



MINISTERIO
DE CIENCIA E INNOVACIÓN



Departamento de
Mejora Genética Animal

INFORME DE RESULTADOS DE:

DEFINICIÓN DE HERRAMIENTAS PARA LA CARACTERIZACIÓN DE LAS POBLACIONES DE VACUNO AUTÓCTONO, IMPUTACIÓN DE GENOTIPOS, DESARROLLO DE LAS VALORACIONES GENÉTICAS INCORPORANDO INFORMACIÓN GENÓMICA EN CADA UNA DE LAS RAZAS ASÍ COMO ELABORACIÓN DE UN INFORME CONJUNTO.

EVALUACIÓN GENÓMICA CONJUNTA DE TODAS LAS POBLACIONES DE VACUNO DE CARNE AUTÓCTONO INVOLUCRADAS EN EL ESTUDIO.

Clara Díaz Martín
Científico Titular de OPIS

Cristina Meneses González
Alejandro Rubio de Juan
Personal Técnico de Apoyo

Departamento de Mejora Genética Animal
INIA



Resumen Ejecutivo

- Para estudiar el efecto de la incorporación de la información genómica en las evaluaciones genéticas del peso al destete en cada una de las razas se han editado los datos de genotipos bajo el mismo criterio y se han generado dos conjuntos de datos de genotipados; el primero, específico de cada raza (snp\$Raza) con un número variable de marcadores para cada una de ellas (entre 42.706 en Retinta y 48.298 en Asturiana de Valles) y, el segundo (snpMETA) constituido por todos los marcadores que han superado el proceso de edición y que son comunes a todas las razas (31.754).
- El periodo de transición entre la utilización masiva de marcadores microsatélites y la de un uso masivo de marcadores de tipo SNP ha generado la necesidad de poder utilizar simultáneamente ambas tecnologías para la verificación de parentescos. El consorcio de bovino de aptitud cárnica dispone de una base de datos de 513 animales de seis de las razas (Asturiana de Valles, Avileña-Negra Ibérica, Morucha, Pirenaica, Retinta y Rubia Gallega) involucradas en esta encomienda, que supone un buen material para permitir el diseño de una herramienta de imputación útil en este periodo.
- Con la formación del grupo de trabajo de Bovino Autóctono de aptitud cárnica (BAC) se han establecido las bases para el desarrollo coordinando de las actividades de mejora genética en vacuno de carne en nuestro país.
- Los resultados obtenidos nos permiten concluir que:
 - La información genómica incrementa la fiabilidad de las evaluaciones genéticas para el peso al destete, tanto en su componente directa (VGP_PD) como, en su componente materna (VGP_LECHE) para la mayoría de las razas, variando la magnitud del cambio entre razas.
 - Como se esperaba, el efecto de la inclusión de la información genómica sobre la fiabilidad, es evidente en los animales genotipados. No obstante, es esperable que conforme se incremente el número de animales genotipados en las razas, se genere un incremento de fiabilidad en el conjunto de la población evaluada, por una mejora en la estructura de la información.
 - La distribución de los cuantiles de las predicciones de valor genómico y de la fiabilidad de dichas predicciones, no se ven afectados, es decir es similar a la de las estimas BLUP convencionales. Ello es probablemente debido a que la proporción de los animales genotipados sobre la población valorada, es aún muy reducida. Las correlaciones entre las predicciones de valor genético bajo el modelo convencional y el que incorpora la información genómica son altas o muy altas (0,782 y 0,963).



- El tamaño del chip podría ser reducido si ello supusiera una disminución sensible del precio de genotipado. Los resultados obtenidos con el set de SNPs propios de raza y los obtenidos con un set de menor tamaño que contiene los SNPs comunes a todas ellas, son similares.
- Las fiabilidades de los GVGP_m obtenidos en la metapoblación a partir de un modelo equivalente, suponen un incremento de fiabilidad para todas las razas. En este modelo donde asumimos que la variabilidad genética es la misma en todas las razas y que ésta es igual a la media ponderada de las varianzas en cada una de ellas, observamos que hay un incremento de la precisión. Sin embargo, las correlaciones de las predicciones de valor genético obtenidas en la metapoblación con las obtenidas en un modelo que no integra la información genómica, varían de unas poblaciones a otras lo que sugiere, que aun existiendo haplotipos comunes entre ellas, las fases podrían cambiar entre las distintas poblaciones.
- El número total de marcadores microsatélites utilizados en el conjunto de las razas participantes en la encomienda es de cuarenta y dos. En los marcadores comunes (17) existen alelos privados de raza. Solo 15 de los micros comunes están presentes en los tríos generados en el proyecto SelGenBeef. Esto es importante porque de estos animales se dispone de genotipados de alta densidad, lo que favorece la búsqueda fiable de los haplotipos de SNPs que nos permiten imputar los alelos de los micros.
- Utilizar como población de referencia el conjunto de todos los animales genotipados con alta densidad nos ha permitido establecer el procedimiento de imputación. Sin embargo, hay que integrar los genotipos obtenidos en esta encomienda para mejorar la probabilidad de imputación correcta.



Tabla de contenido

RESUMEN EJECUTIVO	2
TABLA DE CONTENIDO	4
INTRODUCCION	5
1.- EVALUACIÓN GENÉTICA DE CADA UNA DE LAS RAZAS INCORPORANDO LA INFORMACIÓN GENÓMICA	7
1.1 PROCEDIMIENTO DE TRABAJO	7
1.2 RESULTADOS	9
CONCLUSIONES	17
2.- EVALUACIÓN DE LA METAPOBLACIÓN	18
2.1 PROCEDIMIENTO DE TRABAJO	18
2.2 RESULTADOS	20
CONCLUSIONES	24
3.- EVALUACIÓN GENERAL DEL AUDITOR DE ICAR	26
4.- GENERACIÓN DE HERRAMIENTAS: IMPUTACIÓN DE MICROSATÉLITES PARA LA VERIFICACIÓN DE LAS PATERNIDADES A PARTIR DE LOS CHIPS DE SNPS	28
OBJETIVO	28
4.1 PROCEDIMIENTO DE TRABAJO	28
4.2 RESULTADOS	31
CONCLUSIONES	34
LITERATURA CITADA	35



INTRODUCCION

Los objetivos principales de esta encomienda eran tres. Por un lado, hacer una evaluación de la situación de partida de los Programas de Mejora Genética; por otro, determinar el efecto de la incorporación de la información genómica en los sistemas de evaluación genética de cada raza y, por último, evaluar que beneficio se podría esperar de trabajar en el escenario de una Metapoblación, solución, que al menos teóricamente, podría funcionar, aunque la realidad es que parece que es razas/carácter dependiente (Kachman et al., 2013).

Como ya se ha mencionado, el primer informe de este proyecto abordó el primer objetivo de esta encomienda, evaluándose la situación de partida de los Programas de Mejora Genética desarrollados por cada una de las asociaciones de criadores de razas autóctonas de vacuno de carne participantes. Por orden alfabético del acrónimo, estas asociaciones son la *Asociación Nacional de Criadores de raza Retinta* (ACRE), la *Asociación Criadores de Rubia Gallega* (ACRUGA), la *Asociación Española de Criadores de Raza Avileña Negra-Ibérica* (AECRANI), la *Asociación de Criadores de raza Bovina Parda de Montaña* (ARAPARDA), la *Asociación Española de Asturiana de Valles* (ASEAVA), la *Confederación de Asociaciones de raza Pirenaica* (CONASPI), la *Federación Española de Raza Parda de Montaña* (FERPAM), la *Asociación Nacional de Criadores de Raza Morucha Selecta* (MORUCHA). Dicho estudio puso de manifiesto la heterogeneidad existente entre las distintas Asociaciones/razas/sistemas de producción y percepción de sus ganaderos hacia las herramientas genéticas/genómicas.

En este informe se abordan los dos últimos objetivos. Las actividades en relación a la incorporación de la información genómica en la evaluación genética de estas poblaciones se iniciaron en el año 2018, con el establecimiento de criterios homogéneos para proceder al genotipado de los primeros individuos de cada una de las razas. Al contrario que en el ovino de leche, las asociaciones de carne sólo disponían de algunos genotipados de SNPs (72 o 75 animales por raza, organizados en tríos (individuo, padre y madre), con un total de 513 animales. Estos datos fueron obtenidos en el marco del proyecto SelGenBeef (AGL2010-15903), proyecto financiado por el Ministerio de Ciencia e Innovación. Como ya se apuntó en el informe anterior, dicho proyecto, junto con el proyecto Gen2Farm (KBBE.2011.1.3-06), financiado por la UE, permitieron establecer las bases teóricas de lo que se podría esperar con la incorporación de la información genómica en la evaluación genética de poblaciones de tamaño limitado, y en particular, de seis de las siete razas, participantes en la encomienda. Los datos de genotipados generados en el SelGenBeef, se han utilizado en este proyecto para atender la demanda de las asociaciones de disponer de una herramienta de verificación de paternidades en el periodo de transición del uso masivo de marcadores tipo microsatélites a tipo SNP. Esta actividad la describiremos posteriormente. Durante tres años consecutivos (2018, 2019 y 2020) se han genotipado animales de las distintas razas; los datos de este último año no se incluyen en este informe porque no están aún disponibles para todas las razas.



La evaluación realizada se ha hecho en dos etapas; en una primera fase se han analizado los resultados dentro de cada una de las razas para posteriormente, trabajar en el marco de una metapoblación. Una metapoblación se define como un conjunto de poblaciones de una misma especie que interactúan a un cierto nivel (Levins, 1969). En nuestro caso, denominamos metapoblación al conjunto de todas las razas de la encomienda, aprovechando que, tanto el parentesco genómico entre ellas (González-Rodríguez et al., 2017) como la persistencia de fase (Cañas-Alvarez et al., 2016) son altos. La persistencia de fase es una medida de similitud de los patrones de desequilibrio de ligamiento entre las razas y sabemos que ésta se mantiene a distancias relativamente amplias a lo largo de todo el genoma (Cañas-Alvarez et al., 2016) aunque se reduce a mayores distancias. Así, las actividades que se van a describir en el presente informe serán relativas a:

- 1.- La evaluación genética de cada una de las razas incorporando información genómica para el carácter peso al destete.
- 2.- La evaluación genética/genómica conjunta de todas las razas (Metapoblación) uniendo la información genómica disponible en cada una de ellas para el carácter peso al destete.
- 3.- Evaluación general de ICAR sobre el funcionamiento de los esquemas de selección de las asociaciones de ganaderos participantes de la encomienda.
- 4.- Desarrollo de una herramienta de imputación de los Micros utilizados en la verificación de parentescos a partir de los SNPs de los chips de genotipado. El objetivo es asistir a las asociaciones en la verificación en el periodo de transición.



1.- EVALUACIÓN GENÉTICA DE CADA UNA DE LAS RAZAS INCORPORANDO LA INFORMACIÓN GENÓMICA

Como resultado del estudio previo de situación de los programas de mejora se identificó el peso al destete, como el fenotipo común a todas las razas. Algunas de estas razas ajustan el peso del destete a una edad determinada y otras ajustan una covariable edad en el modelo. Otros fenotipos presentaban una situación heterogénea dependiendo de la raza/tipo de sistema de producción. Por lo tanto, para estudiar el impacto de la información genómica en la evaluación genética de cada una de las razas, se optó por utilizar el peso al destete como fenotipo de consenso, además, de permitirnos evaluar dos caracteres relacionados, el potencial de crecimiento del individuo en el periodo del nacimiento al destete (PD) y la habilidad maternal de su madre (LECHE) para proporcionar el alimento que le va a permitir expresar dicho potencial.

A lo largo de todo este informe nos vamos a referir a las razas por sus abreviaturas Asturiana de Valles (AST), Avileña-Negra Ibérica (ANI), Morucha (MOR), Parda (PAR), Pirenaica (PIR), Retinta (RET) y finalmente Rubia Gallega (RUB), y se van a disponer en las tablas por orden alfabético de la abreviatura.

1.1 Procedimiento de Trabajo

El procedimiento de trabajo implicó varias actividades que se sucedieron en el tiempo:

- **Elección del chip de genotipado:** En una reunión conjunta del grupo de trabajo se estudiaron dos opciones de chips comerciales diseñados por dos grandes empresas en el mundo de la Biología Molecular como son Illumina y Thermofisher. Finalmente se optó por el chip "Axion Bovine Genotyping" de Thermofisher porque ofrecía un precio mucho más competitivo, está extendido su uso y el chip que comercializan es más denso, por lo que a priori interroga un mayor número de posiciones en el genoma.
- **Elección de los animales a genotipar, de la metodología y del software a utilizar:** En este caso los criterios fueron establecidos por el conjunto de genetistas dentro del grupo de trabajo. En general se acordó una estrategia que maximizase la variabilidad partiendo de animales con fiabilidad fundamentalmente para ambas componentes del peso al destete. De esta manera se maximiza la probabilidad de capturar la diversidad haplotípica existente en cada raza.
- **La elección de la metodología** se hizo tomando en cuenta las opciones existentes, el tamaño de las poblaciones de manera que se pueda hacer una buena estima de los efectos de los SNPs y, por último, la posibilidad de genotipar animales de cada



población que es una limitante en todos los sistemas de evaluación genómica. La opción clara en nuestro contexto es el “single step BLUP” (ssBLUP, Legarra et al, 2014) precisamente porque permite incorporar la información genómica a la información ya existente para las evaluaciones genéticas en el mismo proceso. De esta manera dicha información es una fuente adicional de información junto con los datos fenotípicos y los datos genealógicos.

Entre todos se hizo una búsqueda del **software** disponible para aplicar un ssBLUP en las poblaciones de la encomienda; finalmente se optó por el paquete BLUPf90 (Mizstal et al 2018) porque es sencillo de usar, eficiente y se mantiene permanentemente actualizado dado que el grupo desarrollador del mismo tiene una actividad intensa en el área del desarrollo metodológico para la realización de las evaluaciones genómicas. Su uso en la evaluación genética de estas razas está permitido por los autores. Dichos autores han expresado su deseo de que se reconozca su autoría en las publicaciones de las valoraciones genéticas y/o en los informes que las acompañan.

- **Proceso de edición de genotipos** y obtención de los archivos \$raza.ped (archivo que contiene los genotipos) y \$raza.map (archivo que contiene las posiciones de los marcadores):

Con el objetivo de homogeneizar los criterios de edición de datos para generar la metapoblación, el equipo INIA desarrolló un procedimiento de edición y preparación de los archivos de genotipos para la evaluación de cada una de las razas. Mediante dicho procedimiento, INIA generó los archivos de genotipos para cada una de las razas. El procedimiento se diseñó partiendo de los archivos CELL suministrados por Xenética Fontao hasta llegar al archivo de genotipos de cada una de las razas. Por cada una de ellas, se generaron dos archivos de genotipos: el propio de cada raza una vez aplicados todos los filtros (\$raza.ped) y, el de cada raza conteniendo un conjunto de marcadores comunes que habían superado los filtros de edición en todas ellas (\$raza_META.ped).

- **Evaluación genética** incorporando los datos de genotipado para cada una de las razas realizadas por los genetistas correspondientes.

Los genetistas de cada una de las razas han realizado un total de tres valoraciones genéticas; una, la convencional y otras dos incluyendo distinta información genómica; por un lado, el set de SNPs específicos de cada raza y por otro, el set de SNPs común a todas las razas. De esta manera podemos evaluar qué aporta la información genómica a cada raza y, si la pérdida de SNPs que genera pasar de la evaluación intra-raza a la metapoblación, puede afectar los resultados obtenidos a partir de esta última en cada una de las poblaciones.



Cada genetista ha elaborado un informe para cada raza. Para el desarrollo de dicho informe, INIA proporcionó un script en R, rmarkdown (2020). rmarkdown: Dynamic Documents for R. R package version 2.5.3. URL <https://rmarkdown.rstudio.com>.) que distribuyó a todos los genetistas. De esta forma nos hemos asegurado que la información contenida en cada uno ellos, sea homogénea. Dichos informes se adjuntan en el Anejo 1.

1.2 Resultados

1.2.1. – Preparación de los archivos de genotipos

Con el objeto de darle una mayor homogeneidad al análisis de los resultados de las evaluaciones genómicas en el contexto de una metapoblación, se propuso hacer una edición conjunta de datos. Así el proceso parte de los archivos CEL generados en el laboratorio de genotipado hasta llegar a crear los archivos que contienen los genotipos (.ped) necesarios para hacer la evaluación genómica. El objetivo de este procedimiento fue mantener unos criterios homogéneos entre razas, maximizar el número de marcadores SNPs comunes entre ellas, y maximizar el número de animales de genotipados que podían ser incluidos en la evaluación genómica.

Fase 1: “Axion Suite”.

Los archivos CEL de cada una de las razas se procesaron con el procedimiento estándar que recomienda Thermofisher. Dicho procedimiento consiste en mantener una serie de parámetros de calidad y, en una segunda vuelta utilizar los resultados obtenidos en la primera para recuperar las muestras que se han perdido en el proceso anterior. Sin embargo, pudimos observar que dicho procedimiento, que está optimizado para trabajar con datos de genética humana donde lo que se busca mayoritariamente son genes causales de alguna patología, resulta muy restrictivo en el contexto de la Mejora genética animal por lo que se modificó dicho protocolo de manera que, como ya hemos mencionado arriba, nos permitiera maximizar el número de individuos y el número de SNPs comunes a todas las razas. Los parámetros que se modificaron en concreto fueron: “QC_call rate= 90” que mide el número de datos faltantes por marcador, y el “Average Call Rate=97” que es un control de calidad sobre la placa. Las salidas se corresponden a la opción “Best Probe Set” mediante la cual, Axion Suite elige la mejor sonda para cada uno de los marcadores presentes en el chip con una indicación de calidad. De esta forma, el set de SNPs es siempre del mismo tamaño lo que facilita la automatización de los procedimientos para manejar los datos.

En la Tabla 1.1 se presentan los resultados de esta primera parte del proceso. Durante el periodo de ejecución del proyecto ha habido un cambio de la versión del chip utilizado pasando de la versión 2 (v2) a la versión 3 (v3). Como podemos ver en la Tabla 1.1, el número de SNPs autosómicos que comparten las dos versiones es de 61.691. Cuatro (AST, PIR, RET y RUB) de las siete razas tienen animales genotipados en las dos versiones. Por el contrario, en



ANI, MOR y PAR solo se ha utilizado la versión 3. En esta versión el número de marcadores autosómicos es ligeramente superior 61798. En esta fase del proceso se han desechado una muestra en ANI, tres en AST y dos en RET.

Tabla 1.1 Número de individuos, de marcadores por raza y versión del chip anterior y posterior al procesado de los datos con el software Axion Analysis Suite.

		Versión del chip	Raza						
			ANI	AST	MOR	PAR	PIR	RET	RUB
Nº individuos genotipados		V2	742				111	321	688
		V3	1158	1494	533	516	517	127	345
		Total	1158	2236	533	516	628	448	1033
Nº individuos genotipados Axion		V2	742				111	321	688
		V3	1157	1491	533	516	517	125	345
		Total	1157	2233	533	516	628	446	1033
Nº snp Axion		V2	67567				67567	67567	67567
		V3	63644	63644	63644	63644	63644	63644	63644
Nº snp autosómicos		Ambas	61798	61691	61798	61798	61691	61691	61691

Fase 2: Edición de genotipos con PLINK (Purcell et al., 2007)

Utilizando Plink1.9 (<http://pngu.mgh.harvard.edu/purcell/plink/>) se aplicaron una serie de filtros a los sets de datos de genotipos obtenidos en la primera fase.

Tabla 1.2 Número de individuos (ind) y marcadores (snp) por raza, en la secuencia de filtrado de datos

			Raza						
		Valor mínimo	ANI	AST	MOR	PAR	PIR	RET	RUB
Origen	ind	----	1157	2236	533	516	628	448	1033
	snp	----	61798	61691	61798	61798	61691	61691	61691
Geno	snp	0,1	58974	59794	54804	56471	58272	58252	59419
		0,05	56552	55114	48423	50490	54496	54227	55684
Mind	ind	0,1	1152	2228	533	516	628	446	1032
		0,05	1152	2177	326	413	626	446	1024
HW (p<0,0001)	snp	geno 0,1+ mind 0,1	56159	57456	53416	54132	57203	55823	58396
		geno 0,05+ mind 0,05	54150	53739	47465	49284	53786	52307	54951
MAF (min 0,01)	snp	geno 0,1+ mind 0,1	42765	48298	44119	44341	44842	42706	45483
		geno 0,05+ mind 0,05	40786	44392	37061	40077	41458	39271	42050

*Geno: mínimo de valores faltantes por snp, Mind: mínimo snp faltantes por individuo, HW: equilibrio de Hardy-Weinberg, MAF: mínima frecuencia alélica.



En la Tabla 1.2 aparece reflejado este proceso de edición. En esta fase se aplicaron los criterios estándar. Aquellos marcadores con datos faltantes para un 5% o un 10% de los individuos (--geno) se consideraban candidatos para su eliminación. Así mismo, aquellos animales (--mind) con más de un 5% o 10% de datos faltantes también fueron preseleccionados para su eliminación. Por otro lado, se determinad SNPs fijados en las razas y SNPs que no estaban en equilibrio Hardy- Weinberg como un criterio de sesgo en el genotipado. Evidentemente, como se observa en la Tabla 1.2, cuanto más restrictivo era el criterio, mayor era la pérdida de SNPs. Finalmente, la opción fue considerar como salida final la proporcionada por los siguientes parámetros:

--geno 10 --mind 10 --hw 0,0001 --maf 0.01

Como resultado de esta etapa el porcentaje de pérdida de marcadores en cada raza fue distinto oscilando entre 22% en AST y un 31% en ANI y RET. En esta fase se generaron los datos de genotipado propios de cada una de las razas para entrar a la valoración genómica. El número de marcadores propios de raza se muestra en la Tabla 1.3.

Fase 3: Unión de los archivos de genotipos de las distintas razas.

Por último, se unieron los datos de genotipado específicos de raza hasta encontrar el conjunto de marcadores que habían superado los filtros de edición y que eran comunes a todas ellas. Este set de marcadores tiene un total de 31.754 SNPs (Tabla 1.3)

Tabla 1.3 Número total de datos de peso al destete (N_{feno}), de animales en el pedigrí (N_{ped}), de animales genotipados (N_{geno}) y de marcadores (N_{snp}) que aportan información al destete en cada una de las razas y en la META población.

Raza	N_{feno}	N_{ped}	N_{geno}	N_{snp}
ANI	58.625	83.710	1.152	42.765
AST	135.536	214.687	2.153	48.298
MOR	17.815	27.104	533	44.119
PAR	6.625	15.323	516	44.341
PIR	41.177	67.554	623	44.482
RET	34.214	49.121	446	42.706
RUB	78.406	113.337	1.029	45.083
META	372.398	570.836	6.452	31.754



1.2.2. – Resultados de la evaluación genética incorporando información genómica para cada una de las razas.

Como ya hemos mencionado en el apartado anterior por cada una de las razas se realizaron tres evaluaciones genéticas; el BLUP convencional (VGP), un ssBLUP incorporando el set de genotipos propio de raza (GVGP_r), y por último, un ssBLUP con el set de marcadores comunes a todas las razas (GVGP_c). En todas las razas se ha aplicado un modelo animal con efecto genético directo, materno y ambiental permanente materno. Las estimas de varianzas covarianzas utilizadas en cada una de las razas figura en la Tabla 1.4.

Tabla 1.4 Estimaciones de las varianzas aditiva directa (σ^2_a), materna (σ^2_m), covarianza genética directo materna ($\sigma_{a,m}$), permanente materna (σ^2_{pe}), del grupo de comparación (σ^2_{gc}), residual (σ^2_e) y varianza fenotípica total (σ^2_f), heredabilidad aditiva directa (h^2_a), materna (h^2_m), correlación genética directa materna ($r_{a,m}$), cociente de varianza permanente y varianza fenotípica total (p) y cociente de la varianza de los grupos de comparación y varianza fenotípica total (gc) para cada una de las razas.

Parámetro	Raza						
	ANI	AST	MOR	PAR	PIR	RET	RUB
σ^2_a	216.70	684.48	159.77	182.43	1018.6	262.30	779.56
σ^2_m	111.50	298.27	114.12	132.27	299.15	129.4	182.23
$\sigma_{a,m}$	-111.30	-267.15	-54.01	-62.14	-368.63	-132.00	-285.61
σ^2_{pe}	64.49	47.36	68.47	80.36	48.37	103.3	61.51
σ^2_{gc}	-	287.51	-	-	524.03	409.6	393.29
σ^2_e	465.30	775.73	472.45	543.97	673.51	518.20	681.66
σ^2_f	746.69	1826.2	760.80	876.90	2195.03	1290.8	1812.64
h^2_a	0.29	0.37	0.21	0.21	0.46	0.20	0.43
h^2_m	0.15	0.16	0.15	0.15	0.14	0.10	0.10
$r_{a,m}$	-0.72	-0.59	-0.40	-0.40	-0.67	-0.72	-0.76
p	0.09	0.03	0.09	0.09	0.02	0.08	0.03
gc	-	0.16	-	-	0.24	0.32	0.22

Existe una diferencia substancial en la magnitud de las estimas de (co) varianzas y sus ratios entre las distintas razas, en parte debido a un efecto de escala. Este efecto veremos que se pondrá de manifiesto en las estimas de los efectos de los marcadores. Por otro lado, las diferencias de magnitud en los parámetros genéticos implican que, a igualdad de cantidad de información, siempre van a ser más precisas las estimas de los VGPs obtenidas en AST, PIR y RUB que en el resto de las poblaciones. Se ha intentado estimar los componentes directo y materno en MOR y PAR, pero el sistema de ecuaciones no alcanza la convergencia. Por ello,



para poder utilizar este modelo en estas dos razas, optamos por calcular las (co)varianzas partiendo de las estimas de heredabilidad y correlación de ANI y de la varianza fenotípica propia de cada una de ellas. En la Tabla 1.3 aparecía el número de animales con peso al destete, de animales en la genealogía, así como el número de animales genotipados en cada raza. Como se puede apreciar en dicha tabla, la situación en cuanto a disponibilidad de datos es muy heterogénea, tanto, en relación al número de animales con dato propio como al número animales genotipados. Para el primero, los extremos serían las razas AST y PAR; sin embargo, para la cantidad de animales genotipados, estas serían las razas AST y RET. Aunque las diferencias en cuanto a número de animales son substanciales, si nos referimos al porcentaje de animales genotipados en relación al número de animales en la genealogía no existen tantas diferencias. Dicho valor oscila entre un 1% para AST, PIR y RET y un 3% en PAR. El número de animales genotipados con dato propio de peso al destete representan entre un 0% (PIR) y casi un 0% en RUB y un 73% en el caso de RET, siendo del 44%, 60%, 45% y 30% para ANI, AST, MOR y PAR respectivamente.

Tabla 1.5. Fiabilidad media del valor genético predicho (VGP), el valor genómico con el panel específico de raza (GVGP_r) y el valor genómico con el panel común (GVGP_c) para el grupo de animales genotipados y no genotipados en cada una de las razas tanto de la componente directa (PD) como materna (LECHE).

Raza	Componente	Genotipados			No Genotipados		
		VGP	GVGP _r	GVGP _c	VGP	GVGP _r	GVGP _c
ANI	PD	50	54	53	48	48	48
	LECHE	52	54	54	48	48	48
AST	PD	70	71	71	59	59	59
	LECHE	61	61	61	49	49	49
MOR	PD	31	36	36	32	32	32
	LECHE	35	40	41	32	32	32
PAR	PD	29	30	30	24	24	24
	LECHE	28	30	30	25	25	25
PIR	PD	64	65	65	60	59	59
	LECHE	55	55	56	49	49	49
RET	PD	68	66	66	53	52	52
	LECHE	60	57	57	49	49	49
RUB	PD	78	80	80	60	60	60
	LECHE	69	71	71	51	51	51

En la Tabla 1.5 se presentan las fiabilidades medias de los VGPs, GVGP_r y GVGP_c para cada una de las razas, tanto en el grupo de animales genotipados, como de no genotipados. Las razas, en mayor o menor medida, experimentaron un incremento de fiabilidad que se hace evidente entre los animales genotipados. Para la componente de Crecimiento (GVGP_{PD}), este incremento oscila entre 0 y 5 puntos de fiabilidad que representa entre 0% en el caso de la



RET y un 14% en el caso de la MOR de la fiabilidad con respecto al modelo convencional. Aunque es cierto que en la raza MOR, la ganancia de fiabilidad es mayor, también lo es, que la situación de partida, es decir el nivel medio de fiabilidad es más bajo que en las otras razas, con la única excepción de la raza PAR. En general se aprecia una tendencia asintótica del incremento de fiabilidad, siendo dicho incremento mayor cuanto menor es la fiabilidad de partida de los animales genotipados.

Para la componente materna (GVGP_LECHE) el escenario es similar. En este caso, son AST y PIR las razas en las que no se aprecia ningún efecto sobre la fiabilidad; por el contrario, en las otras razas, el incremento oscila entre dos y seis puntos. El máximo incremento de nuevo se produce en MOR. En general, puede observarse que en todas las poblaciones se han elegido animales con una fiabilidad media más alta, que el conjunto de la población. Cabe destacar que, cuando comparamos el cambio o incremento de fiabilidad teniendo en cuenta el set de SNPs específicos de razas, (GVGP_r) o el set común que se utilizará en la metapoblación (GVGP_c), los resultados son prácticamente los mismos. Lo que sugiere que un chip de menor densidad aportaría prácticamente la misma información y podría abaratar los costes de genotipado. Además de la comparativa de las fiabilidades, estimamos las correlaciones de los VGPs, GVGP_r y GVGP_c para cada componente. Los resultados se muestran en la Tabla 1.6. Las estimas de las correlaciones son altas o muy altas (0,782 – 0,963). En general se espera que cuanto mayor es el incremento de la fiabilidad menores sean las correlaciones entre VGPs, y los GVGP_r y GVGP_c. El BLUP es un predictor centrado sobre la media, lo que significa que, cuanto mayor es la fiabilidad de los GVGP, más se distancia de 0 (la media), más dispersión existe y, por lo tanto, más se espera que se afecte la correlación. Ello ocurre particularmente en las poblaciones en las que las fiabilidades de los VGPs son pequeñas. Por la misma razón, cuando los VGPs tienen mucha precisión, no se espera que cambie la estima de GVGP y de su fiabilidad porque, en estos casos, la información genómica, no puede aportar mucha más información, que la que aporta el BLUP convencional. Sin embargo, es probable que la capacidad predictiva en estos casos sea mejor porque el valor de sustitución del SNP estará estimado con mayor precisión.

En la Figura 1.1 se presenta las soluciones de los efectos de los SNPs para la componente directa (PD) y materna (LECHE) del peso al destete en cada una de las razas. Podemos observar que, tanto en MOR como en PAR, la dispersión de los efectos de los marcadores es muy pequeña como sería esperable dada la baja fiabilidad media de los GVGP en estas razas. Por el contrario, las fiabilidades de AST y RUB presentan los valores más elevados de fiabilidad, lo que se traduce en una mayor dispersión de los efectos de sustitución alrededor de la media. La dispersión de los efectos de sustitución de la componente materna, es menor en todas las razas porque a igualdad de información, también es menor la fiabilidad para este componente. Además, se calculó la correlación entre los efectos de los SNPs para todos los pares de razas (Tabla 1.7). Dicho cálculo se hizo globalmente y cromosoma a cromosoma (estos resultados no se muestran). Todas las estimas son cercanas a cero. Esta correlación es un primer indicador de cómo puede comportarse la metapoblación



Tabla 1.6. Correlaciones entre el VGP (BLUP), el valor predicho genómico con el panel específico de raza (GVGP_r) y, el mismo valor, con el set de marcadores (GVGP_c) común a todas las razas para el set de animales genotipados y no genotipados. Las correlaciones se presentan en cada una de las razas para la componente directa (PD) y materna (LECHE).

Raza	Carácter	Genotipados			No genotipados		
		VGP_GVGP _r	VGP_GVGP _c	GVGP _r _GVGP _c	VGP_GVGP _r	VGP_GVGP _c	GVGP _r _GVGP _c
ANI	PD	0.908	0.905	1	0.983	0.982	1
	LECHE	0.919	0.917	1	0.981	0.980	1
AST	PD	0.944	0.947	0.994	0.998	0.998	1
	LECHE	0.844	0.908	0.950	0.991	0.994	0.998
MOR	PD	0.883	0.882	0.999	0.997	0.997	1
	LECHE	0.770	0.782	0.994	0.978	0.978	1
PAR	PD	0.875	0.875	0.999	0.981	0.981	1
	LECHE	0.904	0.903	0.999	0.983	0.983	1
PIR	PD	0.958	0.957	0.999	0.999	0.999	1
	LECHE	0.931	0.930	0.999	0.997	0.997	1
RET	PD	0.962	0.963	0.999	0.993	0.993	1
	LECHE	0.928	0.926	0.998	0.978	0.977	1
RUB	PD	0.963	0.962	0.999	0.994	0.994	1
	LECHE	0.943	0.941	0.998	0.987	0.987	1

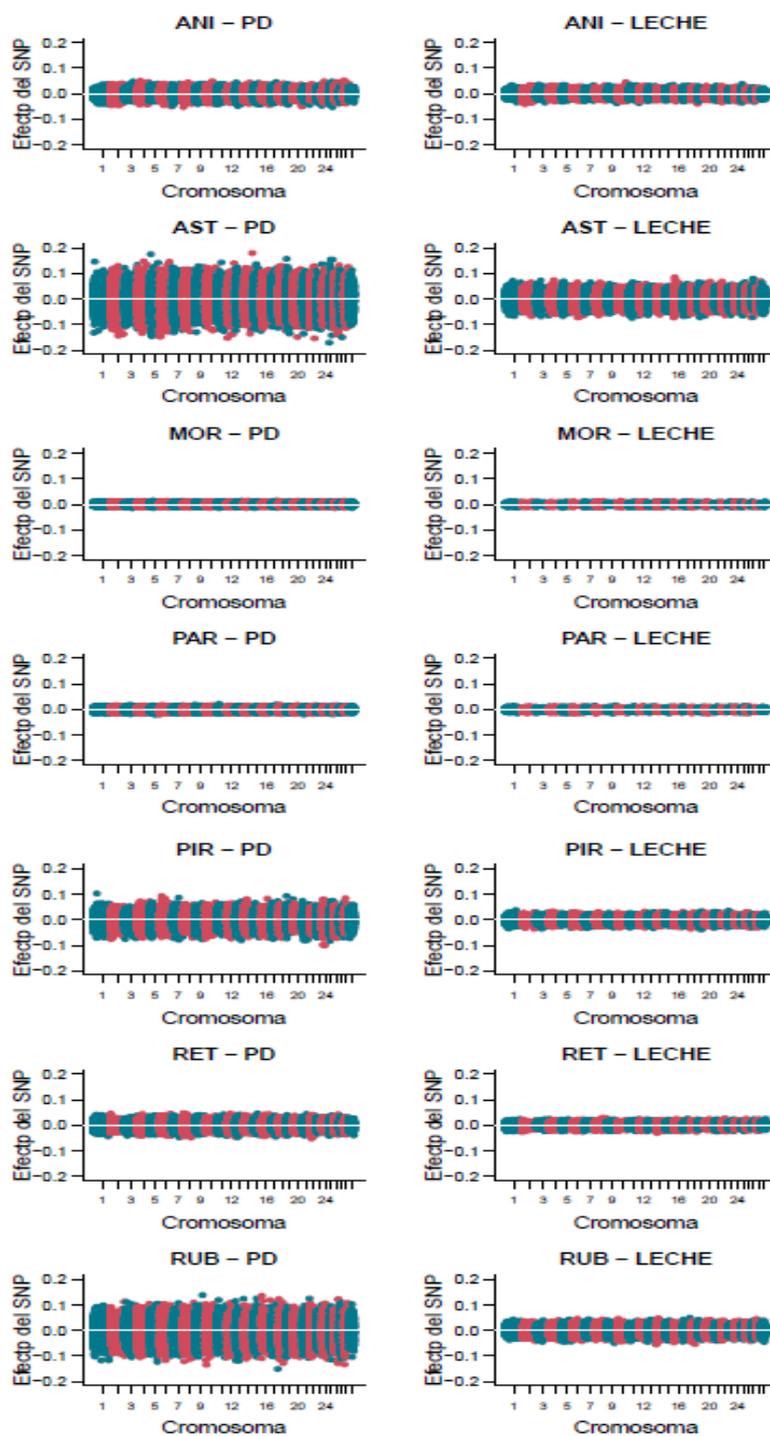


Figura 1.1 Efectos de los SNPs para la componente directa (PD) y materna (LECHE) de peso al destete en cada una de las razas



Tabla 1.7. Correlaciones de las soluciones de los SNPs entre razas para la componente directa (sobre la diagonal) y la componente materna (bajo la diagonal).

	ANI	AST	MOR	PAR	PIR	RET	RUB
ANI		-0,00849	0,00059	0,00530	0,00489	-0,01081	0,00994
AST	0,00073		-0,00123	-0,00366	0,00068	-0,00043	0,00735
MOR	0,00427	-0,00184		0,00759	-0,00303	0,00174	0,01552
PAR	-0,00366	-0,00273	-0,00212		-0,01846	0,00052	-0,00177
PIR	0,00614	-0,00350	-0,01044	0,00050		0,00183	-0,00086
RET	-0,00195	-0,00399	0,00806	0,00004	-0,00021		-0,01055
RUB	0,00841	-0,00769	-0,00176	-0,01567	0,00592	-0,01401	

Conclusiones

- En mayor o menor medida, la mayoría de las razas, han experimentado un incremento en la fiabilidad de ambas componentes a excepción de la AST, en la que sólo se ha observado en la componente PD, y en RET que no se ha observado. El incremento de fiabilidad ha sido mayor cuanto menor es la fiabilidad de los VGP de la población genotipada.
- No se observan diferencias entre los dos sets de SNPs utilizados, por lo que probablemente, si se pudiera disponer de un chip de menor densidad con un precio más reducido, podrían abaratare los costes de genotipado.
- La dispersión de las soluciones de los SNPs sigue el patrón esperado. Las correlaciones medias de dichas soluciones entre razas son cercanas a cero.
- Como ponen de manifiesto los informes de cada una de las razas, por el momento no existe una idea clara de cómo presentar los resultados de la evaluación genómica a los ganaderos de las razas autóctonas de aptitud cárnica (BAC). El mantenimiento de este grupo de trabajo podría ser una oportunidad para gestar herramientas homogéneas que pudieran ser utilizadas por todas las asociaciones.



2.- EVALUACIÓN DE LA METAPOBLACIÓN

Como se puso de manifiesto en la introducción de este informe, las BAC, tienen un tamaño limitado, con lo que constituir en cada una de ellas una población de referencia lo suficientemente grande como para estimar de forma precisa los efectos de los marcadores es difícil, sino imposible. Una solución teóricamente plausible sería generar una metapoblación a partir de la unión de todas ellas (Tootsi et al, 2010). Si en la población de referencia de esa metapoblación fuéramos capaces de capturar todos los haplotipos existentes en el conjunto de las poblaciones, sería viable ganar la precisión esperada. Según los trabajos de simulación desarrollados en el SeGenBeef esa solución podría funcionar en estas poblaciones (Mouresan et al., 2018) porque el parentesco medio entre ellas es elevado (González-Rodríguez et al., 2017) y la persistencia de fase se mantiene a distancias relativamente importantes (Cañas-Alvarez et al., 2016).

2.1 Procedimiento de trabajo

Se construyó un archivo de datos, de genealogía y de genotipos a partir de la unión de la información de cada una de las razas. La Tabla 1.3 presenta una relación de la cantidad de información por raza y del conjunto de las misma (META). El archivo de datos contenía 372.398 datos de peso al destete. El archivo de pedigrí resultante contenía 570.836 individuos y había un total de 6.452 animales genotipados. En set de SNPs empleado había 31.754 marcadores comunes a todas las razas.

Para el análisis de la metapoblación se empleó un ssBLUP con el software BLUPF90 (Misztal et al., 2018) bajo un modelo animal uni-carácter con efecto directo, materno y ambiental permanente materno. El modelo es similar al referido para cada una de las razas con dos grandes diferencias en cuanto a media y varianza; en este modelo se han incluido un efecto de raza que contempla las diferencias medias entre ellas, y la matriz de (co) varianzas es la correspondiente a la media ponderada por el número de datos de cada raza. El modelo contenía los siguientes efectos:

$$Y = GC_{raza} + Edad_destete_{raza} + Sexo + Edad_vaca + Raza + a_d + a_m + pe + e$$

Donde,

- Y= es el dato de peso al destete
- GC: el grupo de comparación definido como rebaño-año-época de nacimiento jerarquizado a raza. (43.074 niveles)
- Edad_destete (covariable): edad del animal al destete en días jerarquizado a raza (7 niveles)
- Sexo (2 niveles)
- Edad_vaca: edad de la madre al nacimiento del ternero, clasificado en 6 niveles en función de la edad en años: 1, 2, 3, 4, 5-9,>9



- Raza: que contempla las diferencias de nivel genético entre las razas (7 niveles)
- a_d : es el efecto aditivo directo (570.836 niveles)
- a_m : efecto aditivo materno (570.836 niveles)
- p : efecto permanente ambiental materno (p) (173.127 niveles)
- e : residuo (e)

Los valores esperados de a_d , a_m y gc para la componente PD y LECHE del peso al destete son **0**. La matriz de (co) varianzas de la metapoblación es la siguiente.

$$Var \begin{bmatrix} a_d \\ a_m \\ gc \\ p \\ e \end{bmatrix} \begin{bmatrix} \sigma_{a_d}^2 & \sigma_{a_d,a_m} & 0 & 0 & 0 \\ \sigma_{a_m,a_d} & \sigma_{a_m}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{gc}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_p^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_e^2 \end{bmatrix} = \begin{bmatrix} 589.64 & -228.16 & 0 & 0 & 0 \\ -228.16 & 214.16 & 0 & 0 & 0 \\ 0 & 0 & 358.19 & 0 & 0 \\ 0 & 0 & 0 & 60.53 & 0 \\ 0 & 0 & 0 & 0 & 655.86 \end{bmatrix}$$

Tabla 2.1 Estimaciones de las heredabilidades, aditiva directa (h^2_d), aditiva materna (h^2_m), correlación genética directa materna ($r_{a,m}$), cociente de varianza permanente y varianza fenotípica total (p) y cociente de la varianza de los grupos de comparación y varianza fenotípica total (gc) para la metapoblación.

Parámetros genéticos	Raza							META*
	ANI	AST	MOR	PAR	PIR	RET	RUB	
h^2_d	0.29	0.37	0.21	0.21	0.46	0.20	0.43	0.36
h^2_m	0.15	0.16	0.15	0.15	0.14	0.10	0.10	0.13
$r_{a,m}$	-0.72	-0.59	-0.40	-0.40	-0.67	-0.72	-0.76	-0.64
p	0.09	0.03	0.09	0.09	0.02	0.08	0.03	0.04
gc	-	0.16	-	-	0.24	0.32	0.22	0.22

*Media ponderada por el número de datos

La media ponderada de las heredabilidades, correlaciones y ratios de otras varianzas (gc y p) con la varianza fenotípica de la metapoblación, es mayor que algunas estimaciones intra-población dependiendo de la raza (Tabla 2.1). Este es el caso de ANI, MOR, PAR y RET para la componente directa, y el caso de RET y RUB para la componente materna. Este cambio en las estimaciones tiene un efecto en los resultados como veremos a continuación. Bajo el modelo empleado para la metapoblación, estamos asumiendo que todas las razas provienen de una misma población original que se ha subdividido a lo largo del tiempo con una varianza intra-población similar.



Este modelo no reconoce que podría existir una varianza entre subpoblaciones (Legarra et al., 2015), que puede ser estimada a partir del parentesco entre los meta-fundadores. Para acomodar el modelo con los meta-fundadores se requiere de un desarrollo teórico que será objeto de un trabajo posterior si las Asociaciones están dispuestas a ceder sus datos.

En paralelo al desarrollo de la evaluación de la metapoblación, se evaluó la diversidad genética entre estas poblaciones siguiendo el mismo procedimiento utilizado en Cañas-Alvarez et al. (2015). Para ello se obtuvieron los Componentes Principales de la matriz de parentesco genómico (“variance-standardized relationship matrix”) y se calculó la diversidad genética entre razas mediante el estadístico F_{st} (Weir y Cockerham, 1984) utilizando Plink1.9 (Purcell et al., 2007)

2.2 Resultados

En la Tabla 2.2 aparecen las fiabilidades medias de los individuos genotipados y no genotipados en cada raza. Estas fiabilidades medias se presentan para el BLUP convencional (VGP), el modelo en el que se utiliza el set de marcadores propios de raza (GVGP_r) y las obtenidas en evaluación de la metapoblación (GVGP_m) en el que se utiliza el set de marcadores común a todas las razas. Los resultados se presentan tanto para la componente directa (PD) como la componente materna (LECHE).

Tabla 2.2. Fiabilidad media del valor genético predicho (VGP), el valor genómico con el panel específico de raza (GVGP_r) y del valor genómico obtenido a partir de la metapoblación (GVGP_m) para el set de animales genotipados y no genotipados en cada una de las razas. Los resultados se presentan para la componente directa (PD) y materna (LECHE).

Raza	Componente	Genotipada			No genotipados		
		VGP	GVGP _r	GVGP _m	VGP	GVGP _r	GVGP _m
ANI	PD	50	54	68	48	49	60
	LECHE	52	54	62	48	49	52
AST	PD	70	71	73	59	59	54
	LECHE	61	61	63	49	49	45
MOR	PD	31	36	57	32	32	46
	LECHE	35	40	47	32	32	35
PAR	PD	29	30	52	24	24	35
	LECHE	28	30	40	25	25	27
PIR	PD	64	65	66	60	59	55
	LECHE	55	55	59	49	49	47
RET	PD	68	66	75	53	52	59
	LECHE	60	57	63	49	49	51
RUB	PD	78	80	80	60	60	57
	LECHE	69	71	72	51	51	51



En este caso, todas las razas experimentan un incremento de fiabilidad de los GVGP_m en la población genotipada, no así en la no genotipada, donde AST y RUB no experimentan ninguna ganancia. El efecto sobre la fiabilidad es tanto mayor, cuanto menor es la fiabilidad de los animales genotipados y no genotipados en cada una de las razas bajo el BLUP convencional. Por una parte, la metapoblación se beneficia de las diferencias de magnitud de las estimas de heredabilidad entre las razas bajo el modelo intra-raza frente al modelo de la metapoblación en ambos componentes. Por otra, este incremento de fiabilidad es debido a uso que hace el modelo de las relaciones de parentesco genómico existente entre estas razas (González-Rodríguez et al., 2017) lo que favorece que la cantidad efectiva de información mejore para todas ellas al menos en la población genotipada, que es la que realmente se puede beneficiar de esta información.

Tabla 2.3. Correlaciones entre las estimas de VGP (BLUP), del valor genómico con el panel específico de raza (GVGP_r) y del valor genómico obtenido a partir de la metapoblación (GVGP_m) para el set de animales genotipados y no genotipados en cada una de las razas. Los resultados se presentan para la componente directa (PD) y materna (LECHE).

Raza	Componente	Genotipada			No genotipados		
		VGP-GVGP _r	VGP-GVGP _m	GVGP _r -GVGP _m	VGP-GVGP _r	VGP-GVGP _m	GVGP _r -GVGP _m
ANI	PD	0.908	0.637	0.694	0.983	0.562	0.593
	LECHE	0.919	0.808	0.861	0.981	0.729	0.754
AST	PD	0.944	0.908	0.943	0.998	0.992	0.994
	LECHE	0.844	0.869	0.888	0.991	0.977	0.977
MOR	PD	0.883	0.543	0.598	0.997	0.763	0.763
	LECHE	0.770	0.593	0.595	0.978	0.698	0.713
PAR	PD	0.875	0.476	0.588	0.981	0.692	0.694
	LECHE	0.904	0.554	0.590	0.983	0.629	0.631
PIR	PD	0.958	0.905	0.945	0.999	0.979	0.981
	LECHE	0.931	0.848	0.914	0.997	0.971	0.972
RET	PD	0.962	0.907	0.945	0.993	0.887	0.893
	LECHE	0.928	0.834	0.908	0.978	0.865	0.876
RUB	PD	0.963	0.937	0.977	0.994	0.984	0.992
	LECHE	0.943	0.854	0.909	0.987	0.899	0.872

En la Tabla 2.3 se presentan las correlaciones entre los VGP (BLUP), el valor genómico con el panel específico de raza (GVGP_r) y las de las estimas de valor genómicos obtenidas en contexto de la metapoblación (GVGP_m). Las estimas de las correlaciones entre VGP y GVGP_r son altas o muy altas para todas las razas, como ya se ha visto en el capítulo anterior. Sin embargo, cuando centramos nuestra atención en las correlaciones entre los VGP y GVGP_m y los GVGP_r y GVGP_m las magnitudes de éstas se reducen. En el caso de los primeros, los valores oscilan entre 0.476 y 0.937 para la componente directa y entre 0.554 y 0.869 para la componente materna, en la población genotipada. En la población no genotipada las diferencias también



se aprecian y en mayor medida en las razas ANI, MOR, PAR y RET. En ambos casos, las razas que representan los límites de los rangos de las correlaciones mencionadas son PAR y RUB para la componente directa y PAR y AST para la componente materna. Cuando observamos las estimas de las correlaciones entre valores genómicos ($GVGP_r$ y $GVGP_m$) la situación es similar. Los resultados obtenidos parecen sugerir que, las relaciones de parentesco y la persistencia de los patrones de desequilibrio de ligamiento entre las razas favorece la presencia de haplotipos comunes que redundan en un incremento de las fiabilidades observadas. Sin embargo, la reducción en la correlación existente entre la predicción de los VGP, $GVGP_r$ y $GVGP_m$ sugieren que las fases de dichos haplotipos sufren cambios entre las distintas razas o que el modelo de la metapoblación introduce de alguna forma sesgos en la predicción que podrían ser corregidos quizás por la inclusión de los meta- fundadores (Legarra et al., 2015).

En la Tabla 2.4 se presentan las correlaciones de las soluciones de los SNPs intraraza con las obtenidas en la metapoblación. En general, se puede observar que cuanto mayor es la correlación entre VGP y $GVGP_m$ mayor es la correlación de las soluciones de los SNPs intraraza con las de la metapoblación. Esto ocurre así para los dos componentes directo y materno. En cualquier caso, las correlaciones de las soluciones de los SNPs de la metapoblación con las de las soluciones intra-raza siempre son de una mayor magnitud que las correlaciones entre razas.

Tabla 2.4. Correlaciones de las soluciones de los SNPs obtenidas con el modelo intra-raza y en la metapoblación (diagonal) para la componente directa (diagonal en negro) y materna (diagonal en verde). Los elementos de fuera de la diagonal son las correlaciones de las soluciones de los SNPs entre razas para la componente directa (sobre la diagonal) y la componente materna (debajo de la diagonal).

	ANI	AST	MOR	PAR	PIR	RET	RUB
ANI	0,21300 0,36366	-0,00849	0,00059	0,00530	0,00489	-0,01081	0,00994
AST	0,00073	0,48628 0,48029	-0,00123	-0,00366	0,00068	-0,00043	0,00735
MOR	0,00427	-0,00184	0,07909 0,08947	0,00759	-0,00303	0,00174	0,01552
PAR	-0,00366	-0,00273	-0,00212	0,15664 0,11091	-0,01846	0,00052	-0,00177
PIR	0,00614	-0,00350	-0,01044	0,00050	0,22714 0,26482	0,00183	-0,00086
RET	-0,00195	-0,00399	0,00806	0,00004	-0,00021	0,22783 0,25804	-0,01055
RUB	0,00841	-0,00769	-0,00176	-0,01567	0,00592	-0,01401	0,34186 0,35026



En la Tabla 2.5 se presenta las estimas de la diferenciación genética (F_{st}) entre las razas de la encomienda. El rango de valores mínimos oscila entre 0.034 entre ANI y MOR y entre AST y PAR, y 0.035 entre AST y RUB. Por el contrario, los valores máximos se presentan entre ANI y PIR, MOR y PIR, PAR y RET y PIR y RET, respectivamente. La raza con una divergencia media menor con el resto de las razas es la AST. El patrón es similar al encontrado en Cañas-Álvarez (2015) con una muestra de 25 tríos las razas participantes en esta encomienda, con excepción de la PAR que no fue incluida en dicho estudio. Por otro lado, la presencia de estos valores nos estaría indicando que buscar una estrategia para definir meta-fundadores que nos permita acomodar esa variabilidad entre razas, que ahora no contempla el modelo, podría mejorar el modelo de predicción (Legarra et al., 2015; Silva-Junqueira et al, 2020)

Tabla 2.5 Estimaciones de la diferenciación genética (F_{st}) entre las razas.

	AST	MOR	PAR	PIR	RET	RUB
ANI	0.050	0.034	0.055	0.075	0.054	0.065
AST		0.040	0.034	0.043	0.040	0.035
MOR			0.048	0.072	0.051	0.058
PAR				0.057	0.067	0.053
PIR					0.064	0.056
RET						0.053

En la Figura 2.2 se representa cada uno de los animales de cada una de las razas según las coordenadas obtenidas con el primer y segundo componente del análisis de componentes principales sobre la matriz de relaciones genómicas existentes entre los 6452 animales genotipados. En líneas generales se puede observar que los animales tienden a agruparse de acuerdo a su raza. Por otra parte, este análisis nos permite posicionar la raza PAR en el contexto de la diversidad existente entre las razas autóctonas de carne dado que ésta no había participado en el análisis que se había realizado previamente (Cañas-Alvarez et al., 2015). Por otro lado, se observa que las mayores distancias, como nos indicaban los F_{st} , se dan entre la raza PIR y ANI, MOR y PIR. Las razas que tienen un mayor número de animales parecen recoger una mayor dispersión, sin embargo, se observa una admixtura clara entre ANI y MOR; entre AST y PAR y entre AST y RUB que indicaría un flujo de genes entre estas poblaciones que, evidentemente, permitiría explicar esa ganancia de fiabilidad en el contexto de la metapoblación, puesto que estaría indicando un parentesco entre ellas que no puede ser rastreado a partir de la información de los Libros Genealógicos. También es interesante resaltar, como se aprecia, una subdivisión en la PAR que no se corresponde con la existencia de las dos Asociaciones. Está claro que hay una estratificación de los animales de acuerdo a las razas pero existe una sub-estratificación que habrá que estudiar para ver como modelizar la varianza entre ellas.

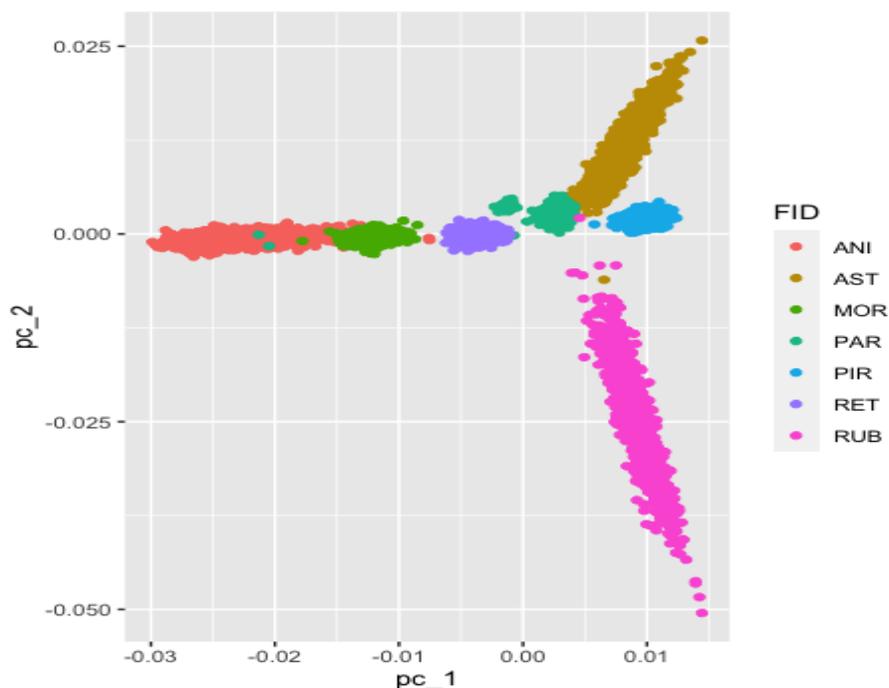


Figura 2.2 Agrupación de los individuos genotipados de acuerdo a los dos primeros componentes principales.

Conclusiones

- El modelo de evaluación genética en el contexto de la metapoblación contribuye al incremento de la fiabilidad de las predicciones de los $GVGP_m$ en todas las razas en la población genotipada. En la población no genotipada hay diferencias en función de la raza.
- La magnitud de las correlaciones entre VGPs y $GVGP_m$ y entre los $GVGP_r$ y $GVGP_m$ son menores que las correlaciones entre los VGPs y las predicciones intraraza lo que sugiere que las fases haplotípicas podrían cambiar de unas razas a otra, al menos en los haplotipos asociados a la componente de crecimiento y materna del peso al destete.
- Los resultados del PCA indican que los animales se pueden asignar a cada una de sus poblaciones. Sin embargo, se observa una admixtura entre ellas que permitiría explicar la ganancia de fiabilidad en la metapoblación.
- El modelo de la metapoblación asume que la variabilidad intra-raza es la misma en todas las poblaciones. Sin embargo, existe una diversidad intra-raza que podría ser modelizada para mejorar el modelo de la metapoblación.



- La evaluación genómica en el contexto de la metapoblación es un área de interés para muchos investigadores. En este sentido, merece la pena proseguir en esta línea de trabajo e ir definiendo un marco metodológico que pudiera favorecer la creación de una metapoblación con las BAC, si finalmente se definiera el interés de la misma para todas las razas.



3.- EVALUACIÓN GENERAL DEL AUDITOR DE ICAR

El método de trabajo de ICAR implica, que una vez los cuestionarios han sido revisados por los auditores, éstos deben analizar la situación *in situ*. Con el objetivo de aclarar todas las dudas además de constatar sobre el terreno que las respuestas se corresponden a la realidad de cada una de las asociaciones. Este procedimiento busca que a través del contacto directo, se genere un clima de intercambio, de manera que, el propio proceso se transforma en una herramienta generadora de ideas para el auditado. Las áreas de actividades para las cuales se ha solicitado este proceso son las que figuran a continuación:

- Identificación animal
- Control de rendimientos
- Control de las genealogías
- Evaluación de la conformación
- Procesado de datos
- Evaluación genética

La situación generada por la alarma sanitaria de la COVID 19 ha imposibilitado que el auditor designado por ICAR haya podido realizar las visitas previstas a las Asociaciones de ganaderos participantes en la encomienda: la *Asociación Nacional de Criadores de raza Retinta* (ACRE), la *Asociación Criadores de Rubia Gallega* (ACRUGA), la *Asociación Española de Criadores de Raza Avileña Negra-Ibérica* (AECRANI), la *Asociación de Criadores de raza Bovina Parda de Montaña* (ARAPARDA), la *Asociación Española de Asturiana de Valles* (ASEAVA), la *Confederación de Asociaciones de raza Pirenaica* (CONASPI), la *Federación Española de Raza Parda de Montaña* (FERPAM), la *Asociación Nacional de Criadores de Raza Morucha Selecta* (MORUCHA). Como alternativa, se organizaron entrevistas vía Zoom con cada una de ellas en las que participaron personal técnico de las mismas, el genetista correspondiente, la Secretaría de ICAR y el auditor. En el caso de bovino de carne esta persona ha sido el Dr. Japie van der Westhuizen del "South African Stud Book and Animal Improvement Association". Las impresiones y recomendaciones generales se pueden resumir en los puntos que figuran a continuación:

- La identificación animal está bien cubierta debido a la legislación nacional y europea.
- La impresión es que hay una amplia variación entre las Asociaciones en cuanto a:
 1. La estructura institucional y de gobernanza
 2. Sofisticación y procedimientos operativos estándar de computación y base de datos de la gestión.
 3. Aplicación y control de la recogida de datos productivos y reproductivos además de controles de plausibilidad.
 4. Capacidad humana y creación de redes o vínculos con los científicos.
 5. Un número adecuado de animales para una comparación justa como candidatos a la selección o prácticas de reposición ya sea por medio de definiciones y/o



- modelos contemporáneos adecuados aplicados en la predicción de méritos genéticos.
6. Prácticas de intercambio de datos y métodos disponibles.
 7. Las relaciones entre los criadores/propietarios y la información y/o los informes.
- Parece haber una falta de cohesión y unidad a nivel nacional. Se sugieren las siguientes recomendaciones generales:
 1. Un organismo o estructura general para el ganado vacuno que ayude a intercambiar ideas, conocimientos y experiencias científicas y técnicas, directrices generales e incluso a compartir recursos.
 2. Mecanismos conjuntos de transferencia de información y tecnología a los ganaderos e incluso a los técnicos.
 3. Establecimiento de directrices universales generales en cuanto a los procedimientos para la recogida de datos, y predicciones/evaluaciones de méritos genéticos.
 4. Algunos niveles de estandarización en los informes y otras publicaciones (definiciones de caracteres y otras) destinados a los ganaderos.
 5. Establecimiento de un enfoque estructurado para la aplicación de la próxima generación de tecnologías, como es el uso de información genómica en la predicción del mérito genético y las decisiones de selección.
 - Se necesitan predicciones de méritos genéticos (BLUP) más frecuentes para que los propietarios, criadores y/u otros actores de la industria puedan tomar decisiones de selección oportunas, también en la selección de animales en las ganaderías. Esto sólo puede ser posible cuando se comparten los recursos y se establece una unidad dedicada a este objetivo. El resultado de esas evaluaciones genéticas también debería ser comunicado a los ganaderos de forma didáctica e informativa.



4.- GENERACIÓN DE HERRAMIENTAS PARA LA IMPUTACIÓN DE MICROSATÉLITES PARA LA VERIFICACIÓN DE LAS PATERNIDADES A PARTIR DE LOS CHIPS DE SNPs

En mayor o menor medida, todas las razas involucradas en la encomienda vienen realizando de manera sistemática genotipados con marcadores tipo microsatélite para la verificación de sus genealogías. En estos momentos las Asociaciones están en una disyuntiva porque por un lado, tienen bases de datos de genotipados con micros con mucha información y una cantidad reducida de datos de genotipados con marcadores SNPs y, por otro, los padres y madres de los nuevos individuos genotipados disponen de genotipos tipo microsatélite y, sus descendientes son genotipados con los chips de SNPs. Este escenario implica que para hacer selección genómica es necesario que se continúe con el genotipado de individuos con los chips de SNPs y, para continuar con las labores de verificación, sería necesario obtener el genotipado de microsatélites de esos mismos individuos, lo que supone un incremento del gasto de genotipado, cuando realmente, por un coste algo superior al de genotipado de los micros, se puede obtener el de un chip de SNPs que, además, es mucho más informativo.

Con la idea de dar apoyo a las asociaciones para que continúen con el genotipado de los candidatos a la selección se propuso diseñar una herramienta informática que cubriese las necesidades de las asociaciones en este proceso. Para ello se planteó como objetivo establecer un procedimiento que nos permitiera imputar los genotipos tipo microsatélite a partir de los genotipos proporcionados por los SNPs. Para ello se inició este estudio partiendo de los genotipados de los tríos de seis de las razas de esta encomienda en HD (High Density) disponibles a través del proyecto SelGenBeef.

Objetivo

Elaborar una herramienta informática para imputar los alelos de los microsatélites a partir de los marcadores SNPs para la verificación del parentesco.

4.1 Procedimiento de trabajo

Material utilizado

La base de datos de individuos genotipados con Chips de Illumina de HD estaba constituida por 513 individuos agrupados en tríos (animal, padre y madre), disponiéndose de 25 tríos en el caso de las razas Asturiana de Valles (AST) Bruna dels Pirineus (BRU) y Morucha (MOR), y de 24 tríos, en las razas restantes (Avileña-Negra Ibérica (ANI), Pirenaica (PIR), Retinta (RET) y Rubia Gallega (RUB). De esta base de datos se extrajeron los genotipos de los animales de los que también se disponía de información de micros. Así finalmente se ha podido trabajar con 404 individuos de los cuales se disponía de ambas fuentes de información.



El procedimiento

- A. En primer lugar, se pidió a las asociaciones/genetistas que nos proporcionaran la identificación de los micros que utilizaba cada una de ellas para la verificación de paternidades y los alelos presentes en cada uno de ellos. En el Anejo 2 de esta memoria se muestran los 42 micros distintos, los alelos dentro de cada micro y la presencia o ausencia en las siete razas.
- B. Posteriormente, se posicionaron los micros en el nuevo mapa bovino ARS-UD1.2 mediante un alineamiento de las secuencias de los primers de cada uno de los marcadores sobre el nuevo genoma bovino (BLAST) y de cálculo de la posición central.
- C. Se creó un procedimiento “pipeline” que combina varias herramientas todas de uso libre: awk, VCFtools, Plink 1.9 (Purcell et al., 2007), Beagle 5.1 (Browning y Browning 2007), Tavix y R. Es un procedimiento que requiere tres pasos:
 - 1) **GENERACIÓN UN ARCHIVO DE REFERENCIA GLOBAL:** donde se posicionan los micros en relación a los SNPs y se codifican, además de buscar las fases.
 - 2) **IMPUTACION Y EVALUACION:** en esta parte del procedimiento se seleccionan los escenarios en función del tamaño de la ventana donde se hace la imputación además de la elección de la población de referencia, frente a la que se quiere imputar. Este proceso nos permite evaluar la bondad de la imputación de los micros a partir de los SNPs realizada. Obtenidos los haplotipos posibles el procedimiento evalúa el resultado de la imputación y realizar la asignación de los haplotipos a los alelos del micro en R. Si no es posible, vuelve al principio para incrementar el tamaño de la ventana de búsqueda.
Para encontrar todos los alelos se realiza un procedimiento recursivo donde la clave es realizar un proceso de muestreo de manera que, los alelos de un micro estén presentes tanto en la población de referencia, como en la población diana, donde están los individuos que “han perdido” la información de los micros que queremos imputar.
 - 3) **DETERMINACION DE LOS HAPLOTIPOS** por cada alelo de cada micro. Esta última fase es la que queda por finalizar.

La Figura 4.1 representa un diagrama de flujo que explica la sistemática del proceso.

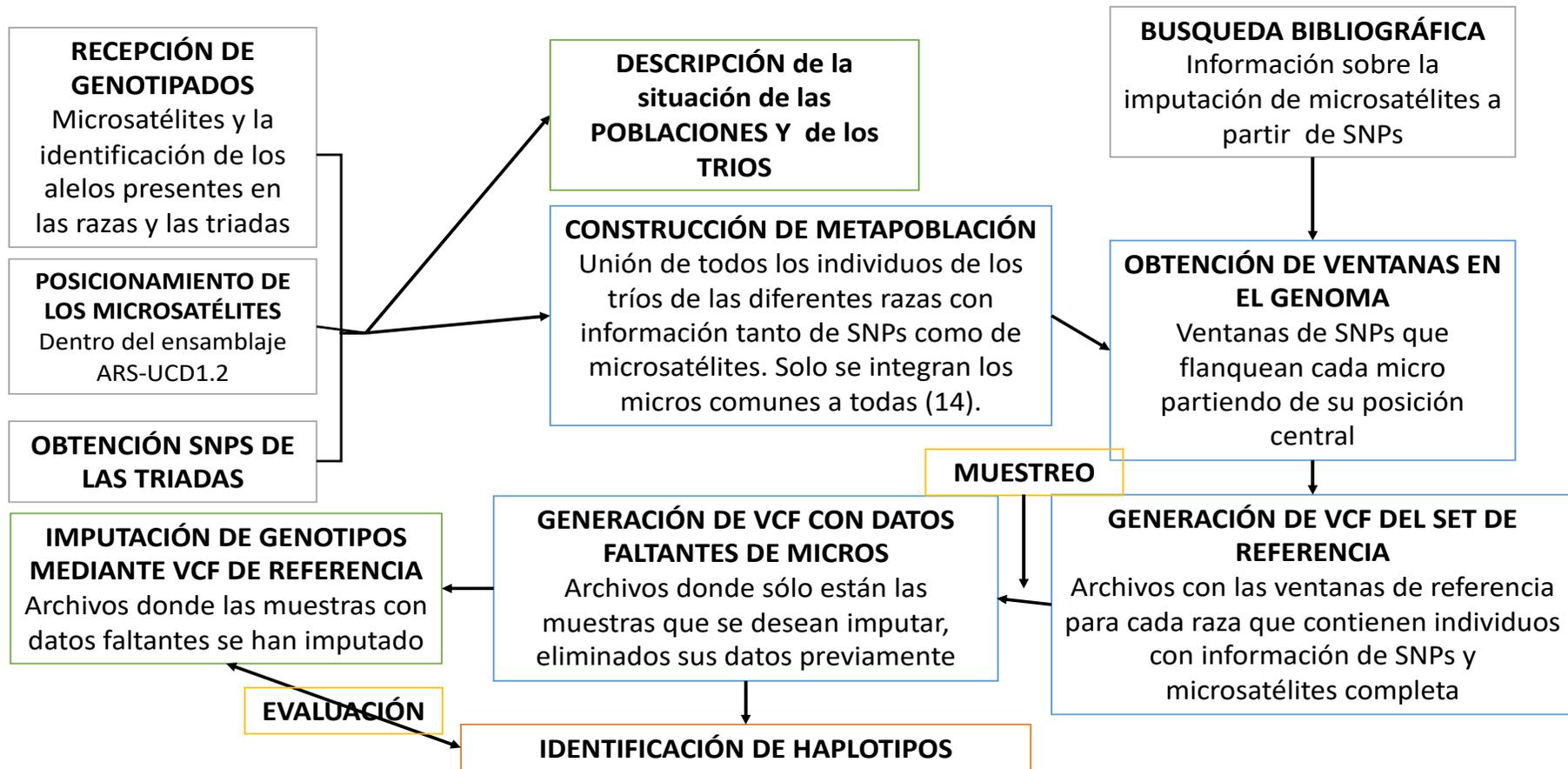


Figura 4.1 Diagrama de flujo del procedimiento de imputación



4.2 Resultados

En la Tabla 4.1 se presenta el número de micros utilizados en cada raza y el número de ellos que comparte con al menos otra raza. Este paso es esencial porque si no compartieran micros (altamente improbable) no tendría sentido utilizar datos conjuntos de varias razas. El número total de micros utilizados por raza oscila entre 17 en PIR y 34 en RET. En total se utilizan un total de 42 micros distintos, siendo 9 de ellos exclusivos de raza.

Tabla 4.1 Número de micros utilizados por cada uno de las razas y número de micros compartidos

Raza	Número de micros	Micros compartidos con al menos otra raza	Micros específicos de la raza
AST	24	21	3
ANI	21	21	0
MOR	21	21	0
PAR	19	19	0
PIR	17	17	0
RET	34	31	3
RUB	30	27	3
TOTAL	42	-	-

En la Tabla 4.2 podemos ver una información similar a la que se presenta en la Tabla 4.1 pero en relación a la información de los micros presentes en los tríos. Como se puede observar no todos los micros están presentes en los tríos, 34 de los 42 existentes en el global de las poblaciones. En total hay 15 marcadores microsatélites que son compartidos por todas las razas representadas en los tríos.

Tabla 4.2 Número de micros utilizados por cada una de las razas y número de micros compartidos presentes en los tríos.

Raza	Número de micros tríos	Micros compartidos con al menos otra raza en sus tríos	Micros específicos de los tríos de la raza
AST	24	20	4
ANI	17	17	0
MOR	17	17	0
PAR	-	-	-
PIR	17	17	0
RET	17	17	0
RUB	30	23	7
TOTAL	34	-	-



En la Tabla 4.3 se muestra el cromosoma y las posiciones inicial y final, así como el punto medio del mismo de los 15 marcadores comunes y presentes en los tríos del proyecto Selgenbeef. También aparecen el tamaño del microsatélite, el número de alelos distintos y una relación de dichos alelos.

Tabla 4.3 Cromosoma (Chr), identificación del micro (Id), posición inicial (Pos Inicial), posición final (Pos final) y punto central estimado, tamaño, número de alelos (Nº alelos) y alelos presentes en los 15 marcadores microsatélites comunes a las razas de los tríos. Las posiciones están dadas en el ensamblaje ARS_UCD1.2 de Bovino.

Chr	Id	Pos. inicial	Pos. final	Pos Central	Tamaño	Nº alelos	Alelos presentes
1	BM1824	131420021	131420162	131420091.5	141	6	178, 180, 182, 184, 188, 190
2	BM2113	126997684	126997782	126997733	98	12	121, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145
3	INRA023	32899374	32899543	32899458.5	169	11	198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218
5	ETH10	56326222	56326401	56326311.5	179	8	209, 213, 215, 217, 219, 221, 223, 225
7	ILSTS006	94303931	94304182	94304056.5	251	10	272, 284, 286, 288, 290, 292, 294, 296, 298, 300
9	ETH225	10621273	10621379	10621326	106	8	140, 142, 144, 146, 148, 150, 152, 156
10	CSRM60	75861828	75861882	75861855	54	8	92, 94, 96, 98, 100, 102, 104, 106
15	SPS115	23831625	23831827	23831726		9	248, 250, 252, 254, 256, 258, 260, 261, 262
16	TGLA53	25079863	25079975	25079919	112	19	154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190
17	ETH185	39852276	39852468	39852372	192	15	218, 220, 222, 224, 226, 228, 230, 232, 233, 234, 235, 236, 238, 240, 242
18	TGLA227	65229175	65229213	65229194	38	13	77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 103
19	ETH3	56013848	56013922	56013885	74	10	109, 115, 117, 119, 121, 123, 125, 127, 129, 131
20	TGLA126	22189045	22189114	22189079.5	69	8	111, 115, 117, 119, 121, 123, 125, 127
21	TGLA122	57053651	57053748	57053699.5	97	19	139, 141, 143, 147, 149, 151, 153, 155, 157, 159, 161, 163, 167, 169, 171, 173, 175, 177, 179
23	BM1818	39516726	39516950	39516838	224	9	258, 260, 262, 264, 266, 268, 270, 272, 278

Como podemos apreciar en la Tabla 4.3 los micros están posicionados en cromosomas distintos, son de un tamaño muy variable y con un grado de polimorfismo también muy



diferente. El número de alelos oscila entre 6 alelos del marcador BM1824 en el cromosoma 1 y los 19 de los marcadores TGLA53 y TGLA122 en los cromosomas 16 y 21, respectivamente. Este grado de polimorfismo, que junto la frecuencia de cada uno de los alelos, van a ser los factores más limitantes en el proceso de imputación (McClure et al., 2013, McClure et al., 2015, McClure et al., 2018).

En la Tabla 4.4 se muestran las probabilidades de imputación partiendo de ventanas desde el centro del microsatélite de 50, 1000 y 2500 pares de bases. La asignación se evalúa haciendo muestreo sin remplazo por lo que se incluye la desviación estándar. Podemos ver que con ventanas de 50 pares de bases la probabilidad de asignación correcta oscila entre 22% de ILTS006 y el 66% del SPS115. Incrementar hasta 1000 o 2500 no aporta prácticamente nada porque el número de haplotipos probables es muy alto y la determinación del haplotipo es prácticamente imposible en una situación como la muestra en la que partimos de un conjunto de datos de 403 individuos.

Tabla 4.4. Probabilidad de imputación correcta y la desviación estándar de esa probabilidad para tamaños de ventana de 50, 1000 y 2500 pares de bases.

Micro	50 bases		1000 bases		2500 bases	
	Probabilidad	Desviación estándar	Probabilidad	Desviación estándar	Probabilidad	Desviación estándar
BM1818	0,453	0,054	0,528	0,065	0,484	0,054
BM1824	0,293	0,043	0,422	0,064	0,415	0,065
BM2113	0,241	0,04	0,431	0,049	0,426	0,057
CSRM60	0,502	0,059	0,543	0,071	0,565	0,052
ETH10	0,338	0,041	0,536	0,071	0,548	0,054
ETH185	0,286	0,029	0,438	0,064	0,479	0,065
ETH225	0,335	0,084	0,466	0,08	0,474	0,062
ETH3	0,39	0,083	0,477	0,075	0,454	0,07
ILSTS006	0,216	0,049	0,364	0,058	0,39	0,05
INRA023	0,406	0,06	0,51	0,067	0,513	0,069
SPS115	0,659	0,043	0,662	0,041	0,631	0,047
TGLA122	0,238	0,047	0,431	0,065	0,467	0,076
TGLA126	0,557	0,048	0,625	0,051	0,607	0,054
TGLA227	0,247	0,068	0,458	0,045	0,439	0,048
TGLA53	0,254	0,048	0,398	0,033	0,387	0,038

El tamaño de ventanas estimado por McClure et al (2015), es de unos 76-77 pares de bases. Sin embargo, en su caso, la población de referencia consta de más de 8000 animales de 39 razas distintas. El procedimiento está establecido, aunque, requiere afinar la búsqueda e incrementar la población de referencia. De esta manera podremos tener todos los alelos de los micros presentes en la



población objetivo, población sobre la cual queremos inferir los micros. Este aspecto es muy limitante en nuestra población de referencia con una nuestra de 403 animales. El objetivo actual es trabajar con los animales genotipados en esta encomienda.

Conclusiones

- Se ha establecido un procedimiento que nos permite imputar los alelos de los micros para verificación de paternidades a partir de los genotipos de SNPs. El procedimiento es general, requiere de una base de datos de genotipos de SNPs, un archivo con sus posiciones, y la misma información referente a los microsatélites. Evidentemente, ambos tipos de marcadores deben tener las posiciones referenciadas en el mismo ensamblaje.
- El procedimiento se ha probado con los animales genotipados en el SelgenBeef que solo cuenta con 403 muestras. Tomando en cuenta la experiencia adquirida, hace falta incrementar el tamaño de la población de referencia por lo que se hará a partir de los genotipados obtenidos durante esta encomienda.



LITERATURA CITADA

- Browning, S.R. and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics* 81:1084-97. doi:10.1086/521987
- Cañas-Álvarez, J.J., González-Rodríguez, A., Munilla, S., Varona L., Díaz C., Baro JA, Altarriba J., Molina A., and J. Piedrafita. 2015. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip. *Journal of Animal Science* 2015.93:5164–5174. doi:10.2527/jas2015-9271
- Cañas-Álvarez, J.J., Mouresan, E.F., Varona, L., Díaz, C., Molina, A., Baro, J.A., Altarriba, J., Carabaño, M.J., Casellas, J., and J. Piedrafita. 2016. Linkage disequilibrium, persistence of phase, and effective population size in Spanish local beef cattle breeds assessed through a high-density single nucleotide polymorphism chip. *Journal of Animal Science* 2016.94:2779–2788. doi:10.2527/jas2016-0425
- González-Rodríguez, A., Munilla, S., Mouresan, E., Cañas-Álvarez, J., Baro, J., Molina, A., Díaz C., Piedrafita, J., Altarriba J., and L. Varona. 2017. Genomic differentiation between Asturiana de los Valles, Avileña-Negra Ibérica, Bruna dels Pirineus, Morucha, Pirenaica, Retinta and Rubia Gallega cattle breeds. *Animal*, 11(10), 1667-1679. doi:10.1017/S1751731117000398
- Kachman, S., Spangler, M., Bennet, G. et al., 2013. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics Selection Evolution*. 45:30. doi: 10.1186/1297-9686-45-30
- Legarra, A., Christensen, O., Aguilar, I., and I. Misztal. 2014. Single Step, a general approach for genomic selection. *Livestock Production Science*. 166:54-65.
- Legarra, A., Christensen, O. F., Vitezica, Z. G., Aguilar, I., and I. Misztal. 2015. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. *Genetics* 200, 455–468. doi: 10.1534/genetics.115.177014.
- Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control, *Bulletin of the Entomological Society of America* 15: 237-240.
- McClure, M. C., McCarthy, J., Flynn, J., McClure, J.C., Dair, E., O’Connell, D.K., and J.F. Kearney. 2018. SNP Data Quality Control in a National Beef and Dairy Cattle System and Highly Accurate SNP Based Parentage Verification and identification. *Frontiers in Genetics*. doi: 10.3389/fgene.2018.00084
- McClure, M. C., McCarthy, J., Flynn, J., Weld, R., Keane, M., O’Connell, K., et al. 2015. SNP selection for nationwide parentage verification and identification in beef and dairy cattle, In *Proceedings, International Committee For Animal Recording Technical Series*, June 2015, eds



- Z. Kowalski, N. Petreny, M. Burke, P. Bucek, L. Journaux, M. Coffey, C. Hunlun, and D. Radzio (Krakow; Rome: ICAR), 175–181.
- McClure, M. C., Sonstegard, T. S., Wiggans, G. R., Van Eenennaam, A. L., Weber, K. L., Penedo, M. C. T., et al. (2013). Imputation of microsatellite alleles from dense SNP genotypes for parentage verification across multiple *Bos taurus* and *Bos indicus* breeds. *Frontiers in Genetics*. 4:176. [10.3389/fgene.2013.00176](https://doi.org/10.3389/fgene.2013.00176)
 - Misztal, I., Tsuruta, S., Lourenco, D., Masuda, Y., Aguilar, I., Legarra, A., and Z. Vitezica. 2018. Manual for BLUPF90 family of programs.
 - Mouresan E.F, Cañas-Álvarez, J. J. , González-Rodríguez, A., Munilla, S., Altarriba, J, Díaz, C., Baró, J.A., Molina, A., Piedrafita, J., 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. doi: [10.1017/S175173111700283X](https://doi.org/10.1017/S175173111700283X)
 - Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and P.C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3): 559-575.
 - Toosi, A., Fernando, R.L., and J.C.M. Dekkers 2010. Genomic selection in admixed and crossbred populations. *Journal of Animal Science* 88, 32–46. doi.org/10.2527/jas.2009-1975
 - Silva Junqueira, V., Lopes, P.S., Lourenco, D., Fonseca e Silva, F., and F. Flores-Cardoso. 2020. Applying the Metafounders Approach for Genomic Evaluation in a Multibreed Beef Cattle Population. *Frontiers in Genetics*, 11: 556399. doi: [10.3389/fgene.2020.556399](https://doi.org/10.3389/fgene.2020.556399).