proposed to assess an optimal selection of individuals to be genotyped. Including genomic information of females selected based on a divergent strategy seems to be better than a random or directional approach (Jenko et al., 2017; Perez et al., 2019), and avoiding the inclusion of only selected individuals would reduce prediction bias (Vitezica et al., 2011; Koivula et al., 2016). In practical terms not all the strategies are feasible for animal breeding programs and even less for small populations. Including the genotypes of high yield females by proportional sampling within detected communities could help to obtain more value from genotypes and phenotypes (Perez et al., 2019).

As the simulation was based on real data, it should be considered that the results obtained are likely to be conditioned by the individuals recorded into pedigree. Only animals that have been selected to remain in the flock are recorded, and there is no information about the ones removed. This is especially relevant for males, because the high selection pressure results in only a few selected individuals recorded in the pedigree. Genotyping males before selection could bring different results, as other studies have shown (Lourenco et al., 2015).

Finally, maximum genotyping scenarios were also simulated under restricted genotyping starting year (scenario + R) and relative accuracy difference results are also summarized in Figure 2 (right). Starting the genotyping in 2010 caused a drop in relative accuracy difference in all the scenarios, and trends were not as clear as in unrestricted scenarios. When the maximum number of males per year was genotyped (MIMaxR), the relative accuracy difference for LCNEUS was 0.1206, genotyping only best females (FBIMaxR) was 0.2393, and genotyping extreme females was 0.2217. Compared with unrestricted scenarios, this reduction in relative accuracy difference was probably due to the reduction of the genotyped population size by almost 60 %, removing relatives of past generations. When only the animals with complete pedigrees were considered, the relative accuracy difference was also reduced, although in a smaller proportion than in unrestricted scenarios. When males with complete

pedigree were combined with females, the relative accuracy difference dropped by 60 % compared with unrestricted scenarios regardless of the applied strategy to select females.

At this moment, the restricted scenarios are closer to the reality of Latxa breeding programs than unrestricted ones. So, the expected gain in accuracy in the short term for Latxa populations would be around 11 %. The lower results showed by restricted scenarios (60 % less accuracy difference than unrestricted scenarios), highlighted the relevance of having genomic data from various previous generations.

Clear examples of recent implementation of GS with gains in accuracy over pedigree-based estimates are French dairy sheep breeds. Lacaune and Manech breeds have genotyped rams since the 1990s, and these breeding programs implemented a genomic scheme in 2015 and 2017, respectively (Baloche et al., 2014; Legarra et al., 2014). By that date, each program had genotyped more than 1,900 and 1,000 individuals, respectively; a genotyped population size over the minimum of 1,000 individuals set by Shumbusho et al. (2013) to achieve benefits from GS over classical selection in dairy sheep. The inclusion of molecular information was described to increase prediction accuracies of milk yield evaluations 0.15 for Lacaune, and 0.16 for Manech Tête Rousse (Baloche et al., 2014; Legarra et al., 2014). These gains in accuracy were similar to the found in Latxa populations when the genotyped of between 1,000 and 2,000 individuals were simulated.

Before implementing GS, the balance between economic cost and benefit should be considered, more complex to reach equilibrium for small ruminant breeding programs. However, simulation studies done for dairy and meat sheep concluded that an optimum balance exists by implementing some strategies to take advantage from a GS scheme (Buisson et al., 2014; Shumbusho et al., 2016). Regarding the Latxa breeding program, it would be interesting to analyze in detail specific costs and benefits of the genotyping strategies.

3.4.2. Genotyping by Percentages

Seeking a balance between prediction accuracy and economic costs, the effect of genotyping a lower number of individuals per year was analyzed by testing different percentages of animals below the maximum number of each breeding program (300 individuals for LCNEUS and LCR or 150 individuals for LCNNAF). Relative accuracy difference results of genomic evaluations in contrast to pedigree-based evaluations are summarized in Figure 3 and Supplemental Tables S8-S13 (Annexe IV). In unrestricted scenarios genotyping only males with indifferent pedigree (MI; Figure 3, left), as might be expected (Meuwissen et al., 2001), the relative accuracy

difference rose according to the number of genotyped individuals in the 3 breeds. An increment of 60 % was achieved from 10 to 70 % male genotyping scenarios. By contrast, by genotyping 70 % of males per year, similar relative accuracy difference to that achieved with 100 % of the maximum number of genotyped males (MIMax) was found. When genotyping the maximum number of males, the relative accuracy difference achievable was around 0.3049 for LCNEUS, 0.1157 for LCR, and 0.0834 for LCNNAF. From 30 % of genotyped males, the relative accuracy difference increase was less noticeable (except for LCNEUS), even though the size of the genotyped population kept increasing.

When those percentages of genotyped males (MI from 10 to 70 %) were combined with females (FBI or FEI) up to the maximum number of genotyped animals per year to make up different proportions of sexes (Figure 3, left), similar relative accuracy differences were found in all the scenarios (best or extreme EBV females). This genotyping scenario showed the highest mean relative accuracy differences, around 0.4999 for LCNEUS, 0.2848 for LCR, and 0.1259 for LCNNAF, which were similar to the results shown on Figure 2 for scenarios with both sexes into the genotyped population. Once again, the benefit of combining males and females into genotyped populations is reflected in prediction accuracy.

The genotyping of different percentages of animals and varying proportions of sexes were also simulated with a restricted start of genotyping, and relative accuracy difference results of genomic predictions comparing with pedigree-based ones are shown in Figure 3 (right). As in previous results, the relative accuracy difference was lower than for unrestricted scenarios, but always keeping the same trends and the combination of both sexes in the genotyped population showed the higher relative accuracy differences.

From an overall perspective of the simulated scenarios and given the fixed genotyping maximum, there appeared to be a limit to the relative accuracy difference achievable, regardless of the genotyping strategy applied, which for LCNEUS was at 0.56, for LCR at 0.33, and for LCNNAF at 0.18. The relative accuracy difference was different depending on the breed. The LCNEUS had the possibility to achieve the biggest benefits from the genomic information, with a mean relative accuracy difference for unrestricted scenarios of 0.4239, whereas LCR had a mean relative accuracy difference of 0.2324 and LCNNAF of 0.1153. These differences could be due to the population size and the amount and characteristics of the recorded information. LCNEUS breed had the biggest amount of data, but compared with LCR, there was no structural differences because the use of AI, knowledge of genealogy and the distribution of offspring and lactations were similar in both programs. Meanwhile LCNNAF was

the smaller population and the breeding program started later, so the gain in accuracy was limited. Despite the different mean relative accuracy differences, the three breeds behave similarly.

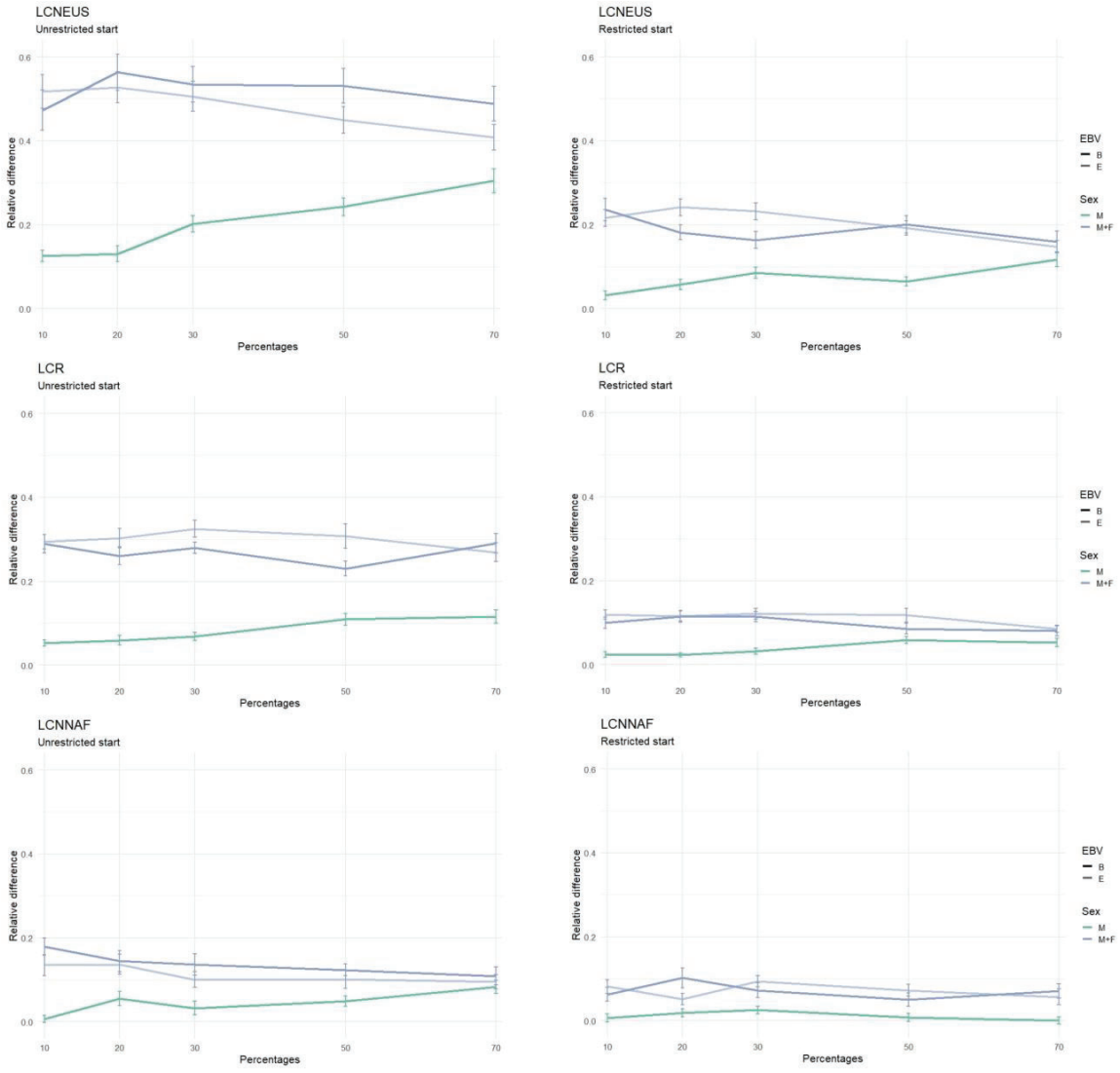

Figure 3: Effect of various combinations of genotyped individual percentages on the relative accuracy difference of genomic evaluations in contrast to pedigree-based evaluations, for unrestricted (left) and restricted (right) start of genotyping of different genotyping strategies. M: males; F: females; M+F: complementation of males with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; R: restricted start of genotyping; for Latxa Cara Negra from Euskadi (LCNEUS), Latxa Cara Rubia (LCR) and Latxa Cara Negra from Navarre (LCNNAF) breeds. Bars indicate standard error.

Restricted scenarios are closer to the current reality of the breeding program and, in spite of lower mean relative accuracy differences (LCNEUS 0.1684; LCR 0.0883; and LCNNAF 0.0560), genomic predictions were shown to be more accurate than pedigree-based predictions. These gains in prediction accuracy would be very interesting, given that the average accuracy is 0.55 for LCNEUS, 0.50 for LCR, and 0.60 for LCNNAF (Granado-Tajada et al., 2020). Moreover, previous studies did not find increases in prediction accuracy by the genotyping of AI rams (Legarra et al., 2014; Granado-Tajada et al., 2020). So, although accuracy gains yield by simulation is known to be bigger than that reported in real data studies (Legarra et al., 2008; de Roos et al., 2009), the results showed a potential benefit of implementing a GS scheme in Latxa breeds, and go in depth into the knowledge of GS applied to small populations.

3.5. Conclusion

The studied genotyped population scenarios showed that breeding values estimated by ssGBLUP were more accurate than pedigree-based BLUP evaluations, reflecting a potential benefit in terms of prediction accuracy of implementing a GS scheme in Latxa breeding programs. Genotyping males was found to be beneficial, when they are not isolated individuals and they have daughters with phenotypic records. Genotyping females with their own lactation records was shown to increase prediction accuracy, regardless of their connection level. No differences were found between the genotyping of genetically best or extreme individuals, probably due to the low selection pressure on females. The combined genotyping of males and females provided intermediate accuracies regardless of the female selection strategy. The importance of genotyping individuals from various generations has been highlighted, although without historical individuals it is also possible to take advantage of accurate genomic predictions. In the same way, larger genotyped population sizes led to greater relative accuracy differences, but by only genotyping males, a maximum relative difference was achievable. Therefore, genotyping rams is of interest, but at some point, it would be more beneficial to focus the efforts of the breeding program on combining males with females. Regarding Latxa breeding programs, the implementation of GS seems to be beneficial and the implementation of a genotyped population of both sexes would be worthwhile.

3.6. References

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93(2):743–752.

Auvray, B., J. C. McEwan, S.-A. N. Newman, M. Lee, and K. G. Dodds. 2014. Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. J. Anim. Sci. 92(10):4375–4389.

Baloche, G., A. Legarra, G. Sallé, H. Larroque, J. M. Astruc, C. Robert-Granié, and F. Barillet. 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. J. Dairy Sci. 97(2):1107–1116.

Boichard, D., V. Ducrocq, P. Croiseau, and S. Fritz. 2016. Genomic selection in domestic animals: Principles, applications and perspectives. C. R. Biol. 339(7-8):274–277.

Buisson, D., G. Lagriffoul, G. Baloche, X. Aguerre, P. Boulenc, F. Fidele, G. Frégeat, B. Giral-Viala, P. Guibert, P. Panis, C. Soulas, J. M. Astruc, and F. Barillet. 2014. Genomic breeding schemes in French Lacaune and Manech dairy sheep: Design and expected genetic gain. In proceedings of the 10th World Congress of Genetics Applied to Livestock Production. 17–22 August. Vancouver, Canada.

Cesarani, A., G. Gaspa, F. Correddu, M. Cellesi, C. Dimauro, and N. P. P. Macciotta. 2019a. Genomic selection of milk fatty acid composition in Sarda dairy sheep: Effect of different phenotypes and relationship matrices on heritability and breeding value accuracy. J. Dairy Sci. 102(4):3189–3203.

Cesarani, A., I. Pocrnic, N. P. P. Macciotta, B. O. Fragomeni, I. Misztal, and D. A. L. Lourenco. 2019b. Bias in heritability estimates from genomic restricted maximum likelihood methods under different genotyping strategies. J. Anim. Breed. Genet. 136(1):40–50.

Christensen, O. F., and M. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42(3):2.

Clark, S. A., J. M. Hickey, H. D. Daetwyler, and J. H. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genet. Sel. Evol. 44:4.

Daetwyler, H., A. Swan, J. van der Werf, and B. Hayes. 2012. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet. Sel. Evol. 44(1):33.

de Oliveira H. R., L. F. Brito, M. Sargolzaei, F. F. e Silva, J. Jamrozik, D. A. L. Lourenco, and F. S. Schenkel. 2019. Impact of including information from bulls and their daughters in the training population of multiple-step genomic evaluations in dairy cattle: A simulation study. J. Anim. Breed. Genet. 136(6):441–452.

de Roos, A. P. W.,B. J. Hayes, and M. E. Goddard. 2009. Reliability of genomic predictions across multiple populations. Genetics. 183(4):1545–1553

Ding, X., Z. Zhang, X. Li, S. Wang, X. Wu, D. Sun,Y. Yu, J. Liu, Y. Wang, Y. Zhang, S. Zhang, Y. Zhang, and Q. Zhang. 2013. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. J. Dairy Sci. 96(8):5315–5323.

Gao, H., P. Madsen, U. S. Nielsen, G. P. Aamand, G. Su, K. Byskov, and J. Jensen. 2015. Including different groups of genotyped females for genomic prediction in a Nordic Jersey population. J. Dairy Sci. 98(12):9051–9059.

Granado-Tajada, I., A. Legarra, and E. Ugarte. 2020. Exploring the inclusion of genomic information and metafounders in Latxa dairy sheep genetic evaluations. J. Dairy Sci. 103(7):6346–6353.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5.

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. 91:47–60.

Ibañez-Escriche, N., and O. González-Recio. 2011. Review. Promises, pitfalls and challenges of genomic selection in breeding programs. Span J. Agric. Res. 9(2):404–413.

Ibañez-Escriche, N., and H. Simianer. 2016. From the editors: Animal breeding in the genomics era. Anim. Front. 6(1):4–5.

Jenko, J., G. R. Wiggans, T. A. Cooper, S. A. E. Eaglen, W. G. L. Luff, M. Bichard, R. Pong-Wong, and J. A. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. J. Dairy Sci. 100(1):439–452.

Jiménez-Montero, J. A., O. González-Recio, and R. Alenda. 2012. Genotyping strategies for genomic selection in small dairy cattle populations. Animal. 6(8):1216–1224.

Kijas, J. W., L. Porto-Neto, S. Dominik, A. Reverter, R. Bunch, R. McCulloch, B. J. Hayes, R. Brauning, and J. McEwan. 2014. Linkage disequilibrium over short physical distances measured insheep using a high-density SNP chip. Anim. Genet.45:754–757.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal. 10(6):1061–1066.

Legarra, A., P. López-Romero, and E. Ugarte. 2005. Bayesian model selection of contemporary groups for BLUP genetic evaluation in Latxa dairy sheep. Livest. Prod. Sci. 93:205–212.

Legarra, A., C. Robert-Granié, E. Manfredi, and J. M. Elsen. 2008. Performance of genomic selection in mice. Genetics. 180(1):611–618.

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(5):3200–3212.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. Genetics. 200:455–468.

Lourenco, D. A. L., B. O. Fragomeni, S. Tsuruta, I. Aguilar, B. Zumbach, R. J. Hawken, A. Legarra, and I. Misztal. 2015. Accuracy of estimated breeding values with genomic information on males, females, or both: An example in broiler chicken. Genet. Sel. Evol. 47(1):56.

Lund, M. S., I. van den Berg, P. Ma, R. F. Brøndum, and G. Su. 2016. How to improve genomic predictions in small dairy cattle populations. Animal. 10(6):1042–1049.

MacCluer, J. W., J. L. Vandeburg, B. Read, and O. A. Ryder. 1986. Pedigree analysis by computer simulation. Zoo Biology. 5(2):149–160.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819–1829.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). In proceedings of the 7th World Congress on Genetics Applied to Livestock Production. 19-23 August 2002. Montpellier. France.

Mouresan, E. F., J. Altarriba, C. Moreno, S. Munilla, A. González-Rodríguez, and L. Varona. 2017. Performance of genomic selection under a single-step approach in autochthonous Spanish beef cattle populations. J. Anim. Breed. Genet. 134(4):289–299.

Mouresan, E. F., J. J. Cañas-Álvarez, A. González-Rodríguez, S. Munilla, J. Altarriba, C. Díaz, J. A. Baró, A. Molina, J. Piedrafita, and L. Varona. 2018. Evaluation of the potential use of a meta-population for genomic selection in autochthonous beef cattle populations. Animal. 12(7):1350–1357.

Perez, B. C., J. C. C. Balieiro, R. Carvalheiro, F. Tirelo, G. A. Oliveira Junior, J. M. Dementshuk, J. P. Eler, J. B. S. Ferraz, and R. V. Ventura. 2019. Accounting for population structure in selective cow genotyping strategies. J. Anim. Breed. Genet. 136(1):23–39.

Pszczola, M., T. Strabel, J. A. M. van Arendonk, and M. P. L. Calus. 2012. The impact of genotyping different groups of animals on accuracy when moving from traditional to genomic selection. J. Dairy Sci. 95(9):5412–5421.

Rupp, R., S. Mucha, H. Larroque, J. McEwan, and J. Conington. 2016. Genomic application in sheep and goat breeding. Anim. Front. 6(1):39–44.

Sargolzaei, M., P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 15:478.

Schöpke, K., and H. H. Swalve. 2016. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal. 10(6):1050–1060.

Shabalina, T., E. C. G. Pimentel, C. Edel, L. Plieschke, R. Emmerling, and K.-U. Götz. 2017. Short communication: The role of genotypes from animals without phenotypes in single-step genomic evaluations. J. Dairy Sci. 100(10):8277–8481.

Shumbusho, F., J. Raoul, J. M. Astruc, I. Palhiere, and J. M. Elsen. 2013. Potential benefits of genomic selection on genetic gain of small ruminant breeding programs. J. Anim. Sci. 91(8):3644–3657.

Shumbusho, F., J. Raoul, J. M. Astruc, I. Palhiere, S. Lemarié, A. Fugeray-Scarbel, and J. M. Elsen. 2016. Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. Animal. 10(6):1033–1041.

Uemoto, Y., T. Osawa, and J. Saburi. 2017. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. Animal. 11(3):382–393.

Usai, M. G., S. Salaris, S. Casu, T. Sechi, S. Miari, P. Carta, and A. Carta. 2018. Feasibility of genomic predictions of Sarda breed rams using a female reference population. In proceedings of the 11th World Congress on Genetics Applied to Livestock Production. 11–16 February. Auckland, New Zealand.

van den Berg, I., T. H. E. Meuwissen, I. M. MacLeod, and M. E. Goddard. 2019. Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. J. Dairy Sci. 102(4):3155–3174.

VanRaden P. M., C. P. Van Tassell, G. R. Wiggans, T. S. G. Sontegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92(1):16–24.

VanRaden, P. M. 2020. Symposium review: How to implement genomic selection. J. Dairy Sci. 103(6):5291–5301.

Vitezica, Z. G., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for populations under selection. Genet. Res. (Camb.) 93(5):357–366.

GENERAL DISCUSSION

GENERAL DISCUSSION

With the development of molecular techniques, evaluation methodologies and bioinformatics, the implementation of genomic selection (**GS**) in breeding programs has been spreading, although at variable rate depending on the species and breed (Ibañez-Escriche and Gonzalez-Recio, 2011; Jonas and de Koning, 2015; Meuwissen et al., 2016). Dairy cattle breeding programs were the first and most successful implementing GS (Hayes et al., 2009a; Van Raden, 2020), while its implementation in dairy sheep breeds is not straightforward and the specific requirements should be considered, such as population size, use of artificial insemination (**AI**) and the linkage disequilibrium (**LD**) pattern (Jonas and de Koning, 2015; Mrode et al., 2018). The work presented in this thesis manuscript was developed after some exploratory results about the prediction accuracy of genomic evaluations in Latxa dairy sheep breeds, which were not conclusive (Legarra et al., 2014). Therefore, an in-depth analysis of the effect of including genomic information in genetic evaluations with more data and better tools was needed, together with the study of the influence of making up the reference genotyped population with different sets of individuals and understanding the current population structure after years of selection and how this could affect the prediction accuracy.

The availability of genomic information allows one to obtain more precise genetic diversity estimates of populations with heterogeneous pedigrees, which is the case of Latxa breed, where estimates based on pedigree are likely to be underestimated. The first study undertaken was the estimation of inbreeding and coancestry of Latxa populations based on pedigree and genomic information (Chapter 1). Despite the decades of selection done by the Latxa breeding program, the estimated moderate inbreeding increase (a rate in inbreeding per generation between 0.0050 and 0.0080) shows that there has been a suitable management to control inbreeding. It was mainly done by mating control and selection of lambs for AI, even though decisions were taken considering pedigree-based inbreeding estimates. Differences between inbreeding, coancestry and subsequent estimated effective population sizes ($\mathbf{N_e}$) based on different information sources and methodologies were expected (Rodríguez-Ramilo et al., 2019). However, using pedigree or genomic information in Latxa breeds results were consistent and showed that black populations (LCNEUS and LCNNAF) have higher inbreeding increases and smaller $\mathrm{N_e}$, whereas in LCR population the importation of genetic material from the French Manech Tête Rousse (MTR) has reduced the inbreeding increase, resulting in bigger $\mathrm{N_e}$ and an increase in coancestry between both blonde populations across borders. The inbreeding and $\mathrm{N_e}$ estimates for Latxa populations obtained showed enough genetic diversity

to ensure long term viability (Meuwissen, 2009), and are in agreement with the estimates of other dairy sheep breeds with similar characteristics (Calvo et al., 2006; Chitneedi et al., 2017; Rodríguez-Ramilo et al., 2019). This study showed the usefulness of genomic information to obtain more precise genetic diversity estimates of populations with heterogeneous pedigrees, such as in the Latxa breed, where estimates based on pedigree are likely to be underestimated.

N_e is known to evolve, affecting the LD pattern between specific alleles of SNPs and quantitative trait loci (**QTL**): when N_e is low, the kinship level among individuals is high and therefore so is the LD (Falconer and Mackay, 1996). The response to GS relies on the fact that markers are in LD with the causal variant (Meuwissen et al., 2001), the stronger the LD, the higher the reliability of genomic predictions (Calus et al., 2008; Solberg et al., 2008). Based on this assumption, the higher prediction accuracies found in cross-validation studies for LCNEUS and LCNNAF (Chapter 2) are expectable because these populations have the lower N_e. However, the comparison of BLUP and ssGBLUP methodologies based on data available until 2017 (section 2.1), showed that the inclusion of genotypes did not increase or decrease the accuracy of predictors for any of the breeds (accuracy difference was 0.01 ± 0.04 in both breeds), independently of their N_e.

Comparing the results from section 2.1 with the previous one obtained by (Legarra et al., 2014), the predictions were more consistent, indicating that breeding values are more correctly estimated. Biases were lower and with considerably smaller standard errors, which were remarkably high in the previous study; and the slopes of the regression for both BLUP and ssGBLUP methodologies were closer to one than previous ones. Accuracy of estimations was positive in all of the cases, whereas negative values were reported by Legarra et al (2014). These were higher than the previous estimates and more coherent with the results of the breeding program. A possible explanation of these better results was associated with a stronger data structure, more genotyped individuals and the implementation of better methodological tools, such as metafounders or validation by the LR method.

The cross-validation study was updated including data until 2019 in section 2.2, and the same methodology of the section 2.1 was applied, except for the use of genetic groups instead of metafounders to model missing pedigree. The new data comprised genealogical, phenotypic and genomic information, where the composition and size of the latter was especially relevant. Natural service rams and females were included and the number of genotyped individuals was 2.4 times bigger. This fact was especially relevant for LCNNAF population, because the

considerable increase in the number of genotyped made a consistent cross-validation study possible.

Prediction accuracies of the updated results (section 2.2) were also positive in all cases, and the comparison of BLUP and ssGBLUP showed that the inclusion of genotypes available until 2019 increased the accuracy of genomic predictions for the three breeds. The populations with smaller N_e, the black ones, showed higher accuracies (around 0.60 vs. 0.50 for LCR), however the bigger differences between genetic and genomic prediction accuracies were found in LCNEUS and LCR populations (14 %). The genotyped population sizes of LCNEUS and LCR were near to 1,000 individuals, which was close to the minimum recommended by Shumusho et al. (2013) of precisely 1,000 genotyped individuals to achieve benefits from GS over classical selection in dairy sheep. For LCNNAF, genomic information also represented a beneficial source of information, but possibly the smaller genotyped population size (near 500) was a limitation to having higher gains in prediction accuracy (restricted to a 5 %).

These better accuracies of genomic estimates may be due to the increase in the genotyped population size, as well as to the type of genotyped animals (Schöpke and Swalve, 2016; van den Berg et al., 2019). Genotyping natural service rams and dams of AI males, together with the genotyping of the new males used in AI, could have been the point to substantially increase the accuracy of predicted breeding values due to the better estimation of the realized relationship matrix. In the same sense, the prediction accuracy of the genotyped individuals with no phenotype of their own increased (Hayes et al., 2009b; Habier et al., 2010), and the pedigree relationships for non-genotyped animals are enhanced by the genomic information of their relatives (Lourenco et al., 2020).

Given the relevance of the choice of individuals to form the genotyped population, and in agreement with the described in the literature, a simulation study was performed based on the real data structure of the Latxa populations (Chapter 3) to have a deeper insight. Optimizing the genotyped population is especially important for the implementation of GS in small ruminant breeding programs due to the usually limited economic resources for genotyping. In our results, genotyping males was found to be beneficial (prediction accuracy increased between 6 and 27 %) if they are not isolated individuals and they have daughters with phenotypic records. When the genotyped populations were composed only of females or included males and females, an increase in prediction accuracy was observed (between 10 and 50 %) provided that females have their own lactation records. Moreover, increasing the

genotyped population size showed bigger prediction accuracies, while different sex proportions did not bring differences in accuracy.

The simulation study made it possible to imagine future opportunities to obtain bigger prediction accuracies in the Latxa breeding program with scenarios with large genotyped population sizes by unrestricted start of genotyping. But also restricted scenarios were simulated to reflect the reality of the breeding program and to analyse genotyping strategies closer to the current situation. Besides the simulated control scenario, which mimicked the genotypes of the same individuals as those included in the cross-validation of section 2.1, other male genotyping scenarios brought similarly low prediction accuracies, while sex combined restricted scenarios were close to the data analysed in the updated cross-validation (section 2.2) and both resulted on gains better prediction accuracies. Therefore, although they are not directly comparable due to substantial genotyped population size differences, including females in the genotyped population was found to increase prediction accuracy both by real data and by simulation studies.

The benefits in genomic prediction accuracy when the genotyped population includes female genomic information has been thoroughly studied (Koivula et al., 2016; Uemoto et al., 2017; Perez et al., 2019), and also as a strategy to increase the genotyped population size (Jimenez-Montero et al., 2012; Lund et al., 2016; Jenko et al., 2017). Therefore, genotyping females could be interesting when the benefit from GS seems to be limited, as is the case of sheep breeds, and in general of small populations. The Italian Sarda breed chose the females' genotyping strategy and genomic predictions of rams for milk yield reached 0.13 more accuracy based on a female reference population of approximately 3,700 ewes (Usai et al., 2018), whereas the reference genotyped population of French dairy sheep breeds was made up of AI rams, due to an extensive use of AI and the consequent better pedigree knowledge. Higher gains in accuracy for using genomic information in milk yield evaluations were described for Lacaune and MTR populations (0.15 and 0.16, respectively), whose genotyped population sizes were of 1,900 and 1,000 individuals (Baloche et al., 2014; Legarra et al., 2014).

Regarding the context of the Latxa breeding program and based on the results obtained, the main conclusion of this study is that the implementation of GS seems to be viable. Current inbreeding levels are moderate and the estimated N_e show enough genetic diversity. Moreover, the genetic evaluations done with genomic information are slightly biased, with small dispersion of EBVs and the prediction accuracies are higher than pedigree based evaluations. Finally, including genotyped females in the reference population increases the

accuracy of genomic predictions, suggesting that the implementation of a genotyped population composed of both sexes would be worthwhile. Therefore, the Latxa breeding program has started to move towards a GS scheme, which after some organizational modifications is going to be fully implemented in the near future.

Future research

After the technical analysis of the implementation of genomic selection in the Latxa breeding program developed in this dissertation, further studies have been identified to address the balance between cost and returns or to analyse in which other aspects of the breeding program it could be beneficial to include genomic information. A proposal of future studies is described below:

- A study about the economic cost required to achieve genetic gain from genomic predictions and about the management to cover the significant number of alive AI rams need at the AI centre due to the fresh semen constraints.
- An analysis of the inclusion of genomic information in evaluations of other traits that are objectives of selection of the breeding program such as milk composition and udder morphology.
- To investigate how the introgression of French Manech is affecting the genetic structure of Latxa populations.

References

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

Calus, M. P. L., T. H. E. Meuwissen, A. P. W. De Roos, and R. F. Veerkamp. 2008 Accuracy of genomic selection using different methods to define haplotypes. Genetics. 178: 553–561.

Calvo, J. H., J. A. Bouzada, J. J. Jurado, and M. Serrano. 2006. Genetic substructure of the Spanish Manchega sheep breed. Small Ruminant Res. 64(1-2):116–125.

Chitneedi, P., J. J. Arranz, A. Suarez-Vega, E. García-Gámez, and B. Gutiérrez-Gil. 2017. Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. Anim. Genet. 48:436–446.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Harlow: Longman Scientific & Technical. New York, NY.

Habier, D., J. Tetens, F. R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009a. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92(2):433–43.

Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009b. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. 91:47–60.

Ibañez-Escriche, N. and Gonzalez-Recio, O. 2011. Review. Promises, pitfalls and challenges of genomic selection in breeding programs. Span. J. Agric. Res. 9(2):404–413.

Jenko, J., G. R. Wiggans, T. A. Cooper, S. A. E. Eaglen, W. G. L. Luff, M. Bichard, R. Pong-Wong, and J. A. Woolliams. 2017. Cow genotyping strategies for genomic selection in a small dairy cattle population. J. Dairy Sci. 100(1):439–452.

Jiménez-Montero, J. A., O. González-Recio, and R. Alenda. 2012. Genotyping strategies for genomic selection in small dairy cattle populations. Animal. 6(8):1216–1224.

Jonas, E., and D. J. de Koning. 2015. Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front. Genet. 6:49.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2016. Effect of cow reference group on validation reliability of genomic evaluation. Animal. 10(6):1061–1066.

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.

Lourenco, D. A. L., A. Legarra, S. Tsuruta, Y. Masuda, I. Aguila, and I. Misztal. 2020. Single-step genomic evaluations from theory to practice: Using SNP chips and sequence data in BLUPF90. Genes. 11(7),790:1-32.

Lund, M. S., I. van den Berg, P. Ma, R. F. Brøndum, and G. Su. 2016. How to improve genomic predictions in small dairy cattle populations. Animal. 10(6):1042–1049.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819–1829.

Meuwissen, T. H. E. 2009. Genetic management of small populations: A review. Acta. Agric. Scand. A. Anim. Sci. 59:71–79.

Meuwissen, T., B. Hayes, and M. Goddard. 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 6(1):6–14.

Mrode, R., G. M. Tarekegn, J. M. Mwacharo, and A. Djikeng. 2018. Invited review: Genomic selection for small ruminants in developed countries: How applicable for the rest of the world? Animal. 12:1333–1340.

Perez, B. C., J. C. C. Balieiro, R. Carvalheiro, F. Tirelo, G. A. Oliveira Junior, J. M. Dementshuk, J. P. Eler, J. B. S. Ferraz, and R. V. Ventura. 2019. Accounting for population structure in selective cow genotyping strategies. J. Anim. Breed. Genet. 136(1):23–39.

Rodríguez-Ramilo, S. T., J. M. Elsen, and A. Legarra. 2019. Inbreeding and effective population size in French dairy sheep: comparison between genomic and pedigree estimates. J. Dairy Sci. 102:4227–4237.

Schöpke, K., and Swalve, H. H. 2016. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal. 10(6):1050–1060.

Shumbusho, F., J. Raoul, J. M. Astruc, I. Palhiere, and J. M. Elsen. 2013. Potential benefits of genomic selection on genetic gain of small ruminant breeding programs. J. Anim. Sci. 91:3644–3657.

Solberg, T.R., A. K. Sonesson, J. A.Woolliams, and T. H. E. Meuwissen. 2008 Genomic selection using different marker types and densities. J. Anim. Sci. 86:2447–2454.

Uemoto, Y., T. Osawa, and J. Saburi. 2017. Effect of genotyped cows in the reference population on the genomic evaluation of Holstein cattle. Animal. 11(3):382–393.

Usai, M. G., S. Salaris, S. Casu, T. Sechi, S. Miari, P. Carta, and A. Carta. 2018. Feasibility of genomic predictions of Sarda breed rams using a female reference population. In proceedings of the 11th World Congress on Genetics Applied to Livestock Production. 11–16 February. Auckland, New Zealand.

van den Berg, I., T. H. E. Meuwissen, I. M. MacLeod, and M. E. Goddard. 2019. Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. J. Dairy Sci. 102(4):3155–3174.

VanRaden, P. M. 2020. Symposium review: How to implement genomic selection. J. Dairy Sci. 103(6):5291–5301.

FINAL CONCLUSIONS

FINAL CONCLUSIONS

This dissertation is presented as an in depth analysis of the implementation of genomic selection in the Latxa breeding program, in which the inclusion of genomic information in milk yield genetic evaluations have been assessed based on real and simulated data, as well as the estimation of several genetic diversity parameters. The final conclusions of this thesis are:

First. The increase in inbreeding, estimated based on pedigree or genomic information, is moderate in the three Latxa populations and seems to have been well controlled so far, although it has been done based on pedigree estimations.

Second. The effective population size of LCR is bigger than LCNEUS and LCNNAF, although in all populations the size is big enough to ensure long-term viability.

Third. Coancestry between LCR and MTR populations has increased to a greater extent than within them, probably due to the systematic semen importation to the Latxa population.

Fourth. The inclusion of genomic information provides nearly unbiased and less dispersed predictions, in contrast with pedigree based evaluations.

Fifth. With the current available data, single-step genomic BLUP evaluations are advantageous in terms of prediction accuracy, in contrast with pedigree based evaluations.

Sixth. Genotyping males is beneficial, when they are not isolated individuals and they have progeny with phenotypic records.

Seventh. Genotyping females with their own lactation records increases prediction accuracy, regardless of their connection level or their breeding value.

Eighth. The combined genotyping of both sexes together with greater genotyped population sizes result in increased prediction accuracy, being beneficial and worthwhile.

CONCLUSIONES FINALES

CONCLUSIONES FINALES

Esta tesis presenta un análisis en profundidad sobre la implementación de la selección genómica en el programa de mejora de la raza Latxa. Se evalúa la inclusión de información genómica en evaluaciones genéticas para producción de leche basándose en datos reales y simulados, así como se estiman parámetros de diversidad genética. Las conclusiones finales de la presente tesis son:

Primera. El aumento de la consanguinidad, estimado tanto mediante datos de pedigrí como información genómica, es moderado en las tres poblaciones de Latxa y parece haber estado bien controlado, a pesar de haberse hecho en base a datos de pedigrí.

Segunda. El tamaño efectivo de LCR es mayor que en LCNEUS y LCNNAF, aunque el tamaño de todas las poblaciones es suficientemente grande como para asegurar la viabilidad a largo plazo.

Tercera. El parentesco entre las poblaciones de LCR y MTR se ha incrementado en mayor medida que dentro de ella, probablemente debido a la importación sistemática de semen hecho por la población de Latxa.

Cuarta. La inclusión de información genómica proporciona predicciones prácticamente no sesgadas y poco dispersas, en comparación con las evaluaciones basadas en datos de pedigrí.

Quinta. Los datos disponibles hasta la fecha reflejan que las evaluaciones genómicas bajo la aproximación “Single-Step” incrementan la precisión sobre la valoración BLUP, en comparación con las evaluaciones basadas en datos de pedigrí.

Sexta. El genotipado de machos es beneficioso, siempre y cuando se trate de individuos no aislados y tengan descendientes con datos fenotípicos.

Séptima. Genotipar hembras con lactaciones registradas aumenta la precisión de predicción, independientemente de su conexión con la población o de su valor genético.

Octava. El genotipado combinando ambos sexos junto con el aumento de la población genotipada incrementa la precisión de predicción, resultando beneficioso.

ANNEXES

ANNEXES

Annexe I

GENETIC EVALUATION MODEL FOR MILK COMPOSITION AND UDDER MORPHOLOGY CHARACTERS

It is known that selection focused only on milk yield can produce changes on genetically correlated traits, so it is interesting to consider other traits as milk composition or functional traits. In Latxa breeding programs exploratory studies were done to analyze the inclusion of milk composition (Legarra and Ugarte, 2001) and udder morphology (Legarra et al., 1999, 2001; Ugarte et al., 2001) traits, whose selection into the scheme stared in 2005. The analyses are performed using the BLUPf90 software suite (Misztal et al., 2002) and the genetic evaluation models used for these traits are as follows:

Genetic evaluation model for milk composition

The genetic evaluation for milk composition traits is done by a multi-trait model for milk yield, fat yield (Kg.), protein yield (Kg.), fat percentage and protein percentage. The evaluation model differs depending on the trait:

Milk, fat and protein yield:

$$Y_{ijklm} = FYS_i + A_j + L_k + I_l + u_m + p_m + e_{ijklm}$$

Fat and protein percentage:

$$Y_{ijklm} = FYS_i + A_j + L_k + C_l + u_m + p_m + e_{ijklm}$$

Where for the evaluated traits ($\boldsymbol{Y_{ijklm}}$), flock-year-season (***FYS***), age-parity number (***A***), number of lambs born alive (***L***) and interval from lambing to first milk recording (***I***) or a combination of controls inside lactation with taken milk sample (***C***) are included as fixed effects; and additive genetic (***u***) with unknown parent groups and individual permanent (***p***) are considered as random effects.

Heritabilities for fat yield, protein yield, fat percentage and protein percentage and genetic correlations between traits are shown in Table 1.

Table 1: Heritabilities (diagonal) and genetic correlations (above diagonal) between milk yield and milk composition traits.

	Milk yield	Fat yield	Protein yield	Fat %	Protein %
Milk yield	**0.19**	0.85	0.93	-0.27	-0.35
Fat yield		**0.17**	0.89	0.25	-0.09
Protein yield			**0.18**	-0.057	0.01
Fat %				**0.17**	0.56
Protein %					**0.47**

Genetic evaluation model for udder morphology

Udder traits are also evaluated by a multi-trait model for milk yield, udder depth and attachment and teat length and verticality. The linear model used is:

$$Y_{ijkl} = FYA_i + A_j + S_k + b \cdot MYDS_{ijkl} + u_l + e_{ijkl}$$

Where for the evaluated traits ($\boldsymbol{Y_{ijkl}}$), flock-year-assessor (***FYA***), age-parity number (***A***), stage of lactation (***S***), and milk yield produced on the day of scoring (***MYDS***) as a covariable are included as fixed effects; and additive genetic (***u***) with unknown parent groups is considered as random effect. Milk yield produced on the day of scoring is included in the model as a covariate, to compensate for the effect of the udder fill.

Heritabilities for milk yield, udder depth and attachment and teat length and verticality, and genetic correlations between traits are shown in Table 2.

Table 2: Heritabilities (diagonal) and genetic correlations (above diagonal) between milk yield and udder morphology traits.

	Milk yield	Udder depth	Udder attachment	Teat verticality	Teat length
Milk yield	**0.22**	0.57	0.07	-0.39	-0.11
Udder depth		**0.23**	-0.43	-0.33	0.01
Udder attachment			**0.20**	0.29	0.14
Teat verticality				**0.40**	0.38
Teat length					**0.36**

References

Legarra, A., J. Arranz, I. Beltrán de Heredia, and E. Ugarte. 1999. Sistema de calificación de morfología mamaria en ovejas de raza Latxa: Resultados preliminares. ITEA, Vol. Extra 20: 345–347.

Legarra, A., and E. Ugarte. 2001. Genetic parameters of milk traits in Latxa dairy sheep. Animal Science. 73:407–412.

Legarra, A., E. Ugarte, I. Beltrán de Heredia, and J. Arranz. 2001. Parámetros genéticos y respuesta a diferentes índices de selección de caracteres de morfología mamaria en la raza Latxa. ITEA, Vol. Extra 22:24–26.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). In proceedings of the 7th World Congress on Genetics Applied to Livestock Production. 19-23 August. Montpellier, France.

Ugarte, E., A. Legarra, I. Beltrán de Heredia, and J. Arranz. 2001. Udder morphology: a new trait to introduce in the Latxa breeding programme. In proceedings of the 52nd Annual Meeting of the European Association of Animal Production. 26-29 August. Budapest, Hungary.

Annexe II

CONSANGUINIDAD Y PARENTESCO EN LA RAZA OVINA DE LECHE LATXA

Published in the technical manuscript: Granado-Tajada, I., S. T. Rodríguez-Ramilo, A. Legarra, and E. Ugarte. 2019. Consanguinidad y parentesco en la raza ovina de leche Latxa. Albeitar. 229:16-18. https://www.portalveterinaria.com/pdfjs/web/viewer.php?file=%2Fupload%2Friviste%2FALBEITAR229_MR.pdf

Introducción

La raza Latxa es una raza de ovino lechero autóctono del País Vasco y Navarra en la que se distinguen tres poblaciones: Latxa Cara Rubia (LCR), Latxa Cara Negra de Euskadi (LCNEUS) y Latxa Cara Negra de Navarra (LCNNAF). El programa de mejora comenzó en 1984 y actualmente tiene resultados consolidados, con una ganancia genética anual para producción de leche del 3-4 %. Actualmente el programa de mejora cuenta con 34.077, 25.717 y 12.087 hembras adultas de LCR, LCNEUS y LCNNAF, respectivamente. Cabría pensar que como resultado del mencionado progreso haya habido un incremento en la consanguinidad y, asociada a la misma, una disminución de la variabilidad genética (Falconer y Mackay, 1996). Sin embargo, hay que tener en cuenta el intercambio de material genético (animales y semen) existente entre la razas Latxa y Manech, concretamente en LCR y LCNNAF. La raza Manech está localizada en la País Vasco francés y se puede considerar como una misma raza con estándares raciales muy similares y en la que también se diferencian dos poblaciones: Manech Tête Rousse (MTR) y Manech Tête Noire (MTN).

Tradicionalmente, los programas de mejora animal han estimado y gestionado la consanguinidad en función de la genealogía, principalmente en el control de apareamientos y al seleccionar corderos para el centro de inseminación. La consanguinidad se entiende como la probabilidad de que dos alelos o genes homólogos elegidos aleatoriamente sean idénticos por descendencia, es decir, que procedan de un ancestro común (Wright, 1922). Sin embargo, en general, el ovino lechero presenta un alto porcentaje de genealogía desconocida debido al uso de la monta natural no controlada y al escaso número de test de filiaciones que se han estado realizando hasta la actualidad.

En el caso de la raza Latxa, en el histórico de genealogía, el porcentaje medio de animales en genealogía con padre y madre desconocidos es del 32 % y el de animales con madre conocida del 36 %.Como consecuencia, es probable que las estimas de consanguinidad basadas en el pedigrí sean infraestimadas. Por ejemplo, dos ovejas pueden ser medias hermanas del mismo padre de monta natural, sin embargo esta información no está registrada.

El aumento progresivo de la información molecular del ADN ha permitido obtener estimas más precisas de la relación entre animales y, como consecuencia, de los coeficientes de consanguinidad y del parentesco (Rupp et al., 2016). A tal efecto, se han desarrollado diferentes métodos que, mediante marcadores, hacen posible evaluar el impacto de la consanguinidad en una población y mejorar el manejo en función de parámetros más precisos (Caballero y Toro, 2002; Curik et al., 2014; Howard et al., 2017).

Material y Métodos

Dentro de este contexto, a través de este trabajo se planteó estimar y comparar el incremento de consanguinidad en las tres poblaciones de Latxa. Por un lado, mediante un enfoque clásico, se realizaron las estimas considerando todo el pedigrí disponible (PED). Por otro lado, se utilizó la información molecular disponible (373, 192 y 427 machos de inseminación artificial en LCNEUS, LCNNAF y LCR, respectivamente) y se obtuvieron estimas en base a la proporción de SNP homocigotos (SNP) y en base a la proporción de segmentos homocigotos (ROH). Se quiso además, analizar la influencia que ha tenido el intercambio genético que ha habido con la raza Manech.

Resultados y Discusión

Estimas de consanguinidad

En la Figura 1 se muestra la evolución de la consanguinidad obtenida en función del pedigrí existente. En todos los casos se observa una tendencia ascendente en la que los machos muestran mayores valores que las hembras. En la gráfica destaca la evolución correspondiente a LCR, que refleja cómo a partir de finales de los años 2000 los valores de consanguinidad se mantienen constantes. Se considera que este hecho está asociado a la importación de semen de la raza MTR que, aunque se ha hecho históricamente, fue a partir de ese momento cuando comenzó a convertirse en una práctica sistemática.

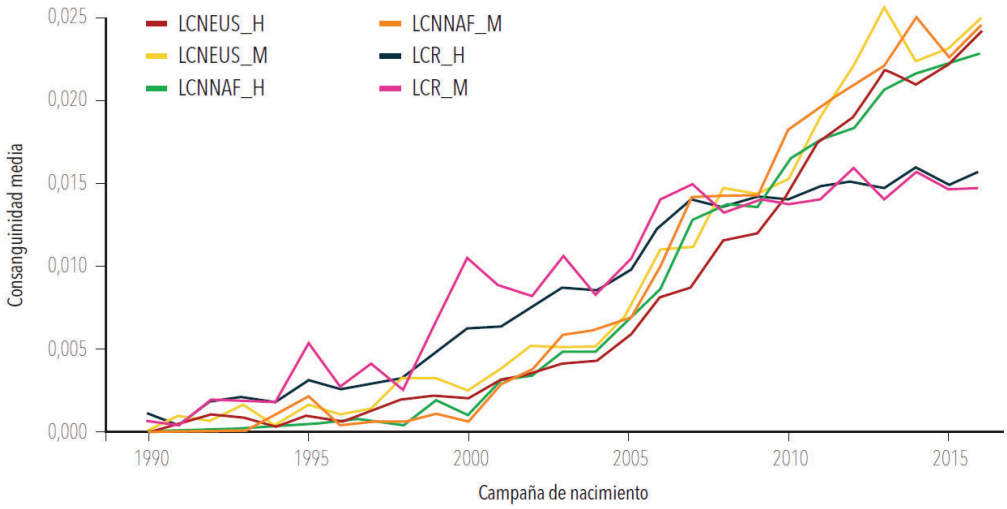

Figura 1: Evolución de la consanguinidad en función del pedigrí de los machos (M) y hembras (H) de Latxa Cara Negra de Euskadi (LCNEUS), Latxa Cara Negra de Navarra (LCNNAF) y Latxa Cara Rubia (LCR).

En la Tabla 1 se muestra la tasa de incremento de consanguinidad por generación (ΔF) estimada en función de las diferentes fuentes de información, y se puede apreciar que hay diferencias en función de la metodología empleada. Al utilizar el pedigrí (ΔF_{PED}), con un alto porcentaje de genealogías desconocidas, se obtuvieron incrementos bajos que no concuerdan con un programa de selección que ha mostrado progresos genéticos constantes. Esto evidencia la importancia de contar con pedigrís completos y con suficiente profundidad para poder reflejar la situación real de la población. Por otra parte, al usar la información del ADN, se pudo ver que en el caso de las poblaciones de Latxa Cara Negra los incrementos mayores eran los obtenidos en base a los ROH, seguidos de los PED y de los SNP. La variabilidad obtenida dentro de cada población con las estimas basadas en la información molecular subraya la importancia de tener un número alto de animales genotipados para poder obtener estimas precisas. En el caso de LCR, se observa que en las estimas obtenidas en base a la información molecular los ΔF son muy bajos y cercanos a cero, debido probablemente al efecto de la importación de semen de MTR. Actualmente, LCNNAF también importa semen de MTN, pero su efecto aún no se ha visto reflejado en la consanguinidad de esta población. Además, se pudo ver que las estimas de ΔF en función de la información molecular eran imprecisas cuando el número de animales genotipados era limitado, como es el caso.

Tabla 1: Tasa de incremento de consanguinidad por generación (ΔF) ± error estándar; calculado con el pedigrí (PED), SNP a SNP (SNP) y segmentos de homocigotos (ROH); de Latxa Cara Negra de Euskadi (LCNEUS), Latxa Cara Negra de Navarra (LCNNAF) y Latxa Cara Rubia (LCR).

Método	Raza		
	LCNEUS	LCNNAF	LCR
ΔF_{PED}	0.0017 ± 0.0001	0.0023 ± 0.0001	0.0020 ± 0.0001
ΔF_{SNP}	0.0012 ± 0.0004	0.0022 ± 0.0009	-0.0001 ± 0.0004
ΔF_{ROH}	0.0054 ± 0.0010	0.0049 ± 0.0018	0.0007 ± 0.0009

Estimas de contribución de fundadores y parentesco

Analizando la contribución que ha tenido la raza Manech en la raza Latxa, en función del pedigrí de los machos de inseminación, se estimó que la raza francesa ha contribuido en un 33 % a la diversidad genética de la población actual, fundamentalmente debido a la mencionada importación sistemática de semen.

Para ver cómo ha evolucionado la relación entre ambas razas, se comparó el parentesco molecular medio estimado para los individuos más antiguos (25 % de los machos genotipados más antiguos) con el estimado para individuos más actuales (25 % de los más jóvenes). En la Figura 2 se puede apreciar como el parentesco entre los individuos LCR y MTR más antiguos, era menor que el existente en la actualidad (aumento de 0,0084). Analizando ese mismo parámetro dentro de cada raza, se observó que, aunque en menor medida, también ha aumentado. Así, la evolución del parentesco en la última década entre individuos ha aumentado 0,0058 en LCR y 0,0046 en MTR, respectivamente, aunque en este caso solo se disponía de información molecular de los individuos que han sido utilizados en LCR y no del total de la población. En las razas Latxa Cara Negra, también ha habido un aumento del parentesco, más marcado en la población de Euskadi (0,0108) que en la de Navarra (0,0019), probablemente debido a que la población de LCNEUS es más cerrada y sin importación de semen, a diferencia de LCR y LCNNAF.

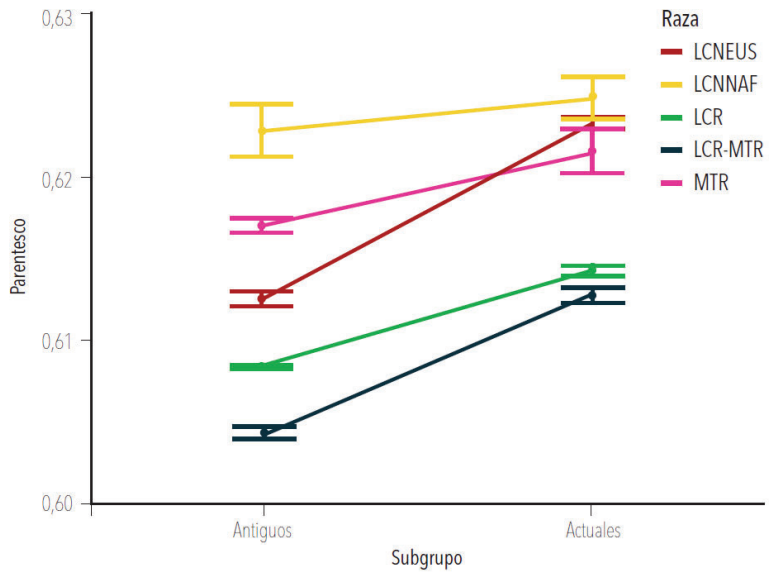

Figura 2: Parentesco dentro y entre razas Latxa Cara Negra de Euskadi (LCNEUS) y Latxa Cara Negra de Navarra (LCNNAF), Latxa Cara Rubia (LCR) y Manech Tête Rousse (MTR) de individuos antiguos y actuales. Las barras indican el intervalo de confianza al 95 %.

Conclusión

Se ha podido comprobar utilizando diferentes metodologías que el programa de mejora ha repercutido en el aumento de la tasa de consanguinidad de cada una de las poblaciones y que las estimas obtenidas son coherentes con la realidad de cada programa. Dado que este parámetro está íntimamente unido con la diversidad genética, esta será menor de lo esperable en función del tamaño poblacional. Además, las estimas de parentesco indican que como consecuencia del propio proceso de selección, el parentesco dentro de cada raza está aumentando y que en el caso de LCR y MTR, debido a la importación de semen, el parentesco entre ambas razas ha aumentado más que el existente dentro de cada raza lo que refleja la importancia y consecuencias del uso de semen francés. Si se continúa de forma sistemática con esta práctica es de esperar que a largo plazo ambas poblaciones se conviertan en una sola.

Referencias

Caballero, A., and M.A. Toro. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet. 3:289–299.

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.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to quantitative genetics. Harlow: Longman Scientific & Technical. New York, NY.

Howard, J.T., J.E. Pryce, C. Baes and C. Maltecca. 2017. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100:1–16.

Rupp, R., S. Mucha, H. Larroque, J. McEwan, and J. Conington. 2016. Genomic application in sheep and goat breeding. Anim. Front. 6:39–44.

Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330–338.

Annexe III

Supplementary Table 1: Γ matrix estimates for MF, defined from 1985 to 2015 every 3 years for Latxa Cara Negra from Euskadi (LCNEUS).

	MF1	MF2	MF3	MF4	MF5	MF6	MF7	MF8	MF9	MF10	MF11
MF1	0.632	0.259	0.384	0.417	0.476	0.465	0.437	0.442	0.410	0.410	0.410
MF2	0.259	1.027	0.392	0.316	0.335	0.363	0.312	0.358	0.349	0.349	0.349
MF3	0.384	0.392	0.843	0.352	0.392	0.417	0.397	0.406	0.393	0.393	0.393
MF4	0.417	0.316	0.352	0.848	0.456	0.365	0.395	0.401	0.380	0.380	0.380
MF5	0.476	0.335	0.392	0.456	0.768	0.415	0.397	0.413	0.390	0.390	0.390
MF6	0.465	0.363	0.417	0.365	0.415	0.863	0.394	0.388	0.393	0.393	0.393
MF7	0.437	0.312	0.397	0.395	0.397	0.394	0.995	0.295	0.356	0.356	0.356
MF8	0.442	0.358	0.406	0.401	0.413	0.388	0.295	0.894	0.338	0.338	0.338
MF9	0.410	0.349	0.393	0.380	0.390	0.393	0.356	0.338	0.858	0.356	0.356
MF10	0.410	0.349	0.393	0.380	0.390	0.393	0.356	0.338	0.356	0.858	0.356
MF11	0.410	0.349	0.393	0.380	0.390	0.393	0.356	0.338	0.356	0.356	0.858

Supplementary Table 2: Γ matrix estimates for MF, defined from 1985 to 2015 every 3 years for Spanish progeny and every 9 years for French origin for Latxa Cara Rubia (LCR).

	MF1	MF2	MF3	MF4	MF5	MF6	MF7	MF8	MF9	MF10	MF11	MF12	MF13	MF14
MF1	0.681	0.258	0.309	0.394	0.433	0.424	0.454	0.446	0.450	0.443	0.322	0.373	0.429	0.442
MF2	0.258	1.053	0.308	0.358	0.225	0.359	0.376	0.380	0.383	0.376	0.242	0.332	0.380	0.364
MF3	0.309	0.308	1.034	0.337	0.302	0.294	0.356	0.363	0.364	0.372	0.265	0.310	0.371	0.353
MF4	0.394	0.358	0.337	0.908	0.402	0.444	0.450	0.431	0.434	0.409	0.297	0.383	0.398	0.406
MF5	0.433	0.225	0.302	0.402	0.827	0.419	0.416	0.415	0.427	0.408	0.290	0.438	0.433	0.429
MF6	0.424	0.359	0.294	0.444	0.419	0.785	0.446	0.419	0.435	0.414	0.293	0.446	0.428	0.428
MF7	0.454	0.376	0.356	0.450	0.416	0.446	0.667	0.464	0.461	0.435	0.317	0.468	0.454	0.459
MF8	0.446	0.380	0.363	0.431	0.415	0.419	0.464	0.599	0.463	0.436	0.318	0.464	0.460	0.468
MF9	0.450	0.383	0.364	0.434	0.427	0.435	0.461	0.463	0.565	0.467	0.312	0.479	0.472	0.469
MF10	0.443	0.376	0.372	0.409	0.408	0.414	0.435	0.436	0.467	0.613	0.322	0.476	0.480	0.472
MF11	0.322	0.242	0.265	0.297	0.290	0.293	0.317	0.318	0.312	0.322	1.450	0.327	0.339	0.317
MF12	0.373	0.332	0.310	0.383	0.438	0.446	0.468	0.464	0.479	0.476	0.327	0.678	0.483	0.483
MF13	0.429	0.380	0.371	0.398	0.433	0.428	0.454	0.460	0.472	0.480	0.339	0.483	0.579	0.506
MF14	0.442	0.364	0.353	0.406	0.429	0.428	0.459	0.468	0.469	0.472	0.317	0.483	0.506	0.601

Annexe IV

Supplementary Table 1: Distribution of markers across chromosomes.

Chr	No. markers	Chr	No. markers
1	4,713	14	956
2	4,511	15	1,341
3	4,122	16	1,241
4	2,206	17	1,125
5	1,921	18	1,169
6	2,074	19	1,028
7	1,849	20	901
8	1,692	21	707
9	1,727	22	921
10	1,479	23	901
11	984	24	627
12	1,398	25	792
13	1,413	26	749

Supplementary Table 2: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error for Latxa Cara Negra from Euskadi with unrestricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with complete pedigree with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; P: only individuals with complete pedigree; Max: maximum genotyping per year.

Scenario	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIMax	9,248 298	9,248 298		0.2729 ± 0.0247
MP	4,205 130	4,205 130		0.2769 ± 0.0271
FBIMax	9,299 300	298 298	9,001 300	0.5369 ± 0.0396
FEIMax	9,299 300	298 298	9,001 300	0.5492 ± 0.0393
FBP	4,817 151	298 298	4,519 151	0.3657 ± 0.0291
FEP	5,322 167	298 298	5,024 167	0.3789 ± 0.0308
MP+FBI	9,299 300	4,205 130	5,094 170	0.5419 ± 0.0351
MP+FEI	9,296 300	4,205 130	5,091 170	0.4801 ± 0.0400
MP+FBP	9,298 300	4,205 130	5,093 170	0.5698 ± 0.0360
MP+FEP	9,295 300	4,205 130	5,090 170	0.5019 ± 0.0358

Supplementary Table 3: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error for Latxa Cara Negra from Euskadi with restricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with complete pedigree with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; P: only individuals with complete pedigree; Max: maximum genotyping per year; R: restricted start of genotyping.

Scenario	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIMaxR	2,348 293	2,348 293		0.1206 ± 0.0118
MPR	1,428 161	1,428 161		0.1215 ± 0.0190
FBIMaxR	2,399 300	298 298	2,101 300	0.2393 ± 0.0204
FEIMaxR	2,399 300	298 298	2,101 300	0.2217 ± 0.0181
FBPR	1,892 228	298 298	1,594 228	0.2223 ± 0.0345
FEPR	1,751 208	298 298	1,453 208	0.1564 ± 0.0167
MP+FBIR	2,399 300	1,428 161	971 139	0.1998 ± 0.0160
MP+FEIR	2,399 300	1,428 161	971 139	0.1896 ± 0.0206
MP+FBPR	2,398 300	1,428 161	970 139	0.1580 ± 0.0152
MP+FEPR	2,398 300	1,428 161	970 139	0.1874 ± 0.0194

Supplementary Table 4: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error for Latxa Cara Rubia with unrestricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with complete pedigree with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; P: only individuals with complete pedigree; Max: maximum genotyping per year.

Scenario	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIMax	8,333 268	8,333 268		0.1729 ± 0.0145
MP	4,775 150	4,775 150		0.1131 ± 0.0154
FBIMax	9,288 300	287 287	9,001 300	0.3322 ± 0.0207
FEIMax	9,288 300	287 287	9,001 300	0.3513 ± 0.0366
FBP	4,652 146	287 287	4,365 146	0.2375 ± 0.0147
FEP	5,084 160	287 287	4,797 160	0.2526 ± 0.0218
MP+FBI	9,288 300	4,775 150	4,513 150	0.2575 ± 0.0208
MP+FEI	9,288 300	4,775 150	4,513 150	0.2522 ± 0.0149
MP+FBP	9,064 293	4,775 150	4,289 143	0.2622 ± 0.0181
MP+FEP	9,065 293	4,775 150	4,290 143	0.2476 ± 0.0186

Supplementary Table 5: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error for Latxa Cara Rubia with restricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with complete pedigree with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; P: only individuals with complete pedigree; Max: maximum genotyping per year; R: restricted start of genotyping.

Scenario	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIMaxR	2,387 300	2,387 300		0.0683 ± 0.0066
MPR	1,822 219	1,822 219		0.0768 ± 0.0105
FBIMaxR	2,388 300	287 287	2,101 300	0.1274 ± 0.0141
FEIMaxR	2,388 300	287 287	2,101 300	0.1154 ± 0.0106
FBPR	2,009 246	287 287	1,722 246	0.1060 ± 0.0121
FEPR	2,006 246	287 287	1,719 246	0.1315 ± 0.0170
MP+FBIR	2,388 300	1,822 219	566 81	0.0722 ± 0.0099
MP+FEIR	2,389 300	1,822 219	567 81	0.0801 ± 0.0109
MP+FBPR	2,387 300	1,822 219	565 81	0.0774 ± 0.0105
MP+FEPR	2,388 300	1,822 219	566 81	0.0811 ± 0.0111

Supplementary Table 6: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error for Latxa Cara Negra from Navarre with unrestricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with complete pedigree with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; P: only individuals with complete pedigree; Max: maximum genotyping per year.

Scenario	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIMax	4,093 132	4,093 132		0.0777 ± 0.0168
MP	2,402 76	2,402 76		0.0607 ± 0.0118
FBIMax	4,626 150	126 126	4,500 150	0.1252 ± 0.0186
FEIMax	4,626 150	126 126	4,500 150	0.1755 ± 0.0190
FBP	2,013 65	126 126	1,887 65	0.1447 ± 0.0169
FEP	2,267 74	126 126	2,141 74	0.1539 ± 0.0289
MP+FBI	4,626 150	2,402 76	2,224 74	0.1672 ± 0.0216
MP+FEI	4,628 150	2,402 76	2,226 74	0.1519 ± 0.0212
MP+FBP	4,416 143	2,402 76	2,014 69	0.1020 ± 0.0177
MP+FEP	4,418 143	2,402 76	2,016 70	0.1570 ± 0.0248

Supplementary Table 7: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error for Latxa Cara Negra from Navarre with restricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with complete pedigree with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; P: only individuals with complete pedigree; Max: maximum genotyping per year; R: restricted start of genotyping.

Scenario	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIMaxR	1,056 133	1,056 133		0.0444 ± 0.0085
MPR	714 84	714 84		0.0289 ± 0.0141
FBIMaxR	1,176 150	126 126	1,050 150	0.0882 ± 0.0227
FEIMaxR	1,176 150	126 126	1,050 150	0.0724 ± 0.0126
FBPR	752 89	126 126	626 89	0.0240 ± 0.0141
FEPR	758 90	126 126	632 90	0.0540 ± 0.0144
MP+FBIR	1,176 150	714 84	462 66	0.0931 ± 0.0173
MP+FEIR	1,176 150	714 84	462 66	0.0533 ± 0.0161
MP+FBPR	1,176 150	714 84	462 66	0.0678 ± 0.0172
MP+FEPR	1,176 150	714 84	462 66	0.0799 ± 0.0191

Supplementary Table 8: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error, by each of the percentages for Latxa Cara Negra from Euskadi with unrestricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge.

Scenario	Percentage	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MI	10	1,114 27	1,114 27		0.1257 ± 0.0134
MI	20	1,930 54	1,930 54		0.1306 ± 0.0185
MI	30	2,746 82	2,746 82		0.2017 ± 0.0196
MI	50	4,378 136	4,378 136		0.2429 ± 0.0214
MI	70	6,010 190	6,010 190		0.3049 ± 0.0286
MI+FBI	10+90	9,299 300	1,114 27	8,185 273	0.4727 ± 0.0480
MI+FBI	20+80	9,299 300	1,930 54	7,369 246	0.5633 ± 0.0430
MI+FBI	30+70	9,299 300	2,746 82	6,553 218	0.5344 ± 0.0428
MI+FBI	50+50	9,299 300	4,378 136	4,921 164	0.5307 ± 0.0415
MI+FBI	70+30	9,299 300	6,010 190	3,289 110	0.4887 ± 0.0411
MI+FEI	10+90	9,301 300	1,114 27	8,187 273	0.5177 ± 0.0399
MI+FEI	20+80	9,299 300	1,930 54	7,369 246	0.5274 ± 0.0370
MI+FEI	30+70	9,297 300	2,746 82	6,551 218	0.5060 ± 0.0362
MI+FEI	50+50	9,301 300	4,378 136	4,923 164	0.4494 ± 0.0316
MI+FEI	70+30	9,297 300	6,010 190	3,287 110	0.4084 ± 0.0308

Supplementary Table 9: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error, by each of the percentages for Latxa Cara Negra from Euskadi with restricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; R: restricted start of genotyping.

Scenario	Percentage	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIR	10	512 31	512 31		0.0316 ± 0.0103
MIR	20	726 61	726 61		0.0572 ± 0.0123
MIR	30	940 92	940 92		0.0856 ± 0.0134
MIR	50	1,368 153	1,368 153		0.0646 ± 0.0106
MIR	70	1,796 214	1,796 214		0.1174 ± 0.0172
MI+FBIR	10+90	2,399 300	512 31	1,887 269	0.2360 ± 0.0265
MI+FBIR	20+80	2,399 300	726 61	1,673 239	0.1818 ± 0.0181
MI+FBIR	30+70	2,399 300	940 92	1,459 208	0.1636 ± 0.0200
MI+FBIR	50+50	2,399 300	1,368 153	1,031 147	0.2006 ± 0.0208
MI+FBIR	70+30	2,399 300	1,796 214	603 86	0.1593 ± 0.0252
MI+FEIR	10+90	2,399 300	512 31	1,887 268	0.2160 ± 0.0199
MI+FEIR	20+80	2,399 300	726 61	1,673 239	0.2414 ± 0.0197
MI+FEIR	30+70	2,399 300	940 92	1,459 208	0.2322 ± 0.0200
MI+FEIR	50+50	2,399 300	1,368 153	1,031 147	0.1920 ± 0.0169
MI+FEIR	70+30	2,399 300	1,796 214	603 86	0.1475 ± 0.0143

Supplementary Table 10: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error, by each of the percentages for Latxa Cara Rubia with unrestricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge.

Scenario	Percentage	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MI	10	919 21	919 21		0.0525 ± 0.0078
MI	20	1,551 42	1,551 42		0,0591 ± 0.0120
MI	30	2,183 63	2,183 63		0.0683 ± 0.0093
MI	50	3,447 105	3,447 105		0.1093 ± 0.0137
MI	70	4,711 147	4,711 147		0.1157 ± 0.0162
MI+FBI	10+90	9,288 300	919 21	8,369 279	0.2891 ± 0.0218
MI+FBI	20+80	9,288 300	1,551 42	7,737 258	0.2603 ± 0.0213
MI+FBI	30+70	9,288 300	2,183 63	7,105 237	0.2794 ± 0.0138
MI+FBI	50+50	9,288 300	3,447 105	5,841 195	0.2304 ± 0.0171
MI+FBI	70+30	9,288 300	4,711 147	4,577 153	0.2904 ± 0.0229
MI+FEI	10+90	9,294 300	919 21	8,375 279	0.2941 ± 0.0170
MI+FEI	20+80	9,288 300	1,551 42	7,737 258	0.3026 ± 0.0235
MI+FEI	30+70	9,282 299	2,183 63	7,099 237	0.3250 ± 0.0196
MI+FEI	50+50	9,294 300	3,447 105	5,847 195	0.3075 ± 0.0296
MI+FEI	70+30	9,282 299	4,711 147	4,571 152	0.2689 ± 0.0225

Supplementary Table 11: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error, by each of the percentages for Latxa Cara Rubia with restricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; R: restricted start of genotyping.

Scenario	Percentage	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIR	10	521 33	521 33		0.0238 ± 0.0065
MIR	20	755 67	755 67		0.0234 ± 0.0046
MIR	30	989 100	989 100		0.0321 ± 0.0073
MIR	50	1,457 167	1,457 167		0.0585 ± 0.0083
MIR	70	1,925 234	1,925 234		0.0522 ± 0.0097
MI+FBIR	10+90	2,388 300	521 33	1,867 267	0.0994 ± 0.0133
MI+FBIR	20+80	2,388 300	755 67	1,633 233	0.1145 ± 0.0138
MI+FBIR	30+70	2,388 300	989 100	1,399 200	0.1147 ± 0.0119
MI+FBIR	50+50	2,388 300	1,457 167	931 133	0.0858 ± 0.0129
MI+FBIR	70+30	2,388 300	1,925 234	463 66	0.0801 ± 0.0122
MI+FEIR	10+90	2,388 300	521 33	1,867 267	0.1193 ± 0.0113
MI+FEIR	20+80	2,388 300	755 67	1,633 233	0.1162 ± 0.0128
MI+FEIR	30+70	2,388 300	989 100	1,399 200	0.1212 ± 0.0126
MI+FEIR	50+50	2,388 300	1,457 167	931 133	0.1178 ± 0.0163
MI+FEIR	70+30	2,388 300	1925 234	463 66	0.0855 ± 0.0088

Supplementary Table 12: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error, by each of the percentages for Latxa Cara Negra from Navarre with unrestricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge.

Scenario	Percentage	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MI	10	633 17	633 17		0.0060 ± 0.0087
MI	20	1,140 34	1,140 34		0.0548 ± 0.0168
MI	30	1,647 51	1,647 51		0.0324 ± 0.0166
MI	50	2,658 84	2,658 84		0.0491 ± 0.0119
MI	70	3,412 110	3,412 110		0.0834 ± 0.0157
MI+FBI	10+90	4,626 150	633 17	3,993 133	0.1784 ± 0.0206
MI+FBI	20+80	4,626 150	1,140 34	3,486 116	0.1442 ± 0.0254
MI+FBI	30+70	4,626 150	1,647 51	2,979 99	0.1363 ± 0.0252
MI+FBI	50+50	4,623 150	2,658 84	1,965 66	0.1233 ± 0.0143
MI+FBI	70+30	4,499 146	3,412 110	1,087 45	0.1089 ± 0.0217
MI+FEI	10+90	4,627 150	633 17	3,994 133	0.1346 ± 0.0252
MI+FEI	20+80	4,626 150	1,140 34	3,486 116	0.1366 ± 0.0238
MI+FEI	30+70	4,625 150	1,647 51	2,978 99	0.1004 ± 0.0193
MI+FEI	50+50	4,624 150	2,658 84	1,966 66	0.1014 ± 0.0218
MI+FEI	70+30	4,500 146	3,412 110	1,088 45	0.0949 ± 0.0166

Supplementary Table 13: Number of genotyped total animals, number of males, number of females, number of animals per year of each case, mean relative accuracy difference ± standard error, by each of the percentages for Latxa Cara Negra from Navarre with restricted start of genotyping and different genotyping strategies. M: males; F: females; M+F: complementation of males with females; B: genetically best individuals; E: combination of genetically extreme animals; I: indifferent pedigree knowledge; R: restricted start of genotyping.

Scenario	Percentage	No. total No. per year	No. males No. per year	No. females No. per year	Relative difference ± SE
MIR	10	222 14	222 14		0.0062 ± 0.0095
MIR	20	318 27	318 27		0.0186 ± 0.0092
MIR	30	414 41	414 41		0.0252 ± 0.0094
MIR	50	606 69	606 69		0.0078 ± 0.0101
MIR	70	798 96	798 96		0.0003 ± 0.0085
MI+FBIR	10+90	1,176 150	222 14	954 136	0.0626 ± 0.0163
MI+FBIR	20+80	1,176 150	318 27	858 123	0.1019 ± 0.0235
MI+FBIR	30+70	1,176 150	414 41	762 109	0.0724 ± 0.0172
MI+FBIR	50+50	1,176 150	606 69	570 81	0.0497 ± 0.0159
MI+FBIR	70+30	1,176 150	798 96	382 64	0.0708 ± 0.0175
MI+FEIR	10+90	1,176 150	222 14	954 136	0.0812 ± 0.0164
MI+FEIR	20+80	1,176 150	318 27	858 123	0.0514 ± 0.0135
MI+FEIR	30+70	1,176 150	414 41	762 109	0.0938 ± 0.0133
MI+FEIR	50+50	1,176 150	606 69	570 81	0.0723 ± 0.0144
MI+FEIR	70+30	1,176 150	798 96	382 64	0.0564 ± 0.0188