
Trazabilidad genética en ganado bovino: estudio comparativo de la
eficacia de microsatélites y SNPs.

Arianne Sanz Fernández

Tesis Doctoral
Universidad de Zaragoza

<http://zaguan.unizar.es>

TDR-UZ [Tesis Doctorales en Red Universidad de Zaragoza]



Biblioteca
Universitaria

Universidad Zaragoza

LABORATORIO DE GENÉTICA BIOQUÍMICA (LAGENBIO)



Dpto. Anatomía, Embriología y Genética Animal
Facultad de Veterinaria, Universidad de Zaragoza



**“TRAZABILIDAD GENÉTICA EN GANADO
BOVINO: ESTUDIO COMPARATIVO DE LA
EFICACIA DE MICROSATÉLITES Y SNPs.”**

Memoria presentada por

ARIANNE SANZ FERNÁNDEZ

para optar al grado de Doctora por la Universidad de Zaragoza

Zaragoza, Marzo de 2010



Departamento de Anatomía, Embriología y Genética Animal

Facultad de Veterinaria

Universidad de Zaragoza

Miguel Servet, 177. 50013 Zaragoza

Tfnos: 976 761620 / 976 761000 (Ext. 4203) Fax: 976 762949



Doña Clementina Rodellar Penella, Profesora Titular, y Doña Pilar Zaragoza Fernández, Catedrática, del Departamento de Anatomía, Embriología y Genética Animal de la Facultad de Veterinaria de la Universidad de Zaragoza,

CERTIFICAN QUE:

La tesis titulada “Trazabilidad genética en ganado bovino: Estudio comparativo de la eficacia de Microsatélites y SNPs” que se recoge en la presente Memoria, de la que es autora Arianne Sanz Fernández, ha sido realizada bajo nuestra dirección y se corresponde con el proyecto de tesis aprobado por la Comisión de Doctorado en Septiembre de 2006, cumpliendo por tanto todos los requisitos exigidos para que su autora pueda optar al Grado de Doctora por la Universidad de Zaragoza.

Lo que suscribimos como directoras del trabajo en Zaragoza a 15 de Febrero de 2010.

Fdo.: Clementina Rodellar Penella

Fdo.: Pilar Zaragoza Fernández

A mis madres: mami y abue.

A mis hombres: Ariel y Daniel.

AGRADECIMIENTOS

AGRADECIMIENTOS

Me gustaría poder mostrar mi agradecimiento en estas páginas a todas y cada una de las personas que han contribuido al desarrollo de este trabajo, y que han estado conmigo en este largo camino que empecé a recorrer el día que dejé de vivir en mi isla.

A Clemen Rodellar por dirigirme en este trabajo, por sus oportunos consejos, claros y concisos.

A Pilar Zaragoza por abrirme las puertas y permitirme formar parte de este grupo. Por su ayuda y por la dirección de este trabajo.

A Ignacio de Blas por hacerme siempre hueco en su apretada agenda y por su activa participación en el desarrollo de este proyecto. Gracias Nacho!

Inma, muchas gracias por tu ayuda, tu comprensión, por compartir tus conocimientos de biodiversidad conmigo, siempre dispuesta a hacerme sitio en el coche para poder llegar a tiempo al cole... en fin.

A Don Isaias Zarazaga que es ejemplo de abnegación y dedicación a la ciencia.

A Charo Osta porque si llegué aquí fue gracias a ella.

A Juan Altarriba por su contribución en esta tesis.

A todas las personas e instituciones que han contribuido a que este trabajo pudiese realizarse: Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA). Asociación Aragonesa de criadores de la raza Parda de Montaña (ARAPARDA). Asociación Aragonesa de ganaderos de raza Pirenaica (ASAPI).

A Joaquín Medina del Servicio de Audiovisuales y a Rodolfo Casellas por su colaboración en el diseño de la portada.

A los compañeros del Departamento de Bioquímica por su disposición a ayudar siempre.

A Odalys Uffo y Atzel Acosta mis colegas de la tierra, que un día casi fuimos mayoría hablando en cubano.

A mis compañeros del laboratorio, con los que comparto el día a día:

Carmi, gracias por estar cerquita.. ¿solo esto?.. lo demás prefiero decírtelo. Carmen, gracias por tu ayuda siempre, por tu cariño y por animarme a descubrir el mundo de la alta cocina. Laura Ordovás que hemos compartido mucho tiempo el espacio vital y siempre has estado dispuesta a echar una mano, aun hoy desde Bélgica. Ana C, Jaber, Bea, Raquel, Fernandinho, Janne, Ana Solanas, Ana Rosa, Sara y Amaya.

A los que hoy andan por otros lugares pero que siempre se recuerdan: David, Rosa, Jesús, Manolito, María, Saray, Ane, Alex, Ainhoa, Salvador y Sonia.

A Laura Bericat que gracias al trabajo nos conocimos y tantos buenos momentos hemos compartido. Gracias a su familia: Juampa, Maricarmen, Javier, la abuela... por hacerme sentir desde el primer día como en casa.

A mi mamá, que aquí o allá, está siempre a mi lado para protegerme.

A mi abuela que desde la distancia no se separa de mí.

A tí Arie, por estar siempre ahí.

A mi Dani, mi bebé que me quita tantas horas de sueño pero me hace sonreír cada día y da los mejores quieros del mundo.

A mi familia que siempre son un gran apoyo, a mis amigos en todas las latitudes imaginables...

...en fin a todos los que me quieren, la mejor de mis sonrisas como muestra de mi infinita gratitud. Muchas gracias!!!!

ÍNDICE

ÍNDICE	
1. RESUMEN	1
2. JUSTIFICACIÓN Y OBJETIVOS	7
2.1. Justificación	9
2.2. Objetivos	10
3. REVISIÓN BIBLIOGRÁFICA	13
3.1. Trazabilidad	16
3.1.1. Conceptos y definiciones	16
3.1.2. Legislación sobre trazabilidad	19
3.2. Trazabilidad convencional	22
3.2.1. Elementos del Sistema de identificación en ganado bovino	23
3.2.2. Etiquetado de la carne de vacuno	28
3.3. Trazabilidad genética	28
3.3.1. Identificación genética	29
3.3.2. Marcadores moleculares aplicados a la identificación individual y trazabilidad	30
3.4. Marcadores moleculares utilizados en el presente trabajo	32
3.4.1. Microsatélites o STR (Short Tandem Repeat)	32
3.4.2. SNPs: Polimorfismos de una base (Single Nucleotide Polymorphism)	35
3.4.2.1. SNPs localizados en regiones exónicas	38
3.4.2.1.1. SNP en el gen de la transferrina	38
3.4.2.1.2. SNP en el gen de la calpaína	40
3.4.2.1.3. SNP en el gen de la Proteína Transportadora de Ácidos Grasos 1 (FATP1)	41
3.4.2.1.4. SNP en el gen de la Sintasa de Ácidos Grasos (FASN)	41
3.5. Aplicación de la identificación individual a la trazabilidad	42
3.5.1. Comparación de perfiles	43
3.5.2. Asignación individual a raza o población	46
4. MATERIAL Y MÉTODOS	49
4.1. Material biológico	51

4.1.1. Animales	51
4.1.2. Muestras	52
4.1.2.1. Muestras de sangre y carne para el estudio de trazabilidad.....	52
4.1.2.2. Muestras de ADN de diferentes razas españolas para cálculos de frecuencias de SNPs	53
4.1.2.3. Muestras de ADN para estudios de asociación de polimorfismos	53
4.2. Extracción de ADN y cuantificación	54
4.2.1. Extracción	54
4.2.1.1. A partir de sangre.....	55
4.2.1.2. A partir de músculo.....	55
4.2.2. Cuantificación	56
4.3. Análisis de marcadores moleculares.....	56
4.3.1. Marcadores moleculares microsatélites	56
4.3.1.1. Amplificación de los fragmentos mediante PCR	57
4.3.1.2. Visualización de los fragmentos y lectura de los genotipos.....	59
4.3.2. Polimorfismos de una base (SNP: Single Nucleotide Polymorphism)	60
4.3.2.1. Elección de los SNPs.....	60
4.3.2.1.1. Búsqueda en bases de datos y bibliografía existente.....	60
4.3.2.1.2. Amplificación de los fragmentos que contienen los SNP.....	60
4.3.2.1.3. Purificación de los productos amplificados	61
4.3.2.1.4. Secuenciación automática.....	62
4.3.2.2. Tipificación de los SNPs	62
4.3.2.2.1. Minisequenciación o SNaPshot	63
4.3.2.2.1.1. Reacciones de amplificación	63
4.3.2.2.1.2. Purificación de los productos	65
4.3.2.2.1.3. Reacción de Minisequenciación o SNaPshot.....	65
4.3.2.2.1.4. Purificación de los productos de	

minisequenciación	68
4.3.2.2.1.5. Genotipado	68
4.3.2.2.2. PCR alelo específica (ASO-PCR: Allele Specific Oligonucleotide-PCR)	68
4.3.2.2.2.1. PCR alelo específica estándar	69
4.3.2.2.2.2. PCR alelo específica en RT-PCR.....	69
4.4. Herramientas y metodología informáticas aplicadas	70
4.4.1. Herramientas informáticas	70
4.4.2. Aplicación web diseñada para gestión de los datos	70
4.4.3. Caracterización de los marcadores microsatélites y SNPs	72
4.4.3.1. Análisis del potencial de los diferentes marcadores en identificación y trazabilidad	72
4.4.3.1.1. Probabilidad de exclusión	72
4.4.3.1.2. Probabilidad de identidad.....	74
4.4.3.2. Análisis estadísticos	74
4.4.3.2.1. Correlación	75
4.4.3.2.2. Regresión.....	75
4.4.3.2.3. Coeficientes de determinación.....	75
4.4.3.2.4 Selección de variables en regresión múltiple.....	76
4.4.4. Asignación de individuos	76
5. RESULTADOS Y DISCUSIÓN.....	79
5.1. Diseño de herramientas informáticas para identificación de individuos y trazabilidad	81
5.1.1. Gestión de datos.....	82
5.1.2. Análisis de filiación.....	83
5.1.3. Análisis de parámetros genéticos	85
5.1.4. Realización de búsquedas específicas.....	85
5.2. Estimación de la eficacia de los marcadores utilizados en identificación y trazabilidad	86
5.2.1. Microsatélites	86
5.2.1.1. Estimación de la variabilidad genética	86
5.2.1.2. Probabilidades de exclusión	91
5.2.1.3. Probabilidades de identidad teóricas y empíricas	93

5.2.1.3.1. Probabilidad de identidad teórica (PI _t)	93
5.2.1.3.2. Probabilidad de identidad empírica (PI _e).....	96
5.2.1.4. Estimación de parámetros genéticos en la descendencia a partir de parámetros genéticos de los progenitores utilizando microsatélites	100
5.2.1.4.1. Análisis de regresión para las probabilidades de identidad empíricas de los descendientes	102
5.2.1.4.1.1. Comparación entre las PI teóricas y las PI estimadas por regresión en los descendientes	105
5.2.1.4.2. Análisis de regresión para las heterocigosidades de los descendientes	108
5.2.1.4.2.1. Comparación entre las heterocigosidades esperadas y estimadas de los descendientes	110
5.2.2. SNPs	113
5.2.2.1 Estimación de la variabilidad genética	113
5.2.2.1.1. Desequilibrio H-W en SNP del cromosoma 26	119
5.2.2.1.2. Análisis de variabilidad en SNP del cromosoma 1 ..	126
5.2.2.2. Probabilidades de exclusión	131
5.2.2.3. Probabilidades de identidad teóricas y empíricas	133
5.2.2.4. Estimación de parámetros genéticos en la descendencia a partir de parámetros genéticos de los progenitores utilizando SNPs	139
5.2.2.4.1. Análisis de regresión para las probabilidades de identidad de los descendientes	141
5.2.2.4.1.1. Comparación entre las PI teóricas, empíricas y estimadas en los descendientes.....	145
5.2.2.4.2. Análisis de regresión para las heterocigosidades de los descendientes	148
5.2.2.4.2.1. Comparación entre las heterocigosidades esperadas, observadas y estimadas en los descendientes ...	150
5.3. Asignación de individuos a población como método complementario a trazabilidad: Microsatélites y SNPs	153
5.4. Comparación de la aplicabilidad a los sistemas de	

trazabilidad de ambos tipos de marcadores moleculares:	
Microsatélites y SNPs.....	158
5.4.1. Comparación de parámetros: PI y PE	158
5.4.2. Costes económicos.....	161
5.4.3. Estimación de parámetros de los descendientes a partir de los progenitores	163
5.4.4. Comparación de la asignación	165
5.4.5. Verificación de la correspondencia carne-sangre de animal de procedencia	166
5.5. Estudio piloto de trazabilidad	168
6. CONCLUSIONES	171
7. BIBLIOGRAFÍA	175

1. RESUMEN

1. RESUMEN

La crisis de confianza de los consumidores desatada en los últimos años por la aparición de distintas enfermedades transmisibles al hombre como la Encefalopatía Espongiforme Bovina (EEB), ha supuesto una creciente sensibilización de todos los sectores en los temas relacionados con la seguridad alimentaria. En este sentido la trazabilidad se ha convertido en una herramienta fundamental al servicio de la calidad alimentaria.

En la presente memoria se propone el establecimiento de un sistema de trazabilidad genética basado en marcadores de ADN. Se ha realizado un estudio comparativo de la eficiencia entre microsatélites y SNPs.

En el estudio se han utilizado un total de 675 muestras de animales procedentes de dos razas bovinas del norte de Aragón: Parda de Montaña y Pirenaica. El muestreo se llevó a cabo en el período comprendido entre el año 2004 y 2005; en una primera fase se tomaron muestras de sangre de los reproductores y descendientes, y posteriormente se tomaron muestras de carne de animales descendientes que fueron sacrificados en matadero. Toda la información obtenida se ha almacenado en una base de datos que permite, conjuntamente con un programa diseñado, gestionar y verificar el sistema de trazabilidad.

Se han estimado las frecuencias alélicas de un gran número de SNPs en diferentes razas españolas, eligiéndose para el estudio aquellos SNPs que presentaban frecuencias intermedias.

Los marcadores de ADN utilizados han demostrado su utilidad en identificación, paternidad y trazabilidad con poderes de exclusión superiores al 99% para microsatélites y superiores al 96% para SNPs. De igual modo se han obtenido valores de probabilidades de identidad del orden de 10^{-9} para microsatélites y de 10^{-7} para SNPs, por lo que las probabilidades de encontrar individuos al azar que compartan los mismos genotipos son 1:100 millones para microsatélites y 1:1 millón para SNPs. Además, y como método

complementario de trazabilidad ha sido demostrada la aplicabilidad de estos paneles en la asignación de individuos a raza o población.

Diferentes parámetros genéticos (probabilidad de identidad y heterocigosidad) han sido calculados tanto en progenitores como en descendientes. Se ha detectado una clara correlación entre las variables de ambos grupos de animales y mediante un análisis de regresión se han obtenido una serie de modelos matemáticos que permiten predecir con una elevada fiabilidad el comportamiento de las variables en los descendientes a partir de la información de los progenitores.

En relación a los SNPs incluidos en el panel utilizado en el presente trabajo, se ha demostrado que el SNP g.329C>T (AJ496781) ubicado en el cromosoma 26, se encuentra en una región duplicada, hecho que podría justificar las frecuencias intermedias detectadas por otros autores. Se recomienda la exclusión de este marcador o, cualquier otro con similares características, de cualquier panel de marcadores genéticos destinado a estudios de identificación, trazabilidad y asociación a determinadas patologías tanto en las especies animales como en la especie humana.

Los paneles de SNPs presentan una ventaja frente a los microsatélites ya que pueden incluirse SNPs de regiones codificantes que sirvan simultáneamente para identificación genética y para establecer posibles asociaciones con caracteres productivos. En el presente estudio se ha analizado el SNP c.1455A>G (U02564) del exón 12 de la transferrina bovina que supone un cambio aminoacídico Asp/Gly, que nos ha permitido asociar el gen de la transferrina a cantidad de grasa en la leche.

SUMMARY

The consumers' lack of confidence after the Bovine Spongiform Encephalopathy (BSE) outbreak has supposed an increasing awareness on issues related to food safety. In this sense, the development of a traceability system has become necessary to ensure food safety.

In this work we proposed the establishment of a genetic traceability system based on DNA markers: microsatellites and SNPs. A comparative study of the efficiency of both methods has been performed.

The study was carried out using a total of 675 samples belonging to animals from "Parda de Montaña" and "Pirenaica" breeds. Sampling was conducted in the period between 2004 and 2005. In the earlier phase blood samples were taken from ancestors and descendants. Later, meat samples were taken from offspring at the slaughterhouse. All the information data was stored in a database linked to software that allows, to manage and to verify the traceability system.

The allelic frequencies of a large number of SNPs in different Spanish cattle breeds have been estimated. Those showing intermediate frequencies were selected.

The DNA markers used in this work have demonstrated their usefulness in identification, paternity testing and traceability with power of exclusion over 99 % for microsatellites and greater than 96 % for SNPs. Match probability values have been obtained in the order of 10^{-9} for microsatellites and of 10^{-7} for SNPs. This implies that the chances of finding random individuals sharing identical allelic profile are 1:100 million for microsatellites and 1:1 million for SNPs. Moreover, and as complementary traceability method it has been demonstrated the applicability of these panels in the assignment of individuals to breeds.

Some genetic parameters (match probability and heterozygosity) have been calculated in parents and offspring. A correlation has been detected

between variables in both groups of animals. Mathematical models were obtained by regression analysis to predict, with high reliability, the offspring's variables using the parental information.

For the SNPs included in the panel used in this work, it has been demonstrated that the g.329C>T SNP (AJ496781) on chromosome 26 is located in a duplicated region. This fact might justify the intermediate frequencies detected by other authors. We recommend discarding this SNP, or any other with similar characteristics, of any molecular marker panel to be used in animal identification, traceability and animal pathology studies.

SNPs represent an advantage over the microsatellites since SNPs located in coding regions could be used, simultaneously, for genetic identification and to determine possible associations with quantitative traits. In the present study the c.1455A>G SNP (U02564) identified at exon 12 of bovine transferrin gene has been associated with milk-fat yield.

2. JUSTIFICACIÓN Y OBJETIVOS

2. JUSTIFICACIÓN Y OBJETIVOS

2.1. Justificación

En la actualidad no se puede hablar de calidad en alimentación sin hablar de seguridad alimentaria, la que se ha convertido en los últimos años en una de las mayores preocupaciones de los consumidores en los países desarrollados. A ello han contribuido de forma importante diversas crisis relacionadas con la producción animal en general y la producción de carne de una forma específica. Dentro de todos los riesgos que preocupan en este momento a los consumidores, ninguno de ellos es comparable con la crisis de confianza, a todos los niveles, desatada por la aparición de la Encefalopatía Espongiforme Bovina (EEB), impulsando así la necesidad de garantizar el origen de los animales y de sus productos y establecer un sistema fiable de seguimiento desde que el animal nace hasta que su carne es comercializada.

Un sistema seguro de identificación individual, que asegure en todo momento el seguimiento del animal vivo o sus productos, no solo resulta imprescindible para asegurar la calidad sanitaria de los productos, sino también como un sistema de garantizar la calidad de los productos por su pertenencia a determinadas “marcas de calidad” o “denominaciones específicas”.

La Unión Europea (UE) mediante la Regulación (EC) No 1760/2000, adoptada en Julio de 2000 establece la obligatoriedad de trazar el origen del animal y de establecer un sistema de trazabilidad integral para la carne de vacuno. Además de todos los sistemas de identificación y registro presentes en la reglamentación comunitaria y la derivada en cada país miembro de la Unión Europea, existe un sistema de identificación de animales y productos basado en el ADN contenido en los mismos.

La identificación individual mediante el ADN ofrece un instrumento potente para controlar y autenticar los sistemas de identificación tradicionales (crotales, tatuajes, sistemas electrónicos...), puesto que la

información genética contenida por un individuo es inalterable durante toda su vida, está presente en cada parte de su organismo y es única en cada sujeto. Esto hace que el sistema posea una objetividad, repetibilidad y seguridad no alcanzable con ningún otro sistema de identificación actual.

Actualmente los métodos genéticos de identificación genética individual se basan en la utilización de marcadores microsatélites aunque en los últimos años y debido principalmente a las posibilidades de automatización de la tipificación de SNPs, estos últimos se están proponiendo como posibles marcadores de elección a aplicar en un futuro próximo. A este respecto la Sociedad Internacional de Genética Animal (ISAG), aunque sigue recomendando para la identificación genética los microsatélites, consciente de la evolución del desarrollo de metodologías para la tipificación de SNPs, pretende promover en un futuro próximo el desarrollo de un panel de SNPs común y estandarizado internacionalmente. Para ello resulta imprescindible el análisis de distintos SNPs en diferentes razas, por lo que el trabajo desarrollado en la presente memoria puede contribuir, y de hecho está contribuyendo, al desarrollo de dicho panel.

Los perfiles de ADN obtenidos para cada individuo, utilizando microsatélites o SNPs, pueden garantizar la trazabilidad de la carne hasta el animal de origen, ya que van a permitir discriminar si dos muestras son la misma o son muestras distintas.

2.2. Objetivos

Dada la importancia de la trazabilidad de la carne de bovino, el objetivo general de esta memoria es el establecimiento de un sistema más seguro de identificación individual mediante el empleo de dos tipos de marcadores moleculares, microsatélites y SNPs, que posibilite la verificación del sistema establecido por la UE.

Para alcanzar este objetivo general nos proponemos los siguientes objetivos específicos:

1. Control de todos los reproductores y descendientes de dos explotaciones de razas Parda de Montaña y Pirenaica del Pirineo Aragonés, en el período comprendido entre los años 2004 y 2005.
2. Creación de una base de datos y software asociado que permita tanto el almacenamiento de datos como el seguimiento de los animales y de las muestras.
3. Genotipado de todas las muestras obtenidas mediante marcadores microsatélites y SNPs.
4. Cálculo de frecuencias alélicas de un conjunto de SNPs en diferentes razas españolas.
5. Elección de un panel de SNPs adecuado a razas españolas basado en sus características: frecuencias alélicas, facilidad de tipificación, posibilidad de asociación a caracteres productivos.
6. Cálculo de diferentes parámetros genéticos con los dos tipos de marcadores: porcentaje de exclusión y probabilidad de identidad.
7. Establecimiento de un sistema de trazabilidad basado en el genotipado de progenitores.
8. Utilización de la asignación a raza como método complementario al sistema de trazabilidad.
9. Comparación de la eficacia de los dos tipos de marcadores utilizados.

3. REVISIÓN BIBLIOGRÁFICA

3. REVISIÓN BIBLIOGRÁFICA

La carne de vacuno es un alimento básico en toda dieta equilibrada ya que nos aporta proteínas y vitaminas esenciales, además de representar una fuente importante de minerales, como el hierro, que desempeñan un papel importante en nuestro organismo.

Las crisis alimentarias ocurridas en la Unión Europea (UE), como son el anuncio de la relación entre la Encefalopatía Espongiforme Bovina (EEB) y una nueva variante de la enfermedad de Creutzfeldt Jacob (Smith *et al.* 2004), la contaminación de alimentos animales con dioxinas (Puntaric *et al.* 2000), los brotes de lengua azul (Mertens *et al.* 2005) o la influenza aviar (Peiris *et al.* 2007), han llevado a una disminución en el consumo de carne particularmente en Europa. Esto ha hecho reaccionar a los gobiernos y a la industria cárnica y trabajar con el objetivo de restablecer la confianza de los consumidores (Boenke 2002; Brambilla & De Filippis 2005; Verbeke *et al.* 2007). Para conseguirlo, los poderes públicos han considerado la necesidad de poder localizar el origen de los alimentos, tanto los destinados directamente al consumo como aquellos dirigidos a la alimentación animal. Las consecuencias derivadas de las crisis en los diferentes sectores implicados plantearon la necesidad de proteger el funcionamiento adecuado del mercado para evitar sobresaltos como el ocurrido en el año 2000, donde se reportó una caída de la producción de vacuno del 70% a raíz de la aparición de la Encefalopatía Espongiforme Bovina, hundiendo el sector de la carne de vacuno en aquel momento. Para restablecer la confianza del consumidor, resultaba fundamental controlar la procedencia de la carne y mostrársela al cliente, ofrecerle los datos sobre su origen y alimentación controlados, así como presentarle los datos de registros de la ganadería, del matadero, etc.

Por otra parte, el creciente interés en las dos últimas décadas en Europa por los productos con identidad local o regional (Negrini *et al.* 2008), impulsó también la creación de un marco legal de protección de alimentos regionales mediante las identificaciones geográficas protegidas (IGP) o las

denominaciones de origen (D.O). Estos sistemas contribuyen al incremento de la confianza de los consumidores, ya que proporcionan la información de procedencia, las características específicas del producto y la protección de estos contra fraudes o imitaciones (Ilbery & Kneafsey 2000).

En la Unión Europea, se han promulgado textos legislativos conducentes a aplicar mecanismos de rastreo que permitan conferir mayor seguridad a la sanidad animal, la salud pública, el comercio internacional y la certificación (Vallat 2001). Así el Libro Blanco de la Seguridad Alimentaria creado como medida comunitaria en enero de 2000 estableció como objetivo principal garantizar un alto grado de seguridad alimentaria y la necesidad de recuperar la confianza del consumidor alterada tras las “crisis alimentarias”. Además, consideraba como un aspecto esencial de la seguridad alimentaria la necesidad de implantar sistemas o procedimientos de rastreabilidad de los alimentos “desde la granja a la mesa”, mediante el establecimiento de un sistema de trazabilidad o rastreabilidad que permitiera mantener el control de origen de los productos y la identidad de los animales desde el criador hasta la venta al pormenor (McKean 2001), así como asegurar el vínculo entre el animal y la carne que proviene de él (Ammendrup & Fussel 2001; Caporale *et al.* 2001).

Aunque un sistema de trazabilidad no puede por sí mismo garantizar la seguridad alimentaria ni la calidad de los productos, resulta una herramienta indispensable para poder conseguir ambas cualidades.

3.1. Trazabilidad

3.1.1. Conceptos y definiciones

El término “trazabilidad” proviene de la traducción literal del inglés “traceability” y aunque la Real Academia de la Lengua Española (RAE) en su “Diccionario panhispánico de dudas” (2005) lo acepta como sinónimo de “rastrear”, desaconseja su uso porque dice que trazabilidad deriva en castellano del verbo “trazar” que no tiene significado análogo al verbo inglés “to trace” que significa “rastrear o seguir la pista”. Por tanto, según la RAE

debería utilizarse el término *rastreabilidad*. Sin embargo, en publicaciones en castellano aparecen ambos términos y aunque no esté oficialmente admitido como tal, se está utilizando el término trazabilidad con mucha más frecuencia que el correspondiente término castellano, es por ello que en esta tesis utilizaremos el término trazabilidad.

Existen diferentes definiciones de trazabilidad, algunas de las cuales serán comentadas a continuación.

La comisión del Codex Alimentarius, creada en 1963 por la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO) y la Organización Mundial de la Salud (OMS) para desarrollar normas y reglamentos alimentarios, define trazabilidad como la capacidad para seguir el movimiento de un alimento a través de etapa(s) especificada(s) de la producción, transformación y distribución.

Para la Organización Internacional de Estandarización (*International Organization for Standardization, ISO*), cuya finalidad es la coordinación de las normas nacionales en consonancia con el Acta Final de la Organización Mundial del Comercio (OMC), trazabilidad es: “la capacidad de rastrear la historia, uso o posición de un artículo o actividad mediante los registros de identificación. Cuando se consideren productos, la trazabilidad puede relacionarse con el origen de materiales y partes, la historia de tratamiento, y la distribución y la posición del producto después de la entrega”.

En este sentido han sido editadas una familia de normas conocidas como la familia de normas ISO 22000 que recoge un conjunto de aspectos necesarios para la consecución de una correcta gestión de la seguridad alimentaria. En ese marco ISO 22005 proporciona un modelo sencillo y ordenado para establecer un sistema de trazabilidad en las empresas alimentarias en base a los conceptos del Codex Alimentarius, por eso ha sido editada recientemente la norma ISO 22005:2007 – “Traceability in the feed and food chain - General principles and basic requirements for system

design and implementations”. Esta norma establece los principios y requerimientos para el diseño e implementación de sistemas de trazabilidad

Por su parte el Reglamento (CE) n° 178/2002, del Parlamento Europeo y del Consejo, de 28 de enero de 2002 (http://eur-lex.europa.eu/pri/es/oj/dat/2002/1_031/1_03120020201es00010024.pdf) por el que se establecen los principios y los requisitos generales de la legislación alimentaria, crean la Autoridad Europea de Seguridad Alimentaria en la cual se fijan procedimientos relativos a la seguridad alimentaria, acoge el concepto de trazabilidad como: “la posibilidad de encontrar y seguir el rastro, a través de todas las etapas de producción, transformación y distribución, de un alimento, un pienso, un animal destinado a la producción de alimentos o una sustancia destinados a ser incorporados en alimentos o piensos o con probabilidad de serlo”.

Todas estas definiciones de trazabilidad nos permiten considerarla como una herramienta de gestión, que lleva consigo la identificación de cualquier producto dentro de la cadena alimentaria conectando materias primas con procesos tecnológicos de transformación. De esta manera se podrá autenticar de forma objetiva el origen de un producto; localizar el origen de la causa de un riesgo manifiesto en la cadena alimentaria de manera rápida y eficaz pudiéndose proceder a retiradas y recuperaciones específicas y precisas de productos. Además, facilitar a los consumidores y a las empresas alimentarias información apropiada y que las autoridades de control puedan llevar a cabo determinaciones de riesgo, evitándose perturbaciones innecesarias del comercio.

Esto permitirá una respuesta efectiva y por tanto una mayor seguridad alimentaria, garantizando el acceso a productos inocuos y de calidad a los consumidores, que tienen derecho a la protección, derecho específicamente reconocido por la Constitución Española y la Ley General de Defensa de los Consumidores y Usuarios.

En conclusión podemos decir que el concepto de trazabilidad se encuentra estrechamente vinculado a garantía, seguridad y calidad, respaldado evidentemente por un marco legal que comentaremos a continuación.

3.1.2. Legislación sobre trazabilidad

Aunque el concepto de trazabilidad no es un concepto nuevo en el campo de la seguridad alimentaria, es la primera vez que existe un texto legal comunitario que lo impone explícitamente a distintos sectores de la cadena alimentaria. En dicho texto (Reglamento 178/2002) se especifica claramente el objetivo y los resultados previstos con la aplicación de los sistemas de trazabilidad, pero sin concretar la forma de hacerlo. Este planteamiento deja a la industria y a los demás sectores un mayor margen de flexibilidad en la aplicación del requisito de trazabilidad.

Desde el punto de vista de trazabilidad los artículos de mayor interés dentro de este Reglamento son el 18 y 19, ya que introducen la obligatoriedad de la trazabilidad y establecen las siguientes obligaciones:

1. En todas las etapas de la producción, la transformación y la distribución deberá asegurarse la trazabilidad de los alimentos, los piensos, los animales destinados a la producción de alimentos y de cualquier otra sustancia destinada a ser incorporada en un alimento o un pienso, o con probabilidad de serlo.

2. Los explotadores de empresas alimentarias y de empresas de piensos deberán poder identificar a cualquier persona que les haya suministrado un alimento, un pienso, un animal destinado a la producción de alimentos, o cualquier sustancia destinada a ser incorporada en un alimento o un pienso, o con probabilidad de serlo. Para tal fin, dichos explotadores pondrán en práctica sistemas y procedimientos que permitan poner esta información a disposición de las autoridades competentes si estas así lo solicitan.

3. Los explotadores de empresas alimentarias y de empresas de piensos deberán poner en práctica sistemas y procedimientos para identificar a las empresas a las que hayan suministrado sus productos. Pondrán esta información a disposición de las autoridades competentes si estas así lo solicitan.

4. Los alimentos o los piensos comercializados o con probabilidad de comercializarse en la Comunidad Europea deberán estar adecuadamente etiquetados o identificados para facilitar su trazabilidad mediante documentación o información pertinente, de acuerdo con los requisitos exigidos en disposiciones más específicas.

El cumplimiento del artículo 18 de este Reglamento puede alcanzarse según diversas metodologías, desde procedimientos simples, basados en sistemas manuales (identificación manual y archivo de los registros asociados), hasta la utilización de herramientas más complejas que aprovechan las nuevas tecnologías y los sistemas de información (códigos de barra para la transferencia de información, identificación por radiofrecuencia, etc.). En esencia, un sistema de trazabilidad no deja de ser una herramienta de comunicación que pone a nuestro alcance información relacionada con la cadena alimentaria, datos que pueden ser utilizados con diversos objetivos.

Por lo que en caso de producirse un fallo de seguridad alimentaria se garantizaría la retirada del producto; ya que en todos los casos e independientemente del procedimiento empleado para asegurar la trazabilidad, podrá conocerse la historia completa del producto alimentario desde su origen hasta el consumo y viceversa.

Aunque en el Reglamento 178/2002 no se especifique cómo hacerlo, parece claro que para establecer un sistema que garantice una relación entre la identificación de la carne y/o del producto de que se trate y el animal o los animales correspondientes a través de todas las etapas de producción,

transformación y distribución, es necesario conservar registros adecuados de todos los eslabones de la cadena como son:

- Registro de explotaciones
- Identificación de los animales
- Registro de los movimientos de los animales
- Controles en los mataderos
- Etiquetado de los productos

Existe además otra legislación horizontal relacionada:

Real Decreto 1808/1991 (http://www.boe.es/g/es/bases_datos/doc.php?coleccion=iberlex&id=1991/30678) por el que se regulan las menciones o marcas que permiten identificar el lote al que pertenece un producto alimenticio.

Real Decreto 2207/1995 (http://www.boe.es/g/es/bases_datos/doc.php?coleccion=iberlex&id=1996/04449&txtlen=1000) sobre normas de higiene relativas a los productos alimenticios.

Real Decreto 1334/1999 (<http://www.boe.es/boe/dias/1999/08/24/pdfs/A31410-31418.pdf>) por el que se aprueba la Norma general de etiquetado, presentación y publicidad de los productos alimenticios.

Real Decreto 1801/2003 (<http://www.boe.es/boe/dias/2004-01-10/pdfs/A00906-00916.pdf>) sobre seguridad general de los productos.

Además de esta legislación de carácter horizontal, existe normativa específica de diferentes sectores en la que ya se considera el concepto de Trazabilidad como es el caso del sector cárnico.

Reglamento 1760/2000 (http://europa.eu.int/eurllex/pri/es/oj/dat/2000/l_204/l_20420000811es00010010.pdf) que establece sistemas de identificación y registro de los animales de la especie bovina y relativo al etiquetado de la carne de vacuno y productos a base de carne de vacuno.

Reglamento 1825/2000 (http://europa.eu.int/eurlex/pri/es/oj/dat/2000/l_216/l_21620000826es00080012.pdf) por el que se establecen disposiciones de aplicación del Reglamento 1760/2000.

Real Decreto 1376/2003 (http://www.boe.es/boe_gallego/dias/2003-11-24/pdfs/A02413-02420.pdf) por el que se establecen las condiciones sanitarias de producción, almacenamiento y comercialización de las carnes frescas y sus derivados en establecimientos de comercio al por menor.

Real Decreto 1698/2003 (<http://www.boe.es/boe/dias/2003-12-20/pdfs/A45345-45350.pdf>) por el que se establecen disposiciones de aplicación de los Reglamentos comunitarios sobre el sistema de etiquetado de la carne de vacuno.

3.2. Trazabilidad convencional

En la actualidad el mercado de la carne de vacuno es de los más avanzados en la implantación de sistemas de trazabilidad. Para aumentar la confianza de los consumidores, mantener un alto nivel de protección de la salud y reforzar la estabilidad sostenible del mercado de carne de vacuno, se elaboró el Reglamento (CE) 1760/2000 del Parlamento Europeo y del Consejo de Europa de 17 de julio de 2000, que establece un sistema de identificación y registro de los animales de la especie bovina y regula el etiquetado de la carne de vacuno y de los productos a base de carne de vacuno.

Este reglamento en su Título I *“Identificación y registro de los animales de la especie bovina”* incluye en su artículo 3 los elementos que forman parte de este sistema, que son:

- Marcas auriculares, destinadas a identificar individualmente a cada animal
- Bases de datos informatizadas
- Pasaporte para los animales
- Registros individuales llevados en cada explotación

En su Título II “*Etiquetado de la carne de vacuno y de los productos a base de carne de vacuno*” en el artículo 11 establece que los agentes económicos u organizaciones están obligados a etiquetar la carne en todas las fases de la comercialización. Esto garantizará una relación entre la identificación de las canales, cuartos o trozos de carne y el animal o el grupo de animales correspondientes. La etiqueta debe contener un número de referencia o código de referencia que garantice la relación entre la carne y el animal o los animales. Dicho número podrá ser el número de identificación del animal del que proceda la carne de vacuno o el número de identificación correspondiente a un grupo de animales, el número de autorización del matadero en el que haya sido sacrificado el animal o grupo de animales, el Estado miembro o tercer país en el que se encuentre el matadero y finalmente el número de autorización de la sala de despiece en la que haya sido despiezada la canal o el grupo de canales junto con el Estado miembro o tercer país en el que se encuentre la sala de despiece.

3.2.1. Elementos del Sistema de identificación en ganado bovino

Uno de los principales requisitos de la trazabilidad de animales y sus subproductos es la implementación de un sistema de identificación animal. Actualmente existen una serie de tecnologías en aplicación o desarrollo de identificación animal o sus productos. Dentro de los métodos clásicos, se pueden destacar el marcaje de la cola, tatuajes, marcas de fuego, marcas de frío, crotales o caravanas, que pueden ser de metal o plástico y los chips electrónicos. También existen métodos biométricos tales como la huella nasal, imágenes digitales de retina e iris y la huella genética (Felmer *et al.* 2006). De todos estos métodos nos centraremos en el empleo de marcas auriculares o crotales por ser las que aparecen recomendadas en el Reglamento (CE) 1760/2000 y las aplicables mayoritariamente en España.

1. Marcas auriculares

El Reglamento (CE) 1760/2000 establece un sistema de identificación y registro de los animales de la especie bovina, que supone que los animales

nacidos después de Enero de 1998, o que después de esa fecha se destinen al comercio intracomunitario, deben estar identificados mediante dos marcas auriculares (crotales). Estas marcas son autorizadas por la autoridad competente y se colocan en cada oreja del animal, tienen el mismo y único código de identificación y son de plástico anaranjado con un código de identificación individual de 14 caracteres que incluye el código del país, un dígito de control, dígito de la comunidad autónoma y el número de identificación del animal (Figura 1):



Figura 1. Marcas auriculares utilizadas en España para identificación individual de animales de la especie bovina.

2. Documento de Identificación de Bovinos (DIB)

Se establece como obligatorio un Documento de Identificación de Bovinos (Figura 2) que debe ser individual para cada animal, expedido dentro de los 14 días posteriores a la notificación del nacimiento, o desde la reidentificación en el caso de animales importados. Se expide también cuando los animales proceden de otro país de la UE, ya que el DIB debe acompañar al animal cuando sea trasladado. En caso de muerte debe ser restituido a la autoridad competente en el plazo de 7 días y en caso de animales exportados, también será restituido a la autoridad que lo expide.

Para los animales nacidos a partir de enero de 2000, los Documentos de Identificación tienen un nuevo formato, con más datos y constarán, al menos, de dos ejemplares, uno que acompañará al animal cuando abandone la explotación y otro que permanecerá siempre en poder del ganadero o, en

su caso, de la autoridad competente y que será la base para solicitar todas las ayudas al sector bovino. Estos documentos son expedidos por la autoridad competente y contienen los datos del animal (número de identificación, fecha de nacimiento, sexo, raza, origen de la madre, explotación de nacimiento y si ha nacido en España); los datos de la explotación (código, titular y la fecha de incorporación del animal a la explotación); datos de la muerte, sacrificio o exportación a país no perteneciente a la U.E; los datos sobre primas, así como todos sus movimientos, donde se incluyen la explotación de origen, de destino y la fecha. Al contener los datos personalizados del propietario del animal y de la explotación en la que reside, en caso de cambio de explotación, el nuevo propietario debe solicitar un nuevo documento.



ESPAÑA

DOCUMENTO DE IDENTIFICACION PARA BOVINOS
ejemplar 1 de acompañamiento de animal

Fecha de expedición: 11/02/2005
Firma o sello de la autoridad competente

NUMERO DE IDENTIFICACION 

Reservado para anotar el identificador de los toros de lidia

DATOS DEL ANIMAL			
FECHA DE NACIMIENTO	SEXO: MACHO	RAZA	NACIDO EN ESPAÑA
Código de la Madre	EXPLORACIÓN DE NACIMIENTO		

DATOS DE LA EXPLORACION	
Código 	FECHA DE INCORPORACION A LA EXPLORACION TITULAR DN /CIF

DATOS DE LA MUERTE, SACRIFICIO O EXPORTACION A PAIS NO PERTENECIENTE A LA U.E.			
ANIMAL	Muerto en explotación <input type="checkbox"/>	Sacrificado en Matadero <input type="checkbox"/>	Exportado a otro país <input checked="" type="checkbox"/>
	Día <input type="text"/> <input type="text"/> Mes <input type="text"/> <input type="text"/> Año <input type="text"/> <input type="text"/> <input type="text"/>	FIRMA O SELLO	

DATOS SOBRE PRIMAS	
SOLICITADA PRIMA ESPECIAL	
PRIMER TRAMO DE EDAD <input type="checkbox"/>	Fecha ____ / ____ / ____
SÉGUNDO TRAMO DE EDAD <input type="checkbox"/>	Fecha ____ / ____ / ____


ES 01 12 0294 6706 01 *2 1305 01 11 13

MOVIMIENTOS DEL BOVINO

FECHA	EXPLORACION DE ORIGEN	EXPLORACION DE DESTINO

Figura 2: Documento de Identificación de Bovinos (DIB).

3. Libros de registro de las explotaciones

Se llevarán a cabo los registros individuales en cada explotación ganadera y se notificarán a la autoridad competente los nacimientos y muertes de animales, así como la entrada o salida de animales de la explotación. El registro podrá ser manual o informatizado, en un formato aprobado por la autoridad competente, ser accesible y conservarlo al menos durante tres años.

4. Bases de datos informatizadas

La información referente a la explotación y a los animales deberá estar en una base de datos informatizada que permita, a partir del código de explotación, conocer el censo en una fecha concreta o dado el número de crotal del animal, conocer todo el historial de sus movimientos desde el momento del nacimiento.

El Real Decreto 1980/1998 (http://www.boe.es/g/es/bases_datos/doc.php?coleccion=iberlex&id=1998/23140) por el que se establece un Sistema de Identificación y registro de los animales de la especie bovina, dispone en su artículo 12, que el Ministerio de Agricultura, Pesca y Alimentación y las Comunidades Autónomas, en sus respectivos ámbitos competenciales, constituirán una base de datos informatizada de conformidad a lo establecido en la Directiva 97/12/CE (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31997L0012:ES:HTML>) relativa a problemas de política sanitaria en materia de intercambios intracomunitarios de animales de la especie bovina y porcina.

Asimismo, el Real Decreto 1716/2000 (<http://www.boe.es/boe/dias/2000/10/25/pdf/A36737-36753.pdf>) sobre normas sanitarias para el intercambio intracomunitario de animales de las especies bovina y porcina dispone que las Comunidades Autónomas implantarán un sistema de redes de vigilancia epidemiológica con el doble objetivo de facilitar los intercambios intracomunitarios de los animales de la especie bovina y porcina, y recoger

datos epidemiológicos que permitan la vigilancia de las enfermedades y el control de su movimiento.

Estas disposiciones constituyen la base legal del Sistema Nacional de Identificación y Registro de los Movimientos de los Bovinos (SIMOGAN) que constituye una base de datos de ámbito nacional controlada por la Administración Central, en la que quedan registradas todas las explotaciones bovinas existentes en España, incluidos los mataderos, todos los animales bovinos y sus movimientos. Permite conocer en un momento dado los datos individuales de todos los animales presentes en cualquier explotación de España, así como, para un animal individual dado, todas las explotaciones por las que ha pasado a lo largo de su vida, desde su nacimiento o importación, hasta su sacrificio o exportación. Recientemente se ha creado el Sistema integral de trazabilidad animal (SITRAN), que es un sistema multiespecie que incorpora las bases de datos para animales vivos SIMOGAN y SIMOPORC que muy pronto dejarán de estar operativas.

En la Comunidad Autónoma de Aragón toda la información está contenida en la base de datos SITRAN, que es el Sistema Integrado de Trazabilidad Animal que contiene la identificación de todas las explotaciones de interés ganadero, sus titulares, datos de animales, movimientos, etc. y está gestionado por la Dirección General de Alimentación. Este sistema agrupa:

- RIIA: Registro de Identificación Individual de Animales. Para aquellas especies para las que la legislación exige identificación individual; contiene toda la información referida a cada animal. (RD 728/2007 <http://www.boe.es/boe/dias/2007/06/29/pdfs/A28189-28200.pdf>).
- REGA: Registro de Explotaciones Ganaderas multiespecie. Contiene los datos de las explotaciones, titulares, ubicación, etc. (RD 479/2004 <http://www.boe.es/boe/dias/2004/04/13/pdfs/A14978-14983.pdf>).
- REMO: Registro de Movimientos de animales (por ejemplo bovino) o de lotes de animales (como en el caso del porcino) (RD 728/2007).

La Comunidad Autónoma de Aragón posee la información de su territorio, mantiene su propio servidor y los datos se actualizan a través de la aplicación desarrollada.

3.2.2. Etiquetado de la carne de vacuno

Según el Reglamento (CE) 1760 la definición de etiquetado es la aplicación de una etiqueta a uno o varios trozos de carne o a su material envasado, o en el caso de productos no preenvasados, el suministro de información por escrito y de manera visible en el punto de venta. Por su parte el Real Decreto 1698/2003 de 12 de diciembre establece las disposiciones de aplicación de los reglamentos comunitarios sobre el sistema de etiquetado de la carne de vacuno (Reglamento CE 1760/2000). Se establecen las menciones obligatorias que deben de figurar en el etiquetado de la carne para asegurar su trazabilidad: número o código de referencia que garantice la relación entre la carne y el animal o los animales, nombre del país de nacimiento, nombre del país de engorde, nombre del país de sacrificio, número de autorización sanitaria del matadero y EM o Tercer.

3.3. Trazabilidad genética

Hasta el momento se han explicado en que consisten los métodos implementados tradicionalmente para el establecimiento de sistemas de trazabilidad que contribuyan a aumentar la seguridad de los consumidores, así como la legislación que respalda este tipo de procedimientos. Sin embargo, estos métodos convencionales de identificación y verificación, como son los clásicos crotales o los registros utilizados por décadas, están sujetos a riesgos de pérdida accidental de información en algunos puntos de la cadena de producción (Capoferri *et al.* 2005). Por tanto se hace necesario el empleo de otros métodos más fiables que permitan identificar el animal o el origen del producto de forma infalible, y que sean seguros contra fraudes permitiéndonos verificar *a posteriori* el origen de la carne (Arana *et al.* 2002), ya que con los métodos tradicionales los fraudes no pueden ser completamente evitados (Orru *et al.* 2006).

En este sentido la identificación individual mediante los marcadores moleculares del ADN ofrece un instrumento potente y seguro para verificar los sistemas convencionales de identificación como crotales, tatuajes, sistemas electrónicos, entre otros. Así como para autentificar la procedencia de determinadas marcas de calidad como son las indicaciones geográficas protegidas (IGP), denominaciones de origen (DO) y especialidades tradicionales garantizadas, ofreciendo máximas garantías e información objetiva a los consumidores y agentes económicos del sector dedicados a la comercialización de los alimentos (Viana *et al.* 2005)

3.3.1. Identificación genética

El principio de identificación genética animal se basa en el hecho de que, a excepción de los gemelos monocigóticos y de los clones, cada animal es portador de una combinación única de bases nitrogenadas en el ácido desoxirribonucleico (ADN) (Cunningham & Meghen 2001). Además esta información genética es única en cada sujeto, inalterable durante toda su vida y está presente en cada parte de su organismo, en cualquier célula nucleada. Esto hace que el sistema posea una objetividad, repetibilidad y seguridad no alcanzable con ningún otro sistema de identificación actual.

Si tenemos en cuenta que el genoma de cada animal contiene aproximadamente 3×10^9 pares de bases (pb), las posibilidades que existen de variación entre los distintos individuos son enormes (Griffiths *et al.* 2000). Esta variación genética se genera y mantiene fundamentalmente por tres procesos:

- ✓ las mutaciones que suponen cambios de bases o bien adición o pérdida de algunas de ellas,
- ✓ la recombinación genética que permite el intercambio de fragmentos de cromosomas homólogos,
- ✓ la combinación al azar durante la meiosis de cromosomas de origen paterno y materno en cada generación en las especies de reproducción sexual.

Todos estos fenómenos son la base que posibilita, estudiando determinadas señales genéticas, la diferenciación de individuos, aún estando muy emparentados, y que sea prácticamente imposible la existencia de individuos genéticamente idénticos (siempre exceptuando a gemelos monocigóticos y clones).

3.3.2. Marcadores moleculares aplicados a la identificación individual y trazabilidad

Los marcadores moleculares son fragmentos de ADN que sirven de referencia para seguir la transmisión de un segmento de cromosoma de una generación a otra, siendo, en un sentido restringido, una entidad genética que manifiesta polimorfismo y se hereda de forma mendeliana. Su uso ha demostrado ser una herramienta muy útil en la identificación individual (Glowatzki-Mullis *et al.* 1995; Usha *et al.* 1995; Heyen *et al.* 1997; Pakstis *et al.* 2007) y por tanto puede ser utilizada en identificación de piezas de carne, permitiendo la verificación de sistemas de trazabilidad. (Meghen *et al.* 1998; Cunningham 2000; Portelle *et al.* 2000; Sancristobal-Gaudy *et al.* 2000; Cunningham & Meghen 2001; Shackell *et al.* 2001; Stanford *et al.* 2001).

El gran avance de la tecnología del ADN está permitiendo el desarrollo de métodos de identificación genética de fácil aplicación, poco costosos, muy sensibles y automatizables. Actualmente los métodos utilizados se basan en el uso de marcadores moleculares de ADN (Sancristobal-Gaudy *et al.* 2000; Sherman *et al.* 2004).

Los marcadores de ADN necesitan reunir una serie de características para ser utilizables en identificación genética, y por tanto poder ser usados como sistema de trazabilidad (Ferreira & Grattapaglia 1995).

- Que sean muy polimórficos, es decir que presenten distintas formas genéticas (alelos) en una población que nos permitan diferenciar los individuos.
- Que estén distribuidos por todos los cromosomas de una especie para obtener información representativa de la posible variación existente.

- Que sean independientes del estado físico y desarrollo del individuo.
- Que tengan un bajo coste tanto en la obtención como en la aplicación para que no se encarezca excesivamente el análisis. Su detección en numerosos individuos debe ser fácil, rápida y barata.
- Que sean públicos y no estén sometidos a ningún tipo de patente.
- Que su interpretación sea sencilla y objetiva.
- Que requieran poca cantidad y calidad de muestra.
- Que sean reproducibles en cualquier experimento de laboratorio, es decir, utilizar marcadores contrastados internacionalmente que permitan la repetición del análisis en cualquier laboratorio y el intercambio de información de animales analizados en diferentes laboratorios.
- Que sean automatizables para aumentar el rendimiento y abaratar los costes.
- Que sean estables para que las mutaciones no puedan influir en los resultados.
- Que tengan preferiblemente herencia codominante. Según el tipo de aplicación del marcador, la tecnología elegida debe ser capaz de detectar las diferentes formas del marcador, es decir, distinguir entre un homocigoto y un heterocigoto.
- Que sean discriminantes, o sea, capaces de detectar diferencias entre individuos estrechamente emparentados.
- Que no estén sujetos a influencias ambientales. La inferencia del genotipo de un marcador debe ser independiente del ambiente en que vive el individuo o de su etapa de desarrollo.
- Que sean neutros. El alelo presente en el locus del marcador es independiente de la presión de selección que se ejerce sobre el individuo y no tiene ningún efecto sobre ella. Esta afirmación suele ser una suposición porque, generalmente, no hay datos disponibles que confirmen o nieguen esta propiedad.

Para que se den muchas de estas características, los marcadores deben ser susceptibles de ser analizados mediante la Reacción en Cadena de

la Polimerasa (PCR) (Mullis & Faloona 1987), esta técnica ha impulsado enormemente el desarrollo de la genética molecular gracias a su gran fiabilidad, rapidez y repetibilidad.

Desde principios del siglo XX ya se viene trabajando con los que probablemente pueden considerarse los primeros marcadores genéticos, los grupos sanguíneos, a los que siguieron otros marcadores como los polimorfismos bioquímicos o las inmunoglobulinas del HLA (Human Leucocyte Antigen). Sin embargo, el escaso número de variantes polimórficas limitó mucho su empleo, además con el descubrimiento de otros tipos de marcadores estos dejaron de emplearse.

3.4. Marcadores moleculares utilizados en el presente trabajo

Existen diferentes marcadores genéticos cada uno con sus ventajas e inconvenientes. En este apartado se van a explicar los utilizados en el presente trabajo por sus especiales características y por ser los que más frecuentemente se están empleando en la actualidad para identificación genética, nos referimos a los Microsatélites y los SNPs.

3.4.1. Microsatélites o STR (Short Tandem Repeat)

Los *loci* microsatélites (Litt & Luty 1989; Koreth *et al.* 1996), también conocidos como SSR ("Simple Sequence Repeat") o STR (Short Tandem Repeat) consisten en repeticiones de secuencias cortas (2-8pb) (Chambers & MacAvoy 2000) que han sido descritos en la mayoría de las especies y poseen una tasa de mutación que oscila entre 10^{-3} y 10^{-5} (Bowcock *et al.* 1994; Forbes *et al.* 1995; Brinkmann *et al.* 1998). Presentan herencia codominante y son fáciles de detectar metodológicamente.

La existencia de microsatélites en genomas eucariotas, concretamente en levaduras y vertebrados fue descrita inicialmente por Hamada y colaboradores (1982), donde son muy frecuentes, bien distribuidos (Tautz 1989; Weber & May 1989) y mucho más polimórficos que los *loci* hipervariables formados por minisatélites, constituyendo la clase de

marcadores moleculares más polimórficos que se conocen (Ferreira & Grattapaglia 1995; Cornide 2002). Aportan un alto grado de información debido al elevado número de alelos por locus que presentan (Baumung *et al.* 2004). Por todo esto hasta el momento aparecen como la herramienta más poderosa de discriminación genética entre animales (Blott *et al.* 1999). El polimorfismo, es decir la variación entre los distintos individuos, se basa en el número de veces que se repite el “motivo”. Se ha señalado que los elementos más repetidos en mamíferos son extensiones de dinucleótidos, siendo la más común la repetición CA.

Los microsatélites han demostrado ser marcadores muy informativos para estudios poblacionales, con un alto grado de polimorfismo (Vignal *et al.* 2002), cuyo número medio de alelos por locus puede variar en un rango muy amplio, en trabajos llevados a cabo en bovinos se han detectado de 5 a 16 (Sanz *et al.* 2007). Actualmente se utilizan también en análisis forenses (Cydne *et al.* 2000), pruebas de paternidad (Mommens *et al.* 1998; Schnabel *et al.* 2000; Sanz *et al.* 2002), estudios de biodiversidad (Peelman *et al.* 1998; Martin-Burriel *et al.* 1999; Martin-Burriel *et al.* 2007), análisis filogenéticos (Ritz *et al.* 2000; Tapio *et al.* 2006), construcción de mapas de ligamiento o como marcadores para la detección de QTLs de importancia económica en producción animal (Arranz *et al.* 1998; Imai *et al.* 2007). Por las ventajas que poseen estos marcadores, se ha trabajado ampliamente en la búsqueda y empleo en diferentes especies de interés en ganadería, tales como ovino (Dietz *et al.* 1993; Buchanan & Thue 1998); caprino (Kemp *et al.* 1995), porcino (Moran 1993) y bovino (Fries *et al.* 1990; Vaiman *et al.* 1994; Ciampolini *et al.* 1995; Usha *et al.* 1995; Moazami-Goudarzi *et al.* 1997; Martin-Burriel *et al.* 1999; Giovambattista *et al.* 2000; Uffo 2003; Liron *et al.* 2006).

Actualmente no existe obligatoriedad impuesta por ningún organismo oficial que indique los marcadores genéticos que deben utilizarse para una correcta identificación de los animales y productos derivados. Sin embargo, la Sociedad Internacional de Genética Animal (ISAG: <http://www.isag.org.uk>), que es una organización que favorece el

intercambio a nivel internacional de información y reactivos entre sus miembros, con el objetivo de estandarizar los marcadores y técnicas utilizadas internacionalmente para la identificación animal en las distintas especies domésticas; teniendo en cuenta los marcadores genéticos bovinos descritos y sus características, sugiere a nivel internacional un grupo de microsatélites (Tabla 1). Estos son los marcadores recomendados como más apropiados para llevar a cabo pruebas de identificación y de filiación, que garanticen una correcta identificación y permitan una estandarización a nivel internacional.

Teniendo esto en cuenta, diversas asociaciones de ganaderos de razas bovinas complementan la información contenida en sus libros genealógicos con datos genotípicos de estos microsatélites, llevando así un control más exhaustivo de sus animales.

Tabla 1: Microsatélites recomendados por la ISAG para identificación genética y chequeos de parentesco en la especie bovina (<http://www.isag.org.uk/journal/comparisonguide.asp>).

Microsatélite	Cromosoma	Tamaño (pb)	Nº alelos	Referencia
ETH225 (D9S1)	9	140-158	8	(Steffen <i>et al.</i> 1993)
INRA023 (D3S10)	3	196-222	10	(Vaiman <i>et al.</i> 1994)
ETH10 ⁴ (D5S3)	5	209-223	7	(Toldo <i>et al.</i> 1993)
BM2113 (D2S26)	2	125-143	10	(Bishop <i>et al.</i> 1994)
BM1824 (D1S34)	1	178-190	5	(Bishop <i>et al.</i> 1994)
TGLA227 (D18S1)	18	77-97	10	(Georges & Massey 1992)
TGLA126 (D20S1)	20	113-125	7	(Georges & Massey 1992)
TGLA122 (D21S6)	21	137-183	15	(Georges & Massey 1992)
SPS115 (D15)	15	246-260	7	(Moore & Byrne 1993)

Un estudio de simulación ha mostrado que un conjunto de 8 o más microsatélites parecen ser suficientes para conseguir alcanzar la identificación individual en ganado bovino sea cual sea la estructura de la población muestreada (Arana *et al.* 2002). Así, la determinación del perfil de ADN mediante el uso de estos microsatélites, puede confirmar o rechazar la

identidad de dos muestras distintas de ADN a un nivel de probabilidad del 99.9% y puede utilizarse como herramienta para el proceso la verificación y revisión al azar de la trazabilidad de los animales y la carne.

3.4.2. SNPs: Polimorfismos de una base (Single Nucleotide Polymorphism)

Son los marcadores de más reciente utilización, pero que se están desarrollando muy rápidamente en los últimos años. Consisten en la sustitución de un nucleótido por otro. Se encuentran distribuidos por todo el genoma y pueden encontrarse tanto en regiones codificantes como no codificantes (Sachidanandam *et al.* 2001). Actualmente sabemos que estos polimorfismos son muy abundantes en el genoma de los mamíferos y se calcula la existencia de un SNP por cada 1000 pares de bases en humanos (Wang *et al.* 1998) o 1 por cada 500 pares de bases en ratón (Lindblad-Toh *et al.* 2000) y bovino (Heaton *et al.* 2001b). Si tenemos en cuenta que el tamaño promedio del genoma de los mamíferos es de 3.000 millones de pares de bases se puede prever que las secuencias de ADN de 2 individuos pueden llegar a diferir en unos 3 millones de posiciones, lo que implica un gran potencial para identificación. Son marcadores muy atractivos desde el punto de vista de su empleo en identificación ya que son genéticamente estables en mamíferos (Markovtsova *et al.* 2000; Nielsen 2000; Thomson *et al.* 2000), abundantes en bovino (Heaton *et al.* 2001a), tienen baja tasa de mutación del orden de 10^{-8} (Nachman & Crowell 2000; Kondrashov 2003), bajas tasas de error en el genotipado (Kennedy *et al.* 2003), la interpretación de los resultados es menos compleja que el caso de otros marcadores (Krawczak 1999), permiten la representación de los genotipos mediante una “firma digital de ADN” (Fries & Durstewitz 2001) y ofrecen un alto potencial para la automatización (Kruglyak 1997; Wang *et al.* 1998; Lindblad-Toh *et al.* 2000).

En este tipo de polimorfismos pueden estar presentes cualquiera de los cuatro nucleótidos en cualquier posición en el genoma, por lo tanto se debe suponer que cada SNP tendría cuatro alelos. Teóricamente esto es posible, pero en la práctica la mayoría de los SNPs tienen solamente dos

variantes, la secuencia original y la versión mutada. Esto se debe a la vía en que estos aparecen y se distribuyen en una población. Un SNP se origina cuando ocurre una mutación puntual en el genoma, convirtiendo un nucleótido en otro. Si la mutación ocurre en los gametos de un individuo, puede heredarse por uno o más descendientes y después de muchas generaciones, el SNP puede establecerse en la población. Para que se produzca un tercer alelo, una nueva mutación debe ocurrir en la misma posición en el genoma de otro individuo, y este individuo y su descendencia deben reproducirse de forma tal que el nuevo alelo quede establecido. Esto no es imposible, pero es poco probable, por lo tanto la mayoría de los SNPs son bialelicos (Brookes 1999).

El método más directo para detectarlos es la secuenciación de segmentos de ADN previamente amplificados por PCR. Una vez localizado el SNP se pueden utilizar diversas técnicas para su análisis, como son: extensión del primer (Pastinen *et al.* 1996), pirosecuenciación (Ronaghi *et al.* 1998), espectrometría de masas MALDI-TOF (*Matrix Assisted Laser Desorption-Time of Flight*) (Andersen & Mann 2000), microarrays o microchips de DNA (Schena *et al.* 1998) y discriminación alélica mediante PCR a tiempo real usando sondas TaqMan (Livak *et al.* 1995) o utilizando cebadores modificados en 3' (Liu *et al.* 1997) que también se puede realizar en PCR estándar.

Debido a las ventajas que ofrecen este tipo de marcadores y a la diversidad de métodos disponibles para su estudio se están utilizando frecuentemente en diversas especies en estudios de identificación y paternidad (Heaton *et al.* 2002; Li *et al.* 2006; Pakstis *et al.* 2007; Van Eenennaam *et al.* 2007), caracterización de poblaciones (Inagaki *et al.* 2002), estudios de biodiversidad (Ben-Ari *et al.* 2005; Twito *et al.* 2007), búsqueda de asociación con caracteres cuantitativos (Page *et al.* 2002; Stone *et al.* 2005), estudios de trazabilidad (Sauer *et al.* 2002; Capoferri *et al.* 2005; Goffaux *et al.* 2005) y elaboración de mapas de ligamiento (Snelling *et al.* 2005), entre otros.

Los SNPs con frecuencias intermedias son los más útiles y los datos actuales indican que es suficiente una cantidad entre 25-50 SNPs para la identificación segura del ganado y de la carne (Heaton *et al.* 2002). Weller y colaboradores (2006) establecieron mediante estudios de simulación que para obtener una probabilidad de identidad equivalente a la obtenida con un panel de 8 microsatélites serían necesarios 18 SNPs, y por cada microsatélite adicional se necesitarían entre 2-2,5 SNPs adicionales.

En el presente trabajo se han utilizado un total de 20 SNPs de los cuales 16 se encuentran en regiones intrónicas y 4 en regiones exónicas. En la Tabla 2 se muestran los SNPs analizados, localización cromosómica, número de acceso del GenBank y referencia.

Tabla 2: SNPs utilizados en este trabajo para identificación.

Nombre SNP	Cromosoma/ Localización	SNP	Nº acceso (GenBank)	Referencia
C1A/G	1/exón	A/G snp1455	<u>U02564</u>	(Laurent & Rodellar 2001)
C2A/G	2/intrón	A/Gsnp109	<u>AF458963</u>	(Stone <i>et al.</i> 2002)
C4C/G	4/intrón	C/Gsnp361	<u>AF458965</u>	(Stone <i>et al.</i> 2002)
C5A/C	5/intrón	A/C snp261	<u>AF458966</u>	(Stone <i>et al.</i> 2002)
C6A/G	6/intrón	A/Gsnp103	<u>AF465155</u>	(Heaton <i>et al.</i> 2001b)
C7C/T	7/exón	C/T	<u>AY738459.1</u>	(Ordovas <i>et al.</i> 2006)
C8C/T	8/intrón	C/Tsnp300	<u>AF465157</u>	(Stone <i>et al.</i> 2002)
C10C/T	10/intrón	C/T snp396	<u>AF465158</u>	(Stone <i>et al.</i> 2002)
C11A/G	11/intrón	A/Gsnp111	<u>AF465160</u>	
C12A/G	12/intrón	A/G snp 107	<u>AJ496770</u>	(Werner <i>et al.</i> 2004)
C13A/G	13/intrón	A/G snp252	<u>AF465161</u>	(Heaton <i>et al.</i> 2002)
C17A/G	17/intrón	A/Gsnp302	<u>AF465165</u>	(Stone <i>et al.</i> 2002)
C18G/T	18/intrón	G/Tsnp453	<u>AF465167</u>	(Stone <i>et al.</i> 2002)
C19C/G	19/exón	C/G	<u>AF285607.2</u>	(Roy <i>et al.</i> 2001)
C20C/T	20/intrón	C/T snp 233	<u>AJ505155</u>	(Werner <i>et al.</i> 2004)
C21C/G	21/intrón	C/G snp137	<u>AF465171</u>	(Stone <i>et al.</i> 2002)
C23C/T	23/intrón	C/Tsnp 338	<u>AF465172</u>	(Stone <i>et al.</i> 2002)
C24A/G	24/intrón	A/G snp 397	<u>AF465175</u>	(Stone <i>et al.</i> 2002)
C26C/T	26/intrón	C/T snp 329	<u>AJ496781</u>	(Werner <i>et al.</i> 2004)
C29A/G	29/exón	A/G snp 4558	<u>AF248054</u>	(Smith <i>et al.</i> 2000)

De estos 20 SNPs los 4 localizados en regiones exónicas (SNP C1A/G, SNP C7C/T, SNP C19C/G y SNP C29A/G) han sido asociados a diferentes

caracteres productivos por otros autores. En el SNP denominado C1 se ha llevado a cabo un estudio de asociación que se mostrará en el Capítulo de resultados. Estos marcadores resultan de interés no solo por su aplicación a la identificación y trazabilidad, sino por su posible asociación con caracteres productivos y uso en trabajos de selección asistida por marcadores (MAS: Marker Assisted Selection)

3.4.2.1. SNPs localizados en regiones exónicas

Los QTL (Quantitative Trait Loci) son considerados regiones del genoma en la que se encuentra uno o varios genes con un efecto significativo sobre un carácter de interés económico. Aunque la mayoría de los caracteres productivos presentan variación continua y están controlados por varios genes con un efecto pequeño pero acumulativo sobre el carácter; existen genes para algunos caracteres que explican un elevado porcentaje de la variación fenotípica (Winter *et al.* 2002; Van Laere *et al.* 2003; Roy *et al.* 2006).

Actualmente se sabe que hay cientos de QTLs distribuidos por todo el genoma bovino (<http://bovineqtlv2.tamu.edu/home.php>) (Polineni 2004). Además la secuencia del genoma bovino ha permitido identificar un elevado número de genes que se encuentran en estas regiones QTL, esta información es muy útil en la identificación de polimorfismos que pudieran influenciar caracteres productivos en el ganado bovino.

Como se ha mencionado anteriormente en este trabajo han sido utilizados 20 SNPs de los que hay 4 que están localizados en regiones exónicas para los que se ha demostrado su relación con caracteres cuantitativos, los cuales pasaremos a describir brevemente:

3.4.2.1.1. SNP en el gen de la transferrina

La transferrina es la proteína que transporta a través de la sangre el hierro absorbido en el intestino y el liberado por el catabolismo de la hemoglobina hacia los sitios de almacenamiento (hígado y sistema retículo-

endotelial). Esta proteína es altamente polimórfica en muchas especies y en bovino se han detectado 10 variantes proteicas de acuerdo a su movilidad electroforética.

Laurent y Rodellar (2001) detectaron un SNP c.1455A>G (U02564) en el exón 12 del gen de la transferrina bovina localizado en el cromosoma 1. Este polimorfismo es detectado en animales con serotipos A/D₂ y D₂/E y produce un cambio aminoacídico Asp/Gly en la posición 469 del péptido, esta mutación permite la distinción de dos grupos de variantes de transferrina: D₂ (c.1455G (U02564)→Gly) y A, E y D₁ (c.1455A (U02564)→Asp). Este gen está ubicado en una región donde se han localizado cuatro QTL (Tabla 3 y Figura 3).

Tabla 3: QTLs en la región del cromosoma 1 donde está ubicado el gen de la transferrina.

QTL Id	Nombre carácter	Marcador inicial	Marcador final
64	Birth Weight	BMS1789 (113.5cM)	BMS4014 (148.21cM)
65	Adjusted Weaning weight	BMS1789 (113.5cM)	BMS4014 (148.21cM)
66	Adjusted Yearling weight	BMS1789 (113.5cM)	BMS4014 (148.21cM)
408	Adjusted Fat	BL28 (121.47cM)	BM3205 (127.441cM)

Al estar ubicado en esta región involucrada en varios caracteres cuantitativos, este polimorfismo ha sido empleado para llevar a cabo un estudio de asociación utilizando una muestra de vacas con valores estimados de producción extremos en cantidad de grasa en la leche.

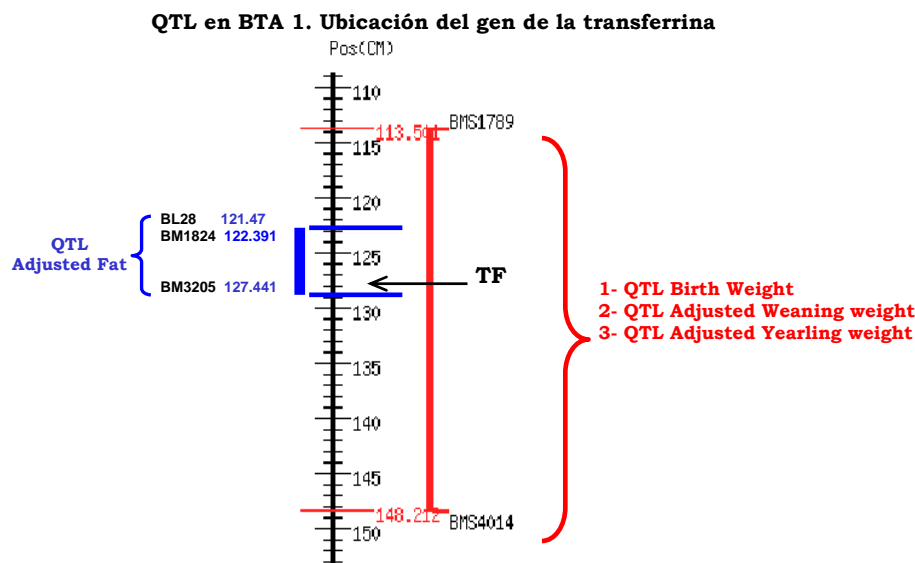


Figura 3: Ubicación relativa del gen de la transferrina bovina relativa a los QTLs encontrados en la región.

3.4.2.1.2. SNP en el gen de la calpaína

Dentro de los factores bioquímicos y genéticos que se han identificado como responsables del ablandamiento de la carne se encuentra el sistema proteolítico de las calpaínas. Las proteasas α -calpaína y μ -calpaína son las dos enzimas responsables del ablandamiento postmortem, aunque se ha indicado que la α -calpaína es la principal enzima. El gen de la α -calpaína se localiza en el cromosoma 29 (Smith *et al.* 2000) en una región QTL para la terneza de la carne (Casas *et al.* 2000; Smith *et al.* 2000) (Figura 4) y se han identificado diferentes polimorfismos en el gen que causan cambios en la composición de aminoácidos en la proteína, los cuales han sido utilizados para asociar este gen con la terneza de la carne (Page *et al.* 2002; Casas *et al.* 2005).

Este polimorfismo ha sido incluido en el panel no solamente por su asociación con el carácter terneza de la carne, sino también con la intención de valorar su potencial en identificación.

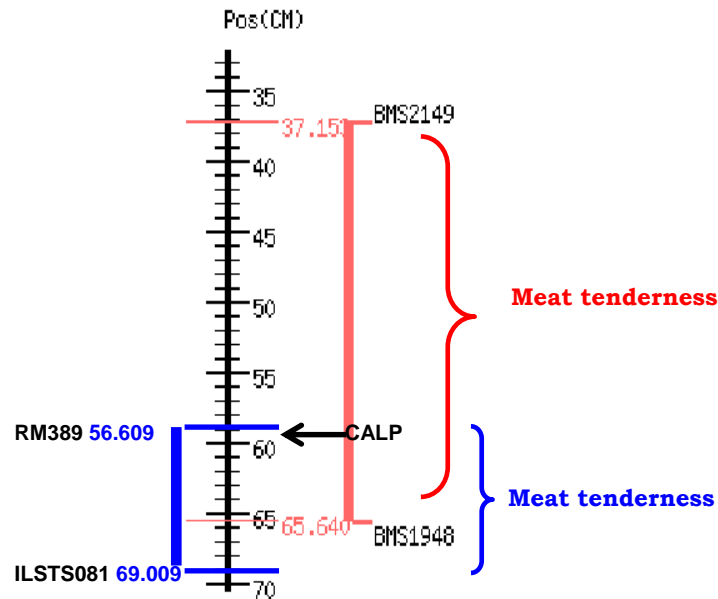


Figura 4: Ubicación relativa del gen de la i-calpaína bovina relativa a los QTL encontrados en la región.

3.4.2.1.3. SNP en el gen de la Proteína Transportadora de Ácidos Grasos 1 (FATP1)

La Proteína Transportadora de Ácidos Grasos 1 (Fatty Acid Transport Protein, FATP1) es una proteína integral de membrana que facilita el transporte de ácidos grasos de cadena larga a través de la membrana plasmática. El gen de la FATP1 ha sido mapeado en el cromosoma 7 (Ordovas *et al.* 2005) donde han sido descritos numerosos QTL relacionados con la grasa. En este trabajo hemos utilizado un SNP C/T localizado en el nucleótido 789 de la secuencia de mRNA (Ordovas *et al.* 2008), que tanto por su función fisiológica como su localización cromosómica hacen que pueda considerarse como gen candidato involucrado en la determinación de engrasamiento en ganado bovino (Ordovas 2006)

3.4.2.1.4. SNP en el gen de la Sintasa de Ácidos Grasos (FASN)

La Sintasa de Ácidos Grasos (FASN) es una enzima citosólica que cataliza la síntesis de ácidos grasos saturados de cadena larga, por lo que es clave en el metabolismo lipídico. En bovino ha sido mapeado el gen en el

cromosoma 19 (Roy *et al.* 2001) cercano a regiones QTL (Figura 5). El polimorfismo G/C ubicado en el exón 1 no traducible ha sido utilizado en este trabajo ya que ha sido demostrada además su asociación con contenido de grasa en leche (Roy *et al.* 2006), debido a la modificación de un sitio de unión del factor de transcripción Sp1 (Ordovas *et al.* 2008).

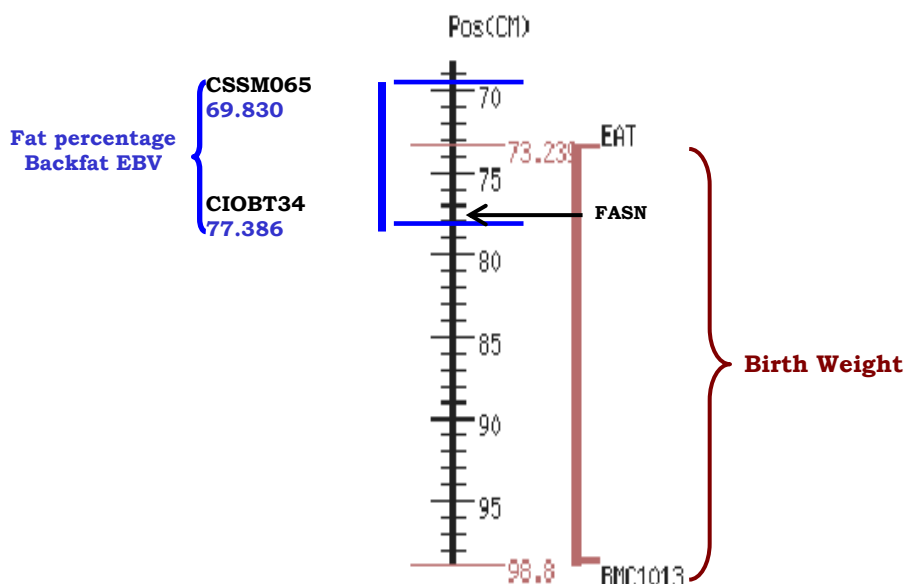


Figura 5: Ubicación relativa del gen de la FASN bovina relativa a los QTL encontrados en la región.

3.5. Aplicación de la identificación individual a la trazabilidad

Los sistemas de trazabilidad basados en identificación convencional de animales no son infalibles, ya que una vez que el animal va al matadero comienza un proceso de cambio de tipos de identificación que sustituyen los crotales por un sistema de etiquetado, pudiendo inducir a errores en la identificación. Es decir, el seguimiento de la carne tiene un punto en la cadena alimentaria que resulta de vital importancia, y es a partir del momento en que el animal llega al matadero y la canal va ser dividida en numerosas partes que pueden tener diferentes destinos.

Para evitar errores debido a la pérdida de la identificación, se están desarrollando sistemas de verificación que puedan asegurar totalmente la

trazabilidad del animal en cada una de las etapas de la cadena de producción, en especial después del despiece del animal.

A continuación se exponen dos de dichos sistemas: la comparación de perfiles y la asignación individual a raza o población.

3.5.1. Comparación de perfiles

La trazabilidad molecular se basa específicamente en la comparación de perfiles de marcadores moleculares de ADN obtenidos para cada individuo, como son los microsatélites (Sancristobal-Gaudy *et al.* 2000; Arana *et al.* 2002; Vazquez *et al.* 2004) o SNPs (Capoferri *et al.* 2005; Goffaux *et al.* 2005), que pueden garantizar la trazabilidad de la carne hasta el animal de origen permitiendo discriminar si dos muestras son la misma o son muestras distintas.

La lógica de un sistema de trazabilidad molecular opera con la toma de una muestra biológica de origen de cada animal antes de que se pierda la identidad individual del mismo (ej.: antes de la pérdida de los crotales). Esta muestra, convenientemente identificada, se almacena por un período no inferior a la “vida útil” del producto carne, hasta que el animal es consumido por completo. Posteriormente y en el eventual caso que en algún punto de la cadena (frigorífico, supermercado o consumidor) se quiera establecer el origen de un determinado corte de carne, se toma una segunda muestra del mismo (muestra problema). Tanto para la muestra de origen como para la muestra problema se determina un perfil específico de ADN. Estos perfiles son comparados entre sí, pudiendo establecerse si ambas huellas genéticas son idénticas o no. Si se trata de muestras pertenecientes al mismo animal, el perfil genético va a ser el mismo, independientemente del tipo de muestra de que se trate (Figura 6). Puede tratarse de cualquier tejido del animal vivo (sangre, pelo, piel, cartílago, semen, etc.) o bien un trozo de tejido obtenido de una pieza de la canal (carne, grasa, etc.), o de alimentos procesados, que han podido ser sometidos a diversos tratamientos, tanto físicos como

químicos. En todos esos supuestos, el perfil genético será el mismo si se trata de muestras del mismo animal.

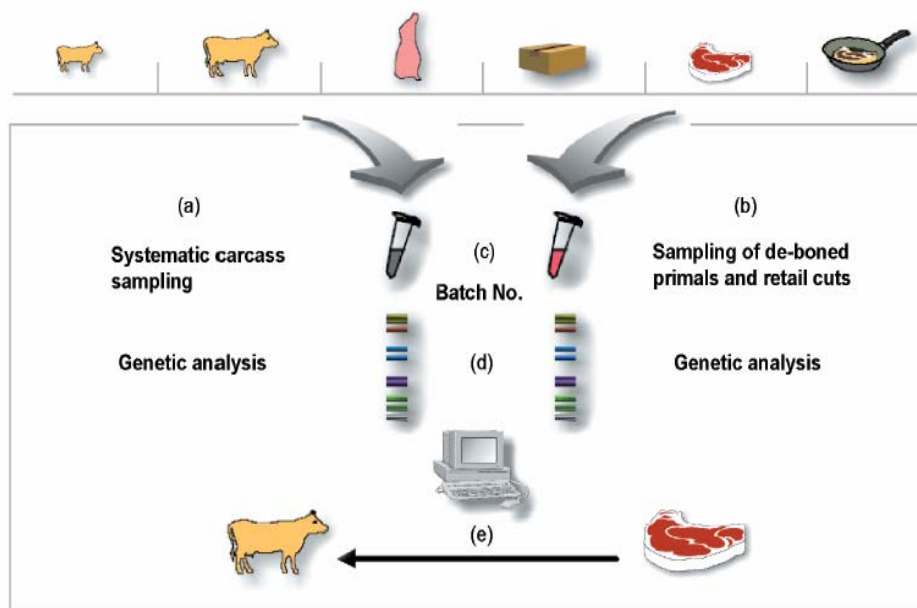


Figura 6: Esquema que representa tanto la toma de muestras de forma sistemática de animal vivo y canales (A) como de productos procesados (B) para su posterior genotipado e identificación (Cunningham & Meghen 2001).

Si dos muestras corresponden al mismo animal, deberán tener necesariamente el mismo genotipo para todos los marcadores estudiados. La probabilidad de que dos muestras que no pertenecen al mismo animal tengan el mismo genotipo para todos los marcadores analizados por efecto del azar depende del número de marcadores estudiados, de forma que al aumentar el número de marcadores, dicha probabilidad disminuye exponencialmente (San Cristobal-Gaudy *et al.* 2000; Arana *et al.* 2002). La Figura 7 nos muestra la probabilidad de que dos muestras coincidan al azar según el número de microsatélites analizados. Como puede observarse, con un número de 9 microsatélites (los recomendados por la ISAG), la probabilidad de que el genotipo de dos individuos coincida por azar es de 10^{-9} , lo que nos indica que es prácticamente imposible que dos muestras presenten el mismo genotipo a no ser que se trate de la misma muestra o de gemelos monocigóticos.

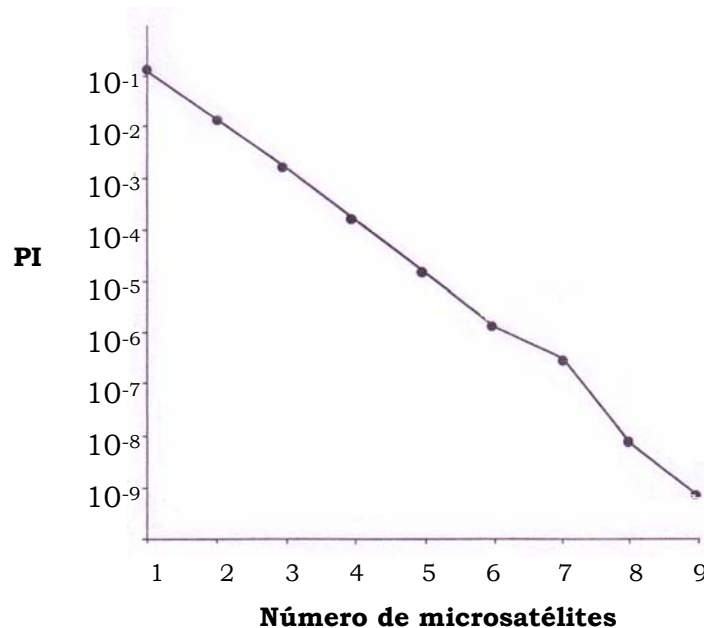


Figura 7: Probabilidad (en ordenadas, escala logarítmica) de que el genotipo de dos muestras coincida por azar según el número de microsatélites analizados (Vazquez *et al.* 2004).

El sistema de trazabilidad genética ideal pasaría por el análisis de los marcadores moleculares en todos los animales vivos. Sin embargo, y aunque el análisis de los marcadores de ADN resulta cada vez más económico, resultaría extremadamente caro hacer analizar todos los individuos que van destinados a matadero.

Una alternativa que se está aplicando en algunos países consiste en tomar muestras para identificación de todos los individuos, pudiendo ser un momento idóneo al marcar el ternero con el crotal correspondiente. Para este cometido existen ya sistemas en el mercado que facilitan la obtención de una muestra de pabellón auricular en el momento de colocación del crotal (Vazquez *et al.* 2004), aunque se puede obtener cualquier tipo de muestra como pelo o papel impregnado en sangre, etc, siempre que cumplan los requisitos especificados con anterioridad. Los sistemas de obtención de la muestra facilitan asimismo una sistemática que permite el almacenado de las muestras. De esta forma, la identidad de un determinado individuo

durante el proceso de cría o de cualquier pieza de carne durante el proceso de comercialización se puede comparar con la muestra inicial en controles rutinarios o en casos excepcionales que así lo aconsejen o requieran, sirviendo para control y apoyo de los otros sistemas de identificación que están establecidos legalmente.

Existen otras alternativas, como la posibilidad de identificar mediante marcadores genéticos a todos los reproductores de las explotaciones y tratar de llevar a cabo un seguimiento, sin necesidad de analizar al individuo, sino a partir de la información genética de sus progenitores (Zaragoza *et al.* 2003; Rodellar *et al.* 2006).

3.5.2. Asignación individual a raza o población

Otra alternativa que se puede tener en cuenta es la de llegar mediante el análisis de los marcadores moleculares del producto animal a la raza de origen del producto, a lo que se le puede llamar trazabilidad de raza. La asignación de un sujeto a una raza utilizando métodos moleculares puede efectuarse esencialmente a través de dos estrategias (Ajmone *et al.* 2004):

I) aproximación determinada: contempla la búsqueda de marcadores moleculares específicos de una raza y/o genes con diferentes variantes de alelos fijadas en razas diferentes. La disponibilidad de tales secuencias específicas permitiría la utilización de pruebas diagnósticas sencillas y la asignación de un individuo a la raza de modo seguro, sin la necesidad de ninguna interferencia estadística. Por ejemplo en cerdo el análisis de la variación de los loci *MC1R* y *KIT* (Kijas *et al.* 1998; Giuffra *et al.* 2002) implicados en el color de la capa muestra una prueba simple que puede ser empleada en la verificación de la población de origen (Carrión *et al.* 2003).

II) aproximación probable: esta estrategia comporta la creación de una base de datos de razas con marcadores moleculares y por lo tanto que contenga información sobre los alelos y las frecuencias alélicas características de cada raza. El individuo a asignar se analizará con los

marcadores moleculares y la atribución se efectuará de manera probable, a partir de las frecuencias alélicas entre las razas. Para llevar a cabo la aproximación probable se pueden aplicar diferentes estrategias, como son los métodos basados en frecuencias alélicas, métodos Bayesianos o los basados en distancias genéticas (Paetkau *et al.* 1995; Rannala & Mountain 1997; Cornuet *et al.* 1999), que serán comentados a continuación.

Los métodos basados en frecuencias alélicas consisten en asignar un individuo a una población, basándose en el valor de probabilidad de su genotipo individual de pertenecer a ella. Utilizan algoritmos estadísticos basados en funciones de máxima verosimilitud (Paetkau *et al.* 1995) que contemplan esencialmente tres partes: i) cálculo de las frecuencias alélicas en todas las poblaciones candidatas; ii) cálculo de la función de verosimilitud de cada genotipo multilocus en el interior de la población de referencia; iii) asignación del individuo a la población, en la que la función de verosimilitud de su genotipo es máxima. En este trabajo se ha utilizado esta metodología para realizar la asignación de individuos a raza usando datos genotípicos multilocus mediante la utilización del programa WHICHRUN v4.1 (Banks & Eichert 2000) basado en funciones de máxima verosimilitud.

Los métodos Bayesianos son semejantes a los anteriores, pero utilizan particulares funciones de probabilidad para estimar las frecuencias de las poblaciones de referencia a partir de las frecuencias observables en la muestra. La atribución se realiza calculando la probabilidad marginal al observar el genotipo de los individuos innatos en el interior de cada una de las poblaciones de referencia, en base a las frecuencias alélicas estimadas. Este método es el implementado en el software Structure (Pritchard *et al.* 2000) que está siendo utilizado frecuentemente en asignación (Canon *et al.* 2008; Dalvit *et al.* 2008).

Por su parte los métodos basados en distancias genéticas asignan los individuos desconocidos a la población genéticamente más cercana o próxima. Este método contempla la construcción de una matriz de distancias entre las posibles parejas de individuos y la atribución se

ejecutará después, sobre la base de la distancia entre el individuo a asignar y la media de las distancias entre los sujetos de las poblaciones de referencia. Estos métodos son aplicables a diversos tipos de marcadores con resultados fácilmente interpretables, que pueden ser representados gráficamente aplicando análisis multivariados o algoritmos que construyen árboles filogenéticos. En este trabajo se ha utilizado la distancia entre individuos (Bowcock *et al.* 1994) basada en la proporción de alelos compartidos (Chakraborty & Jin 1993) en la construcción de un árbol filogenético utilizando el panel de marcadores microsatélites para ilustrar la capacidad de asignación de individuos a su raza.

Utilizando estas metodologías es posible realizar la asignación de muestras problema a la población que le dio origen. Existen numerosos trabajos que avalan la aplicación de estos métodos de asignación en diversas especies como en bovino (Maudet *et al.* 2002; Canon *et al.* 2008), en equino (Kruger *et al.* 2005; Thirstrup *et al.* 2008) o en salmónidos (Rengmark *et al.* 2006).

4. MATERIAL Y MÉTODOS

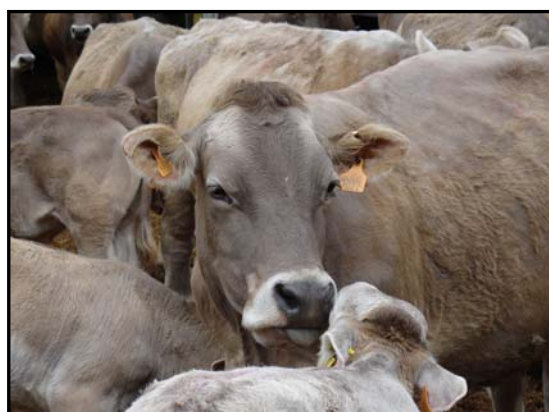
4. MATERIAL Y MÉTODOS

4.1. Material biológico

4.1.1. Animales

Para el estudio de trazabilidad, principal objetivo de esta memoria, se han utilizado animales de dos razas bovinas autóctonas aragonesas *Parda de Montaña* y *Pirenaica*, procedentes de dos explotaciones destinadas a la producción de carne del Pirineo Aragonés (Figura 8).

Se analizaron los progenitores y descendientes presentes en las dos explotaciones del Pirineo Aragonés en el período comprendido entre marzo de 2004 y septiembre de 2005. Se obtuvieron muestras de un total de 564 animales, 257 animales de raza de Montaña, todos pertenecientes a la explotación I y 307 animales de raza Pirenaica de los que 126 animales pertenecían a la explotación I y 181 a la explotación II.



Parda de Montaña



Pirenaica

Figura 8: Ejemplares de raza Parda de Montaña y Pirenaica.

4.1.2. Muestras

4.1.2.1. Muestras de sangre y carne para el estudio de trazabilidad

Entre sangre y músculo se analizaron 675 muestras, 446 pertenecían a la explotación I (302 de raza Parda de Montaña y 144 de Pirenaica) y 229 a la explotación II, todas procedentes de animales de raza Pirenaica (Figura 9).

Las muestras fueron tomadas a tres niveles:

- I. Sangre de todos los progenitores
- II. Sangre de todos los descendientes
- III. Músculo de los descendientes sacrificados en matadero

Las muestras fueron identificadas con un código interno de laboratorio e incluidas en la base de datos del Laboratorio de Genética Bioquímica (LAGENBIO) desarrollada en Microsoft Access 2003 que fue vinculada a una aplicación web y que será descrita en el apartado 4.4.2. Se incluyó en cada caso toda la información disponible como: número de identificación, raza, sexo, categoría, propietario, tipo de muestra, explotación, fecha de nacimiento y datos de padre y madre en el caso de todos los descendientes.

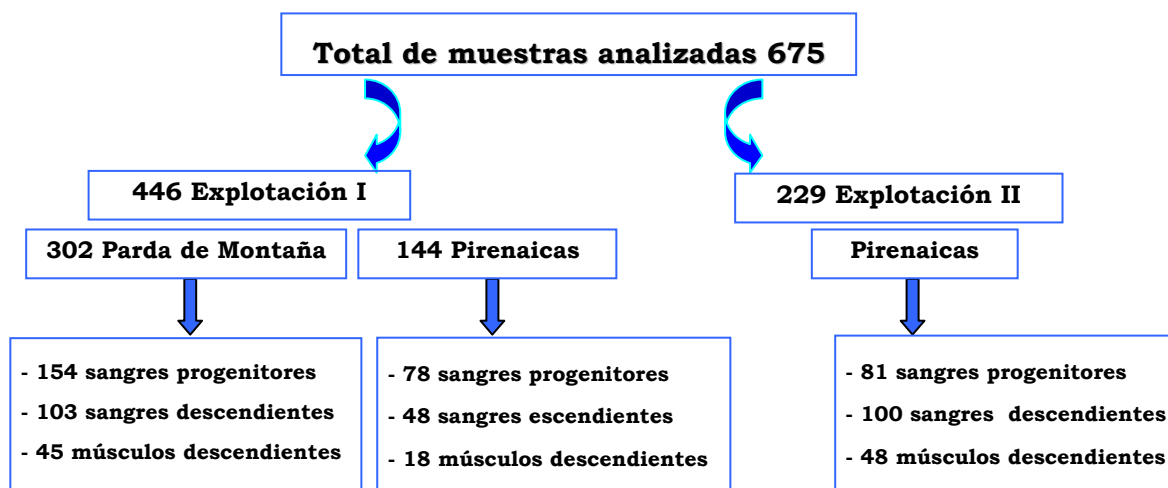


Figura 9: Distribución de las muestras analizadas para el estudio de trazabilidad.

4.1.2.2. Muestras de ADN de diferentes razas españolas para cálculos de frecuencias de SNPs

También se calcularon las frecuencias alélicas de los SNPs analizados. Para ello se utilizaron muestras existentes en el banco de ADN del Laboratorio de Genética Bioquímica de la Facultad de Veterinaria (LAGENBIO) de la Universidad de Zaragoza. Se emplearon 184 muestras, la mayoría de razas autóctonas españolas que se describen a continuación: Asturianas de los Valles (23), Frisona (23), Menorquina (23), Limousine (23), Monchina (23), Serrana Negra de Teruel (23), Parda de Montaña (23) y Pirenaica (23).

4.1.2.3. Muestras de ADN para estudios de asociación de polimorfismos

El estudio de asociación de polimorfismos con caracteres cuantitativos se realizó analizando muestras de ADN de 211 vacas Holstein-Friesian en lactación con valores extremos de EBV (Estimated Breeding Values) para cantidad de grasa en la leche.

Se emplearon muestras de 117 vacas (cola H) con valores altos de EBV (45,19) y fiabilidad del 54% y 94 (cola L) con bajo EBV (-24,45) y fiabilidad de 58%. Todas ellas seleccionadas de un total de 2573 pertenecientes a cuatro explotaciones diferentes localizadas en el Norte de Aragón con fiabilidad de EBV>47%. Los valores de EBV para el contenido de grasa en leche fueron evaluados por la Confederación Nacional de Asociaciones de Holstein-Friesian de España (CONAFE). Cada cola estaba constituida por el 1% de la población de animales con valores extremos de EBV (Figura 10).

En la Tabla 4 se muestran los parámetros genéticos para la cantidad de grasa en la leche para la población completa aragonesa de Holstein-Friesian y los animales seleccionados para las colas con alto (cola H) y con bajo (cola L) valor de EBV.



Figura 10: Muestras seleccionadas para el estudio de asociación del SNP del gen de la transferrina con animales con reportes de cantidad de grasa en leche.

Tabla 4: Promedio, desviación estándar (S.D.) y rango de EBV (máximo y mínimo) para producción de grasa en leche (kg) en la población aragonesa de Holstein-Friesian y en los animales que forman parte de la cola alta (H) y baja (L) para la distribución de valores de EBV para contenido de grasa en la leche.

Población	Promedio	S.D.	Máximo	Mínimo
Aragón Holstein-Friensian (n=7631)	13.67	14.58	63	-53
Cola alta (H) baja (L) (n=211)	14.17	35.35	63	-53
Cola alta (H) (n=117)	45.19	5.53	63	39
Cola baja (L) (n=94)	-24.45	8.08	-17	-53

4.2. Extracción de ADN y cuantificación

4.2.1. Extracción

El ADN de las muestras fue extraído siguiendo diferentes protocolos dependiendo del tipo de muestra de partida.

4.2.1.1. A partir de sangre

Para la extracción del material genético a partir de sangre se utilizaron dos métodos: el descrito por Lahiri y colaboradores (1992) y el protocolo standard utilizando el kit *GFX™ Genomic Blood DNA purification kit* (Amersham Biosciences).

El método descrito por Lahiri y colaboradores se basa en el empleo de una mezcla de detergentes (Tritón X-100 2% y SDS 10%) y precipitación de las proteínas celulares en un medio saturado de cloruro de sodio. Este método evita el uso de solventes orgánicos (fenol/cloroformo) y de proteinasa K.

El kit *GFX™ Genomic Blood DNA purification kit* (Amersham Biosciences), también utilizado para extraer ADN a partir de sangre completa, usa un agente caotrópico para obtener las células sanguíneas nucleadas y promueve la unión del ADN genómico a una fibra de cristal contenida en la columna MicroSpin. Los contaminantes son eliminados mediante lavados con etanol y el ADN purificado es eluido en un buffer con baja fuerza iónica contenido en el kit. Con este kit se obtienen en apenas 20 minutos y a partir de 100µl de sangre aproximadamente 2-4µg de ADN de elevada calidad.

4.2.1.2. A partir de músculo

La extracción del ADN a partir de músculo se realizó siguiendo el protocolo descrito por Taboada (1995). En este caso partimos de 0,1g de músculo el cual fue pulverizado con nitrógeno líquido; se mezcló con 1ml de buffer RSB1x (10mM Tris-HCl ph 7,4, 10mM NaCl y 25mM EDTA), posteriormente se adicionó SDS y proteinasa K. Finalmente se utilizó una mezcla de fenol/cloroformo/alcohol isoamílico y precipitación con etanol.

4.2.2. Cuantificación

Una vez obtenido el ADN se cuantificó mediante espectrofotometría por lectura de absorbancia a 260nm, considerando que una unidad de densidad óptica a esta longitud de onda equivale a 50µg/ml de ADN de doble cadena. La pureza fue estimada mediante lecturas a 260, 230 y 280nm, la relación de lecturas a 260nm y 280nm proporciona una idea del grado de pureza del material, el cual debe estar entre 1,7 y 2.

4.3. Análisis de marcadores moleculares

Los marcadores moleculares elegidos para el estudio de trazabilidad han sido los microsatélites y los SNPs. Aunque se trata de marcadores con características muy diferentes, se consideran los más idóneos para la aplicación en sistemas de identificación y trazabilidad.

A continuación pasaremos a describir todas las etapas de trabajo con cada uno de estos:

4.3.1. Marcadores moleculares microsatélites

Los microsatélites elegidos para desarrollar este trabajo han sido los que forman parte del Panel de Microsatélites recomendados por la Sociedad Internacional de Genética Animal (International Society for Animal Genetics: ISAG) para identificación de bovinos. Los nombres de los 9 microsatélites estudiados, la localización cromosómica, las secuencias de los oligonucleótidos utilizados para la amplificación de los fragmentos y el sentido, así como el rango de talla de los fragmentos obtenidos se muestran en la Tabla 5.

Tabla 5: Microsatélites recomendados por la ISAG para identificación genética bovina (www.isag.org.uk/ISAG/all/02_PVpanels_LPCGH.doc).

Locus	Cromosoma	Sentido	Secuencia del cebador (5'-3')	Tamaño fragmento (pb)
BM1824	1	Directo	GAGCAAGGTGTTTTCCAATC	178-190
		Inverso	CATTCTCCAACCTGCTTCCTTG	
BM2113	2	Directo	GCTGCCTTCTACCAAATACCC	125-143
		Inverso	CTTCCTGAGAGAAGCAACACC	
INRA023	3	Directo	GAGTAGAGCTACAAGATAAACTTC	196-222
		Inverso	TAACTACAGGGTGTAGATGAACTC	
SPS115	15	Directo	AAAGTGACACAACAGCTTCTCCAG	246-260
		Inverso	AACGAGTGTCTAGTTTGGCTGTG	
TGLA122	21	Directo	CCCTCCTCCAGGTAAATCAGC	137-183
		Inverso	AATCATATGGCAAATAAGTACATA	
TGLA126	20	Directo	CTAATTTAGAATGAGAGAGGCTTCT	113-125
		Inverso	TTGGTCTCTATTCTCTGAATATTC	
TGLA227	18	Directo	CGAATTCCAAATCTGTTAATTTGCT	77-97
		Inverso	ACAGACAGAACTCAATGAAAGCA	
ETH10	5	Directo	G TTCAGGACTGGCCCTGCTAACA	209-223
		Inverso	CCTCCAGCCCCTTCTCTTCTC	
ETH225	9	Directo	GATCACCTTGCCACTATTTCCT	140-158
		Inverso	ACATGACAGCCAGCTGCTACT	

4.3.1.1. Amplificación de los fragmentos mediante PCR

Para llevar a cabo el genotipado de las muestras para estos loci se realizó en primer lugar la reacción de amplificación de los fragmentos que contenían los microsatélites para lo que se empleó PCR fluorescente.

La PCR fluorescente consiste en la amplificación de una secuencia específica de ADN utilizando cebadores marcados en la posición 5' con fluorocromos. Los fragmentos que se sintetizan emitirán señal fluorescente a diferentes longitudes de onda dependiendo del tipo de fluorocromo que posean una vez que sean excitados por un láser. De este modo se pueden

visualizar todos los fragmentos y por tanto hacer la correcta lectura de los genotipos.

En este trabajo se diseñó una reacción múltiplex para amplificar los 9 fragmentos que contenían los microsatélites de interés, para ello se combinaron los oligonucleótidos marcados con cuatro fluorocromos diferentes (6-FAM, VIC, NED, PET: *Applied Biosystems*) para obtener fragmentos con una longitud variable de modo que pudieran ser detectados todos los genotipos. Se marcó en cada caso el oligonucleótido reverso para la posterior visualización de los resultados. En la Figura 11 se representan los oligonucleótidos marcados con sus fluorocromos y el rango de talla de los fragmentos amplificados de modo que no se solaparan a la hora de realizar la lectura de los genotipos.

PCR con oligonucleótidos marcados fluorescentemente con 6-FAM, VIC, NED y PET.

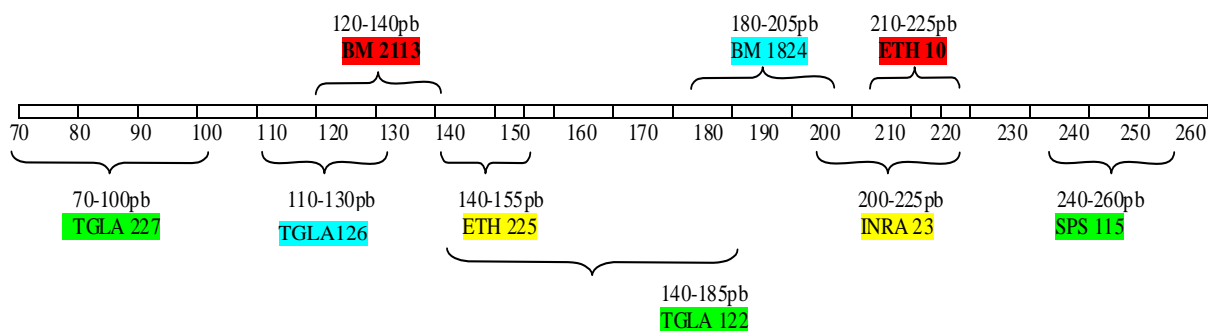


Figura 11: Oligonucleótidos marcados con los fluorocromos 6-FAM, VIC, NED y PET (*Applied Biosystems*) y rangos de talla de los fragmentos producidos.

La PCR fluorescente se llevó a cabo usando el *Quiagen Multiplex PCR kit* formado por el 2x QUIAGEN Multiplex PCR Master Mix que contiene los dNTPs, tampón de PCR, $MgCl_2$ y una *HotStarTaq* DNA Polymerase. Dicha enzima solo puede ser activada después de un ciclo de 15 minutos a 95°C previniendo la formación de productos no deseados en el tiempo que media entre la preparación de la reacción y el primer paso de desnaturalización, con lo que se consigue una elevada especificidad. Además contiene la Q-Solution 5x que es un aditivo que facilita la amplificación de regiones del

ADN con alto grado de estructura secundaria o zonas ricas en GC. La reacción se llevó a cabo en un volumen de 6 μ L con 50ng de ADN genómico, concentraciones finales 1x de Master Mix y de Q-solution y la mezcla adecuada de oligonucleótidos cuyas concentraciones finales se muestran en la Tabla 6.

El programa de amplificación consistió en un paso inicial de desnaturalización de 15 minutos a 95°C para activar la enzima y 30 ciclos de (94°C 30seg, 58°C 1min y 72°C 1min) seguido de un paso final de extensión a 72°C de 30 minutos, en un termociclador ABI 2720 (*Applied Biosystems*).

Tabla 6: Concentraciones de cebadores utilizadas en la PCR fluorescente (P1: cebador sentido directo y P2*: cebador inverso modificado en 5').

Oligonucleótidos (10 μM)	Concentración Final (μM)
TGLA 122 P1 y P2*	0,03
TGLA 126 P1 y P2*	0,1
TGLA 227 P1 y P2*	0,04
ETH 10 P1 y P2*	0,025
ETH 225 P1 y P2*	0,025
BM 1824 P1 y P2*	0,035
BM 2113 P1 y P2*	0,025
INRA 23 P1 y P2*	0,0425
SPS 115 P1 y P2*	0,03

4.3.1.2. Visualización de los fragmentos y lectura de los genotipos

La visualización de los fragmentos amplificados por PCR fluorescente se realizó mediante electroforesis capilar. Con cuidado de no exponer la reacción demasiado tiempo a la luz, el producto final de PCR se diluyó con 50 μ L de H₂O_mQ. Se tomó 1 μ L del producto diluido y se adicionaron 12 μ L de

formamida Hi-Di y 0,25µL de Std 500 LIZ (*Applied Biosystems*) como marcador interno de talla, finalmente se desnaturalizó 5 minutos a 95°C. La electroforesis capilar se realizó usando como soporte polímero Pop 4 en el caso que se trabajó con secuenciador automático ABI Prism 310 (*Applied Biosystems*) o con polímero POP7 usando un secuenciador automático ABI PRISM 3130 (*Applied Biosystems*). Los genotipos fueron interpretados usando el programa GeneScan v3.7 y el programa GeneMapper v3.7 ambos desarrollados por *Applied Biosystems*. En todos los casos se usaron muestras control estandarizadas internacionalmente para la correcta lectura de los genotipos.

4.3.2. Polimorfismos de una base (SNP: Single Nucleotide Polymorphism)

4.3.2.1. Elección de los SNPs

Para la elección de los SNPs utilizados en este trabajo se han tenido en cuenta dos criterios considerados fundamentales: la distribución de los mismos por todo el genoma y sus frecuencias génicas en distintas razas.

4.3.2.1.1. Búsqueda en bases de datos y bibliografía existente

A partir de la información obtenida de las bases de datos del NCBI (<http://www.ncbi.nlm.nih.gov/Genbank>) de SNPs que cumplieran los criterios mencionados anteriormente y de los descritos por distintos autores (Laurent & Rodellar 2001; Heaton *et al.* 2002; Werner *et al.* 2004; Roy *et al.* 2006; Ordovas *et al.* 2008), se procedió a realizar la metodología de obtención y caracterización, con el fin de seleccionar los que serían empleados en este trabajo.

4.3.2.1.2. Amplificación de los fragmentos que contienen los SNP

A partir de los números de acceso de las secuencias que contenían los SNP (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=Nucleotide>) se diseñaron oligonucleótidos utilizando el programa Primer Express (*Applied Biosystems*)

para amplificar los fragmentos y posteriormente secuenciarlos con objeto de verificar la presencia del polimorfismo en las razas estudiadas y que en caso de ser elegidos sirvieran como controles en el posterior genotipado.

Se realizó la reacción de PCR siguiendo un protocolo estándar según se ha descrito previamente. La reacción se realizó en 25µl de volumen para amplificar cada uno de los fragmentos (Tabla 7). Se comprobó la reacción de amplificación por electroforesis en gel de agarosa a diferentes porcentajes dependiendo del tamaño del fragmento obtenido.

Tabla 7: Componentes de la PCR.

Reactivos	Volúmenes
H ₂ O	17,8µl
Tampón 10x	2,5µl
MgCl ₂ [50mM]	1µl
dNTPs [1,25mM]	2µl
Cebador F [10µM]	0,3µl
Cebador R [10µM]	0,3µl
Taq polimerasa	0,1µl
ADN [50ng/µl]	1µl
Volumen final	25µl
T_m PCR	60°C

4.3.2.1.3. Purificación de los productos amplificados

Los fragmentos que mostraron una calidad y concentración adecuada se purificaron por el método enzimático utilizando el “ExoSap-it PCR Clean up Protocol” en el cual se combina un buffer especial y la acción de dos enzimas hidrolíticas, la Exonucleasa I y una fosfatasa alcalina (SAP: Shrimp alkaline phosphatase) para eliminar los dNTPs que queden en el medio así como los oligonucleótidos que no son utilizados en la reacción. La Exonucleasa I degrada las cadenas simples que aparecen en la reacción (cebadores residuales y cualquier otro producto extraño de cadena simple que se produzca durante la PCR) y la fosfatasa alcalina hidroliza los dNTPs remanentes de la mezcla de PCR. La mezcla de enzimas se adiciona

directamente al producto de PCR (cada 5 µl de producto de PCR se adicionan 2 µl de ExoSap-IT) combinada con el tampón adecuado y se calienta a 37°C durante 15min; en caso de usar la enzima diluida es conveniente dejarla actuar por más tiempo (aproximadamente durante 40min); finalmente se inactiva la enzima por calentamiento durante 15min a 85°C.

4.3.2.1.4. Secuenciación automática

Una vez obtenido el producto de PCR purificado se realizó la reacción de secuenciación (Sanger *et al.* 1977) empleando el BigDye Terminator v3.1 Cycle Sequencing Kit (*Applied Biosystems*) siguiendo el programa estándar recomendado por Applied Biosystems. La diferenciación de los productos se realizó mediante electroforesis capilar en un secuenciador automático ABI Prism 3130 (*Applied Biosystems*). En esta electroforesis se produce una inyección electrocinética de los fragmentos de menor a mayor tamaño. Al pasar por un punto específico del capilar estas cadenas son excitadas con un láser y dependiendo del fluorocromo que posea el fragmento en posición 3' este emitirá una señal a una longitud de onda determinada que es captada por una cámara. La información recogida por esta cámara se traduce en electroferogramas a partir de los que podemos hacer la lectura de la secuencia de bases.

4.3.2.2. Tipificación de los SNPs

Teniendo en cuenta los criterios descritos para la selección de los SNPs usados en identificación, así como los resultados de las amplificaciones y la verificación de las frecuencias en animales de razas españolas, se eligió un grupo de 20 SNPs para llevar a cabo el estudio en todas las muestras de la experiencia.

La tipificación se ha llevado a cabo mayoritariamente mediante la reacción de minisequenciación, también se realizó PCR alelo específica para el estudio de asociación de SNP con caracteres cuantitativos.

4.3.2.2.1. Minisequenciación o SNaPshot

La técnica se basa en la extensión de oligonucleótidos no marcados en presencia de ddNTPs marcados con fluorescencia. Para llevar a cabo esta metodología es necesaria la realización de una serie de pasos que serán descritos a continuación.

4.3.2.2.1.1. Reacciones de amplificación

Para llevar a cabo el genotipado de estos 20 SNPs, en primer lugar se procedió a la amplificación de los fragmentos que contenían los marcadores. En la Tabla 8 se muestran las secuencias de los cebadores utilizados para la amplificación de los fragmentos, número de acceso del GenBank, así como su localización cromosómica.

Se realizó un gran esfuerzo en la consecución de condiciones óptimas que permitieran la amplificación en el menor número de reacciones posibles. Se consiguió incluir los 20 SNPs en dos reacciones multiplex de PCR (Tabla 9)

Las reacciones de PCR multiplex se realizaron usando en cada caso concentraciones adecuadas de los componentes del QUIAGEN Multiplex PCR kit, que como se ha explicado anteriormente ofrece numerosas ventajas en la realización de reacciones de amplificación de varios fragmentos a la vez. En el caso de los oligonucleótidos se empleó en la mayoría de los casos concentraciones finales de 0,125 μ M y en algunos casos en los que la amplificación del fragmento no era la óptima se empleaban 0,25 μ M del cebador en cuestión. Ambas reacciones múltiplex se llevaron a cabo en volúmenes finales de 6 μ l y en ambos casos el programa de PCR consistió en un paso inicial de desnaturalización a 95 °C durante 15 minutos seguido de 40 ciclos de (95°C 30seg, 60°C 1 min, 72°C 1min) y un paso final de extensión a 72°C durante 30minutos.

Tabla 8: SNPs utilizados, secuencias de los cebadores para amplificar los fragmentos, nº acceso (GenBank) y localización cromosómica.

Nombre/Cromos.	Secuencia	Acceso (GenBank)
SNP_C1_Fsec SNP_C1_Rsec	GTATCTTGCTGTAGCTGTGGTTAAAA ACCCATGGGGATGTCCA	U02564
SNP_C2_Fsec SNP_C2_Rsec	CCAGCTTGTGCTGGCAAT GTATGGATGGCCAGAGGG	AY776154
SNP_C4_Fsec SNP_C4_Rsec	CCTCACAGGCCTTCAGCTG CTATCCTAGCCACCAAGGGTAGA	AF458965
SNP_C5_Fsec SNP_C5_Rsec	GTATTTCCGATGAGCGACTGG GAGTGAGGCAGGAAATCGATG	AF458966
SNP_C6_Fsec SNP_C6_Rsec	AAACAAATCTACTTCAGTGCCTCATG TTCAACAGGCACTCGGGAAT	AF465155
SNP_C8_Fsec SNP_C8_Rsec	CGCCGGTGAGTCTGATCG AGCTTGTTCCTCTGGAGCCA	AF465157
SNP_C10_Fsec SNP_C10_Rsec	GAACAATCTGAACCCAGCTG CTTGCTGAACTTAATCCTTTGGC	AF465158
SNP_C11_Fsec SNP_C11_Rsec	CCTGGTGAAGAGCAAGAAGG AGATCCAATCAGGATGCGGT	AF465160
SNP_C13_Fsec SNP_C13_Rsec	CTCACCATTTCGAATACATCCT TCAACATCACATAGATCCTCCTCA	AF465161
SNP_C17_Fsec SNP_C17_Rsec	CCCAGGACTCCCTGCA CCACCCTGAGCCTGGGC	AF465165
SNP_C18_Fsec SNP_C18_Rsec	GAGCATCTCTAAACCGGATG GTCTGGGAAGAGCCGGCT	AF465167
SNP_C21_Fsec SNP_C21_Rsec	GGCTTTCGTTTCTGTTCTGG TCTGAGTTTGGGTTTGGCT	AF465171
SNP_C23_Fsec SNP_C23_Rsec	GCTCCAACGGCACCAGAG GGAACAGAGGGACAAGAGACTG	AF465172
SNP_C24_Fsec SNP_C24_Rsec	CAGAAGTCTTGACTGTCCGG AGCAGAGGCAATGGTACAGTAATTA	AF465175
SNP_C12_Fsec SNP_C12_Rsec2	GGTATTGTTAGGCAATGGGGA GGAGCTGGCCCTTCACTTG	AJ496770
SNP_C20_Fsec SNP_C20_Rsec	TCCAAAAGAGTCACAGCCAAGAAC AGGCAAAAGAAGATGGCAGCAG	AJ505155
SNP_C26_Fsec SNP_C26_Rsec2	TTGACAAAGGCACAGACGTTG TGGGAGTCAGGAGTCTTCTAGCA	AJ496781
SNP_C29CAPN_F SNP_C29CAPN_R	CACCACACCCTTGCTGCC GCGCCATTACTATCGATCGC	AF248054
SNP_C7Ex3_F SNP_C7Ex3_R	AGGGGCACCTGTCCATCTTC ACGAGGGGACAGGATAGGCA	AY738458.
SNP_C19Ex1_R SNP_C19Ex1_F	AAGCTGTCAGCCCATGTG GCAGGCTGTGTGAGATGAG	AF285607

Tabla 9: SNPs incluidos en cada una de las reacciones de PCR.

Multiplex I	Multiplex II
SNP_C1	SNP_C10
SNP_C2	SNP_C11
SNP_C4	SNP_C13
SNP_C5	SNP_C17
SNP_C6	SNP_C18
SNP_C8	SNP_C12
SNP_C21	SNP_C20
SNP_C23	SNP_C26
SNP_C24	SNP_C29CAPN
SNP_C7Ex3	SNP_C19Ex1

4.3.2.2.1.2. Purificación de los productos

Una vez concluida la reacción de PCR y verificadas por electroforesis en agarosa al 3% la amplificación de los fragmentos en multiplex, se purificó con ExoSap como se ha descrito previamente y se pasó a realizar la reacción de minisequencing o SNaPshot.

4.3.2.2.1.3. Reacción de Minisequenciación o SNaPshot

Estas reacciones se realizaron utilizando el SNaPshot Multiplex Kit (*Applied Biosystems*) y el secuenciador automático ABI Prism 3130. La técnica de minisequencing (Pastinen *et al.* 1996) se basa en la extensión de oligonucleótidos copiando una cadena molde en presencia de ddNTPs marcados con fluorocromos. En este caso y a diferencia de la reacción de secuenciación la cadena se extiende únicamente un nucleótido por lo que se generan picos del mismo tamaño que, dependiendo de si existe polimorfismo o no se pueden traducir en señales fluorescentes diferentes. Es decir, si el SNP es dimórfico y el individuo heterocigótico se generarán dos tipos de fragmentos de igual tamaño pero que al ser excitados por el láser en el secuenciador automático emitirán a dos longitudes de onda diferentes por lo que en el electroferograma aparecerán dos picos de diferente color. (Figura 12)

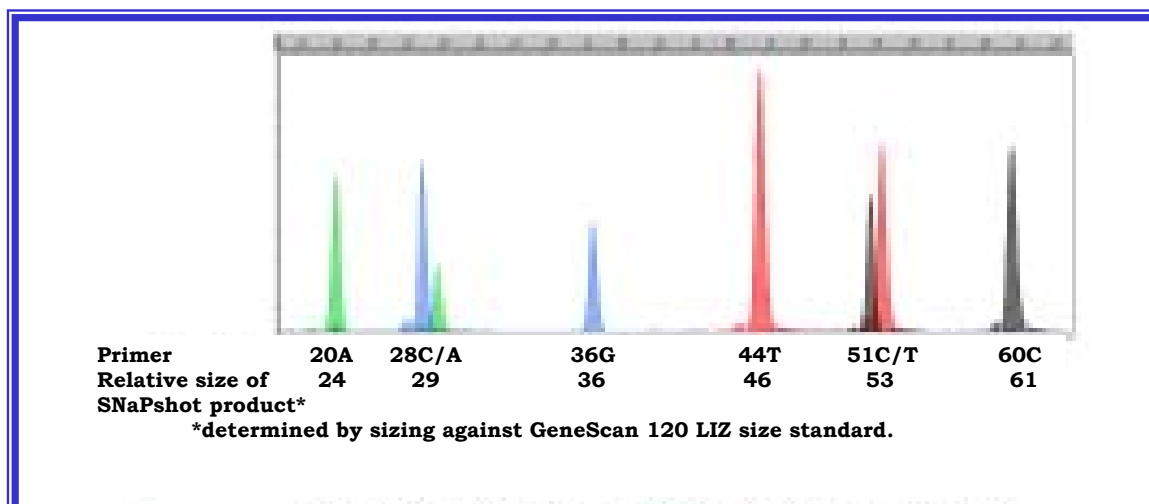


Figura 12: Imagen de reacción de minisequenciación (*Applied Biosystems*).

Para llevar a cabo el genotipado de estos 20 SNPs fueron diseñadas dos reacciones multiplex de minisequencing utilizando mezclas de oligonucleótidos de diferentes tamaños para evitar el solapamiento de los fragmentos, en la mayoría de los casos fue necesario adicionar colas de poliT para obtener fragmentos de diversas longitudes. En la Tabla 10 se muestran las secuencias de los cebadores de minisequencing y sus tamaños relativos, que fueron diseñados de modo que no existiera solapamiento entre fragmentos y teniendo cuidado de mantener la diferencia recomendada de pares de bases entre un fragmento y otro para evitar falsas lecturas de los genotipos.

La reacción de minisequencing se realizó en un volumen de 6µl de los cuales 1µl era del SNaPshot Multiplex Kit (*Applied Biosystems*), 1,5µl de los productos de amplificación puros y concentraciones finales de los oligonucleótidos de 0,2µM; el resto del volumen se completó con H₂O. El programa de PCR seguido consistió en 25 ciclos de (10seg 96°C, 5seg 50°C y 30seg 60°C) en un termociclador ABI 2720 (*Applied Biosystems*).

Tabla 10: Oligonucleótidos de minisequencing empleados en las dos reacciones múltiples.

Nombre primer	Cromos.	SNP	Secuencia primer	Tamaño	Sentido
C1C/T_Rev	1	C/T reverse	(33T)GCAGTGTGGCAGGACTTCTTG	21+33T	Inverso
C2A/G	2	A/Gsnp109	(8T)GACTAGTTCCTCCAGTCTCAGGTCT	24+8T=32	Directo
C4C/G	4	C/Gsnp361	(T)CCAGCCCCTCTCACACCAG	19+T=20	Directo
C5A/C	5	A/C snp261	(15T)GACCAGCCAGCTGAGTGAAGTAG	23+15T=38	Directo
C6A/G	6	A/Gsnp103	(17T)CATGTACTGTGTGGGTCTGGTGTAG	25+17T=42	Directo
C8C/T	8	C/Tsnp300	TTTTCCGTGGAAAAGACTGTAGAGAGCA	24+4T=28	Directo
C21C/G_Rev	21	C/G snp137	(27T)GAGAGGACGGGTCCCTGTC	19+27T=46	Inverso
C23C/T	23	C/Tsnp 338	(24T)CCCCTCTCAAAATCTGAGACTTGTA	26+24T=50	Directo
C24A/G	24	A/G snp 397	(31T)GTGTTTCCCTGTATTTTCCTGCA	23+31T=54	Directo
C10C/T	10	C/T snp396	(9T)TGTCCTGCACTCCCTGTG	19+9T=28	Directo
C11A/G	11	nuevo A/G snp111	(10T)CAGAGCCCAGGCTGGTCC	18+10T=28	Directo
C13A/G	13	A/G snp252	CCTTTGTTGTCTAATGTCAGAACTATAAC TATCTA	36	Directo
C17A/G	17	A/Gsnp302	(T)GTGGCTCCCAGCCATGTCT	19+T=20	Directo
C18G/T	18	G/Tsnp453	(16T)CTCTTCTCAGATTGCCTGATTTC	24+16T=40	Directo
C7Ex3FATP1C/T	7	C/T	(15T)GAGATGTGGGGCCTGATGG	19+15T=34	Directo
C19Ex1FASNC/G	19	C/G	(33T)CAGCCGTTTCGCACAGCC	17+33T=50	Directo
C12A/G	12	A/G snp 107	(37 T)CGGAGACAGCTGGCAGC	17+37 T=54	Directo
C20C/T	20	C/T snp 233	(24T)CCCTGACCTTCACCACCTCC	20+24T=44	Directo
C26 C/T	26	C/T snp 329	(34T)GTCCCTGTTCTGTCCCTGCC	20+34T=54	Directo
C29A/G_CAPN	29	A/G snp 4558	(37T)TGCAGAGAGCTGGATGACCAG	21+37T=58	Directo

4.3.2.2.1.4. Purificación de los productos de minisequenciación

Una vez realizada la reacción de minisequenciación fue necesario purificar los fragmentos para eliminar los ddNTPs residuales que pudieran interferir en la electroforesis capilar. Para esto se adicionó a cada pocillo 1u de SAP (Shrimp alkaline phosphatase) y se sometió la mezcla 1h a 60°C con posterior inactivación de la enzima 15min a 75°C. Un vez finalizado este ciclo se tomó 1µl y se adicionó a una mezcla que contenía 9,7 µl de formamida Hi-Di (altamente desionizada) y 0,3 µl de Standard LIZ 120 (Applied Biosystems), se desnaturalizó 5min a 95°C y se procedió a realizar la electroforesis capilar.

4.3.2.2.1.5. Genotipado

Se asignaron los diferentes genotipos usando el software GeneMapper v3.7 (Applied Biosystems) y usando como controles de lectura muestras previamente secuenciadas en el laboratorio.

4.3.2.2.2. PCR alelo específica (ASO-PCR: Allele Specific Oligonucleotide-PCR)

Este tipo de PCR se utilizó para estudiar la asociación del SNP A/G en la posición 1455 de la secuencia del mRNA de la transferrina bovina con número de acceso de GenBank (U02564) (Laurent & Rodellar 2001). Este polimorfismo está ubicado en el exón 12 del gen de la transferrina bovina y que produce un cambio aminoacídico Asp/Gly.

ASO-PCR es una variante de la PCR estándar, en la que se diseñan dos oligonucleótidos en un sentido que difieren en 3' dependiendo del polimorfismo que se quiere testar y un cebador común en el otro sentido. La reacción se realiza en dos pocillos cada uno con una pareja de oligos y se puede realizar en equipo de PCR estándar o en PCR a tiempo real en ABI 7000 (Applied Biosystems). Las secuencias de los oligonucleótidos se muestran en la Tabla 11.

Tabla 11: Oligonucleótidos para PCR alelo específica.

Nombre del Primer	Secuencia	Tm (°C)
C1_TF_F	GGGTATCTTGCTGTAGCTGTGG	57,5
C1_TF_R2	CAGTGTGGCATGACTTCTTG T	54,9
C1_TF_RevC1	AGTGTGGCATGACTTCTTG C	54,9

4.3.2.2.2.1. PCR alelo específica estándar

La ASO-PCR estándar se realizó en 25µl de volumen final, 0,06µM de cada oligonucleótido, 50ng de ADN, concentraciones adecuadas de tampón de PCR, MgCl₂, dNTPs y 0,125 µl de Taq Polimerasa (*Invitrogen*). Se siguió un programa de PCR en dos pasos que consistía en un paso inicial de desnaturalización y posteriormente 27 ciclos de 15seg a 61°C. Finalmente se realizó la electroforesis en agarosa al 3% para la lectura de los genotipos.

4.3.2.2.2.2. PCR alelo específica en RT-PCR

La PCR a tiempo real se realizó en un equipo ABI 7000 (*Applied Biosystems*) usando la metodología SYBR green que es un agente intercalante que se une con gran afinidad al surco menor del ADN bicatenario y se acopla energéticamente a él. A medida que va transcurriendo la reacción de amplificación y se van formando nuevas moléculas bicatenarias este agente se va intercalando y emitiendo fluorescencia la cual puede ser monitorizada por el equipo. Se empleó el Master Mix SYBR Green 2x a concentración final 1x en 10µl de volumen final; 0,3µM de cada primer y 50ng de ADN.

Se utilizó un programa a dos pasos de RT-PCR que consistió en un paso inicial de desnaturalización a 95°C durante 10 minutos y 40 ciclos de 15seg a 95°C y 30seg a 63°C. La lectura de los genotipos se pudo hacer a tiempo final mediante la lectura de la fluorescencia en cada ciclo de la PCR.

4.4. Herramientas y metodología informáticas aplicadas

4.4.1. Herramientas informáticas

Los programas que se han empleado son los siguientes:

- ENTREZ (<http://www.ncbi.nlm.nih.gov>) para búsqueda de información.
- BLAST (<http://www.ncbi.nlm.nih.gov/BLAST>) para alineamiento de secuencias.
- Primer Express Software v2.0.0 (*Applied Biosystems*) para diseño de oligonucleótidos.
- BioEdit v7.0.5.3 (Hall 1999), Sequencing Analysis v5.2 (*Applied Biosystems*) para análisis de secuencias.
- 7000 system SDS software (*Applied Biosystems*) para análisis de fragmentos amplificados por PCR en tiempo real.
- GeneScan v3.7 (*Applied Biosystems*), GeneMapper v3.7 (*Applied Biosystems*) para análisis de fragmentos.
- Genepop v3.1 (Raymond & Rousset 1995) para caracterización genética.
- Paquete estadístico SPSS 14.0 (<http://www.spss.com>) para análisis estadísticos.
- MICROSAT (Minch 1995) para cálculo de distancias genéticas.
- PHYLIP (Felsestein 1995) y TREEVIEW (Page 1996) para construcción de árboles filogenéticos.
- Wichrun v4.1 (Banks & Eichert 2000) para asignación de individuos a raza.

4.4.2. Aplicación web diseñada para gestión de los datos

La información correspondiente a todos los animales incluidos en el estudio, así como el perfil genético obtenido con ambos paneles de marcadores moleculares, fue almacenada en la base de datos del laboratorio vinculada a una aplicación web especialmente diseñada para la gestión de los datos obtenidos en la presente tesis doctoral, que será descrita a

continuación. Esta aplicación web se desarrolló en lenguaje ASP sobre un servidor Microsoft Internet Information Service 5.0 con soporte para bases de datos Microsoft Access. Las consultas a la base de datos se realizaron en SQL.

La introducción y gestión de los registros se realizó a través de un formulario implementado localmente en la base de datos. La aplicación web permite el acceso a dicha base de datos e incluye tres apartados:

- Gestión de datos: mediante formularios HTML se facilita la realización de consultas y/o modificaciones de los diferentes registros, así como la comprobación de errores en la introducción de los datos correspondientes a los loci, ya que los algoritmos de cálculo implementados en los otros apartados precisaban que estuvieran en orden alfanumérico creciente.

- Filiación: este apartado incluye diversas funciones de búsqueda de coincidencias (teniendo en cuenta microsatélites y SNPs), de búsqueda de progenitores o descendientes compatibles y de búsqueda de progenitores para un individuo dado un progenitor indubitado (estas dos últimas funciones teniendo en cuenta sólo los microsatélites). En todos los casos se ofrece la alternativa de hacer la búsqueda para un individuo o para un perfil genético.

- Análisis: en esta sección se implementaron diversos algoritmos de cálculo basados en las frecuencias alélicas de los microsatélites y SNPs (la visualización de las frecuencias alélicas también se incluye como función específica) con el fin de calcular la probabilidad de identidad teórica (Weir 1996) y la heterocigosidad. Igualmente se diseñaron algoritmos para el cálculo empírico de la probabilidad de identidad. La aplicación permite la realización de las estimaciones y los cálculos sobre subconjuntos de individuos seleccionados en función de la raza, rebaño, sexo, tipo de muestra y tipo de animal (progenitor o descendiente) considerando un conjunto de marcadores moleculares definido por el usuario.

4.4.3. Caracterización de los marcadores microsatélites y SNPs

Para la caracterización de ambos tipos de marcadores se realizó el cálculo de frecuencias alélicas, genotípicas, análisis del equilibrio de Hardy-Weinberg usando el programa GENEPOP v.3.1 (Raymond & Rousset 1995). Con estos valores de frecuencias se estimaron los promedios de heterocigocidades (H) y el Contenido de Información Polimórfica (PIC) (Botstein *et al.* 1980), parámetros que nos dan una idea de la calidad del marcador para su aplicación en identificación. Estos cálculos se realizaron en una hoja de cálculo Excel a partir de los resultados del Genepop, utilizando las fórmulas:

$$H = 1 - \sum_{i=1}^{i=n} p_i^2$$

$$PIC = 1 - \left(\sum_{i=1}^n p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

siendo p_i y p_j las frecuencias de los alelos i y j , respectivamente, y n el número total de alelos.

4.4.3.1. Análisis del potencial de los diferentes marcadores en identificación y trazabilidad

Con el objetivo de verificar el potencial de cada tipo de marcador en su aplicación a estudios de identificación y trazabilidad se realizaron los cálculos de las probabilidades de exclusión y las probabilidades de identidad con los dos grupos de marcadores utilizados.

4.4.3.1.1. Probabilidad de exclusión

La probabilidad de exclusión se define como la habilidad de un marcador o de un set de marcadores de excluir una relación de parentesco dada, y es determinada a partir de los genotipos, las frecuencias alélicas en los diferentes loci y el número de loci independientes testados (Jamieson &

Taylor 1997). Estas probabilidades fueron calculadas asumiendo tres situaciones diferentes usando tres fórmulas expresadas como potencias de frecuencias, empleadas para los cálculos cuando se trabaja con marcadores codominantes.

1) Dados dos progenitores y un descendiente: excluir un progenitor. La fórmula se obtiene a partir de la descrita por Jamieson en 1965 (Jamieson 1965), expresada como potencia de las frecuencias (Jamieson 1994)

$$\mathbf{PEP1} = 1 - 2\sum_{i=1}^n p_i^2 + \sum_{i=1}^n p_i^3 + 2\sum_{i=1}^n p_i^4 - 3\sum_{i=1}^n p_i^5 - 2(\sum_{i=1}^n p_i^2)^2 + 3\sum_{i=1}^n p_i^2 \sum_{i=1}^n p_i^3$$

2) Dado un progenitor y un descendiente, excluir la relación entre ellos. Un ejemplo de esto es la falta de genotipos de uno de los progenitores. La fórmula se obtiene a partir de la descrita por Garber y Morris en 1983 (Garber & Morris 1983), expresada como potencia de las frecuencias (Jamieson & Taylor 1997)

$$\mathbf{PEP2} = 1 - 4\sum_{i=1}^n p_i^2 + 2(\sum_{i=1}^n p_i^2)^2 + 4\sum_{i=1}^n p_i^3 - 3\sum_{i=1}^n p_i^4$$

3) Dados dos progenitores y un descendiente: excluir ambos progenitores. La fórmula se obtiene a partir de la descrita por Grundel y Reetz en 1981 (Grundel & Reetz 1981), expresada como potencia de las frecuencias (Jamieson & Taylor 1997)

$$\mathbf{PEP3} = 1 + 4\sum_{i=1}^n p_i^4 - 4\sum_{i=1}^n p_i^5 - 3\sum_{i=1}^n p_i^6 - 8(\sum_{i=1}^n p_i^2)^2 + 8(\sum_{i=1}^n p_i^2)(\sum_{i=1}^n p_i^3) + 2(\sum_{i=1}^n p_i^3)^2$$

El poder de exclusión total fue calculado según:

$$PE_{BC} = 1 - \prod_{i=1}^M (1 - PE_{BCi})$$

donde PE_{BCi} es el PE del marcador i y PE_{BC} es el PE usando todos los M marcadores.

4.4.3.1.2. Probabilidad de identidad

La evaluación del poder de los microsatélites y de los SNPs para discriminar entre diferentes individuos se realizó mediante el cálculo de la probabilidad de identidad (Weir 1996) que se define como la probabilidad de que al escoger dos individuos al azar estos compartan los mismos perfiles genéticos. El cálculo se hizo según la fórmula:

$$PI = \prod_{k=1}^m \left(\sum_{i=1}^n (p_{ki})^4 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n (2p_{ki}p_{kj})^2 \right)$$

donde m es el número de loci; n el número de alelos del locus k y $p_{ki(j)}$ es la frecuencia del alelo $i(j)$ del locus k .

4.4.3.2. Análisis estadísticos

La inferencia estadística nos permite realizar estimaciones de variables en los descendientes a partir de determinados parámetros de los progenitores.

A partir de la información introducida en la base de datos y utilizando el paquete estadístico SPSS 14.0 para Windows se llevaron a a cabo los análisis de correlación de diferentes variables calculadas en descendientes y progenitores mediante los coeficientes de correlación de Pearson. Con las variables que mostraran estar correlacionadas se desarrollarían los análisis de regresión para intentar explicar el comportamiento de la variable correlacionada de los descendientes en función de los progenitores.

4.4.3.2.1. Correlación

La correlación indica el grado y la dirección de la relación existente entre dos variables aleatorias.

Existen diversos coeficientes que miden el grado de correlación, adaptados a la naturaleza de los datos. En este trabajo hemos utilizado el coeficiente de correlación de Pearson, que es un índice estadístico que mide la relación lineal entre dos variables cuantitativas y se obtiene dividiendo la covarianza de dos variables por el producto de sus desviaciones estándar.

4.4.3.2.2. Regresión

Como en el caso del análisis de correlación, la regresión lineal simple es una técnica para analizar la relación entre dos variables aleatorias continuas. Nos permite investigar el cambio en una de las variables, denominada respuesta, que corresponde a un cambio dado en la otra, conocida como variable explicativa. El objetivo primordial de análisis de regresión consiste en predecir o aproximar el valor de la respuesta asociado a un valor fijo de la variable explicativa.

En estadística la regresión lineal o ajuste lineal es un método matemático que modeliza la relación entre una variable dependiente Y , las variables independientes X_i y un término aleatorio ε . Este modelo puede ser expresado como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

donde β_0 es la intersección o término "constante", las β_i son los parámetros respectivos a cada variable independiente, y p es el número de parámetros independientes a tener en cuenta en la regresión.

4.4.3.2.3. Coeficientes de determinación

Para comprobar el ajuste del modelo a los datos observados se calcula el coeficiente de determinación (R^2) que es una medida de la bondad de

ajuste del modelo de regresión hallado. El coeficiente de determinación es el cuadrado del coeficiente de correlación de Pearson y sus valores estarán entre 0 y 1, por lo que cuanto más cercano a 1 sea este valor mayor será el ajuste del modelo. R^2 indica qué porcentaje de la variabilidad de la variable de respuesta Y es explicada por su relación lineal con X.

4.4.3.2.4. Selección de variables en regresión múltiple

Una buena propiedad de un modelo de regresión lineal es que permita explicar el comportamiento de la variable de respuesta Y lo mejor posible, haciendo uso del menor número de variables predictoras posibles, esta propiedad es llamada “parsimonia”. Existen dos métodos generales de lograr este objetivo: los métodos “stepwise” y el método de los mejores subconjuntos. Entre los métodos "stepwise" están:

- Método de eliminación hacia atrás (“Backward Elimination”)
- Método de Selección hacia adelante (“Forward Selection”)
- Método Paso a Paso ("Stepwise")

En este trabajo se ha utilizado el método Paso a Paso ("Stepwise"), donde una variable que ha sido incluida en el modelo en un paso previo puede ser eliminada posteriormente. En cada paso se cotejan si todas las variables que están en el modelo deben permanecer allí.

4.4.4. Asignación de individuos

La asignación probable de individuos a raza o a una población determinada puede resultar una herramienta muy útil en la verificación de la trazabilidad. A partir de este método probabilístico se podría detectar alguna acción fraudulenta en el sistema de trazabilidad.

Uno de los métodos empleados para valorar el poder de la técnica en la asignación de individuos para verificar la trazabilidad se ha basado en el empleo del programa WHICHRUN v4.1 (Banks & Eichert 2000), que permite la asignación de individuos a la población más probable utilizando datos

genotípicos multilocus basado en funciones de máxima verosimilitud. De esta manera podríamos valorar la probabilidad de que una muestra determinada pertenezca a una raza o incluso a una población que forme parte de un lote determinado de carne. Para esto es imprescindible contar con datos de frecuencias alélicas, por lo que en este trabajo para estimar las probabilidades de asignación de individuos se emplearon además resultados de frecuencias previamente obtenidos en nuestro laboratorio con otras razas como Casta Navarra, Betizu, Serrana de Teruel, Mallorquina, Menorquina, Monchina, Pirenaica y Toro De Lidia (Martin-Burriel *et al.* 1999; Martin-Burriel *et al.* 2007; Sanz *et al.* 2007).

Por otra parte, a partir del análisis de los alelos compartidos (Bowcock *et al.* 1994) y considerando cada animal de forma individual como unidad taxonómica operativa, se construyó una matriz de distancias utilizando el programa MICROSAT (Minch 1995), y se construyó un árbol con la metodología Neighbor-Joining (Saitou & Nei 1987) mediante el programa PHYLIP (Felsestein 1995). La visualización del dendograma se realizó con el programa TREEVIEW (Page 1996). Este método gráfico nos permite visualizar la diversidad genética de las poblaciones y realizar la asignación de individuos a su raza.

5. RESULTADOS Y DISCUSIÓN

5. RESULTADOS Y DISCUSIÓN

En este capítulo se presentan los resultados obtenidos para dar cumplimiento a los objetivos planteados en este trabajo. Estos incluyen el análisis de la utilidad de dos paneles de marcadores moleculares para ser empleados como método de verificación de trazabilidad, diseño y aplicación de mecanismos adicionales para gestionar el sistema de trazabilidad, así como resultados complementarios obtenidos en el análisis de algunos marcadores estudiados.

5.1. Diseño de herramientas informáticas para identificación de individuos y trazabilidad

Como se ha explicado en el capítulo de material y métodos, todas las muestras recibidas para el estudio de trazabilidad fueron incluidas en una base de datos desarrollada en Microsoft Access 2003 con vínculo a una aplicación web que se desarrolló en colaboración con Dr. Ignacio de Blas de la Unidad de Epidemiología de la Facultad de Veterinaria de la Universidad de Zaragoza. Esta aplicación permite gestionar el sistema de trazabilidad genética a través del enlace www.revistaaquatic.com/trazas al cual solo se puede acceder introduciendo nombre de usuario y contraseña de manera que se garantice la seguridad de los datos.

En la Figura 13 se puede ver una imagen de la interface diseñada para gestionar el sistema. Esta aplicación ofrece diferentes posibilidades como son:

- Gestión de datos
- Análisis de filiación
- Análisis de parámetros genéticos
- Realización de búsquedas específicas

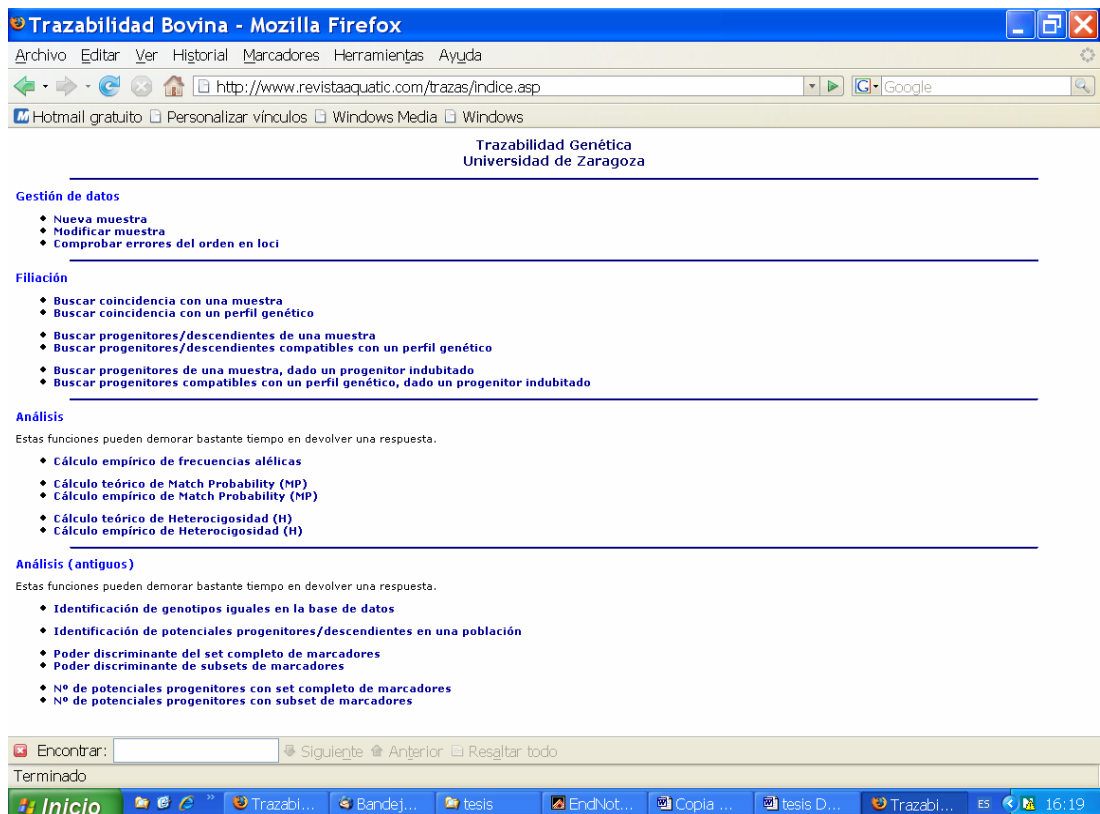


Figura 13: Interface para gestionar el sistema de trazabilidad de LAGENBIO.

A continuación analizaremos cada una de las posibilidades ofrecidas por la aplicación.

5.1.1. Gestión de datos

En el apartado de gestión de datos es posible realizar la inclusión de nuevas muestras o modificar las existentes. Todas las muestras que forman parte de esta base de datos se identifican con un código interno de laboratorio y están acompañadas de toda la información disponible como número de identificación, raza, sexo, categoría, propietario, tipo de muestra, explotación, fecha de nacimiento, fecha de recepción y datos de los progenitores en todos los casos que estén disponibles. Es posible además incluir los perfiles genéticos obtenidos para los paneles de marcadores moleculares empleados en este trabajo, así como observaciones en el caso de que existan.

Se puede modificar una muestra existente accediendo mediante el código interno de laboratorio asignado y como análisis único es posible comprobar si existen errores en el orden de los loci ya que este punto es de vital importancia para la realización de análisis posteriores.

5.1.2. Análisis de filiación

Este apartado de filiación permite realizar la búsqueda de coincidencias, de progenitores o descendientes y la búsqueda del segundo progenitor dado uno indubitado. Estos análisis se pueden llevar a cabo mediante la selección de una muestra o de un perfil genético.

I- Búsqueda de coincidencias con:

- una muestra: introduciendo el número de identificación de una muestra e indicando un número de discordancias admitidas definidas por el usuario para microsatélites y/o SNPs, podemos realizar la búsqueda en la base de datos y obtener un listado de las muestras coincidentes que comparten los mismos requisitos.
- un perfil genético: a partir de un perfil genético definido por el usuario es posible realizar la búsqueda en la base de datos de los individuos que compartan los mismos genotipos para los marcadores seleccionados con un número de discordancias permitidas. Cuando no se seleccionan todos los marcadores del panel en la definición del perfil genético, se debe poner como mínimo que el número de discordancias debe ser igual al número de marcadores en blanco.

En general este apartado permite buscar el origen de una muestra de carne, contando con los genotipos de las muestras de carne y las sangres correspondientes, y por tanto verificar el sistema de trazabilidad.

II- Búsquedas de progenitores o descendientes compatibles con:

- una muestra: seleccionando una muestra mediante su número de identificación e indicando un número de discordancias admitidas

definidas por el usuario, es posible realizar la búsqueda de las muestras que compartan al menos un alelo por locus (0 discordancias admitidas). El número de loci a tener en cuenta en la búsqueda depende del número de discordancias admitidas definidas por el usuario.

- un perfil genético: a partir de un perfil genético definido por el usuario es posible comprobar todas las muestras en la base de datos que compartan al menos uno de los alelos para cada locus que forma parte del panel. Indicando el número de discordancias admitidas podemos excluir de la búsqueda algunos loci.

Esta aplicación es de gran utilidad pues a partir del perfil genético de una muestra de carne problema, sería posible realizar la búsqueda de los posibles progenitores mediante la comparación de perfiles genéticos. En caso de que se contara con datos genotípicos de progenitores de varias poblaciones, se podría hacer una aproximación a la población de origen de la muestra problema. Esto sería posible sin contar con los perfiles genéticos obtenidos a partir de cualquier otra muestra del animal que fue a matadero, por lo que sería una ventaja en el sentido de disminuir los costes de verificación del sistema de trazabilidad. En general esta opción permite realizar la asignación de posibles progenitores.

III- Búsqueda del segundo posible progenitor, dado un progenitor indubitado:

- de una muestra: seleccionando una muestra mediante su número de código interno que contenga reportes de progenitores e indicando un progenitor indubitado (obtenido a partir de los datos depositados en la base), permite listar los posibles progenitores compatibles.
- compatibles con un perfil genético: introduciendo un perfil genético y un progenitor indubitado es posible buscar el otro progenitor, mediante la comparación de alelos.

Esta es una aplicación muy útil en análisis de paternidad, pues permite realizar la asignación del segundo progenitor partiendo de uno indubitado. Este análisis solamente se puede llevar a cabo en los casos donde se cuente con todos los datos de progenitores de los descendientes.

5.1.3. Análisis de parámetros genéticos

Esta opción nos permite realizar cálculos de parámetros genéticos como:

- Frecuencias alélicas.
- Probabilidad de Identidad teórica (Match Probability teórica).
- Probabilidad de Identidad empírica (Match probability empírica): se realiza a partir del conteo de los pares que compartan los mismos genotipos y se divide por todas las combinaciones posibles.
- Heterocigosidad esperada.
- Heterocigosidad observada.

5.1.4. Realización de búsquedas específicas

- Identificación de genotipos iguales en la base de datos: esta opción nos permite buscar por poblaciones las muestras que tienen en toda la base de datos perfiles genéticos coincidentes ya sea dentro de la misma población o en las otras poblaciones. Esta opción nos permite conocer en general si existen muestras que están duplicadas o hacer un análisis general de los posibles errores en el genotipado de las muestras.
- Identificación de potenciales progenitores o descendientes de una población: seleccionando una población y el sexo del progenitor (para los casos de identificación de posibles descendientes) es posible obtener el número de descendientes que comparten al menos un alelo por locus con cada progenitor de la población.

En general el sistema diseñado permite gestionar toda la información obtenida. Está siendo de vital importancia pues nos permite llevar un control

desde la recepción de la muestra hasta la obtención de los perfiles genéticos, facilitando la verificación del sistema de trazabilidad así como el cálculo de parámetros genéticos.

5.2. Estimación de la eficacia de los marcadores utilizados en identificación y trazabilidad

El trabajo de identificación y verificación de trazabilidad se ha llevado a cabo mediante el empleo de los marcadores moleculares microsatélites y SNPs, por lo que en este apartado se expondrán todos los resultados obtenidos usando ambos tipos de marcadores con el fin de estimar su eficacia. En primer lugar se describirán los resultados para los microsatélites, posteriormente los obtenidos para los SNPs y por último se hará una comparación entre ambos.

5.2.1. Microsatélites

5.2.1.1. Estimación de la variabilidad genética

Las 675 muestras de sangre y músculo que forman parte del estudio fueron amplificadas satisfactoriamente para los 9 microsatélites utilizados en el presente estudio, excepto dos de ellas en las que fue imposible realizar la amplificación debido a la falta de ADN. En la Figura 14 podemos ver una imagen donde aparecen los 9 microsatélites amplificados.

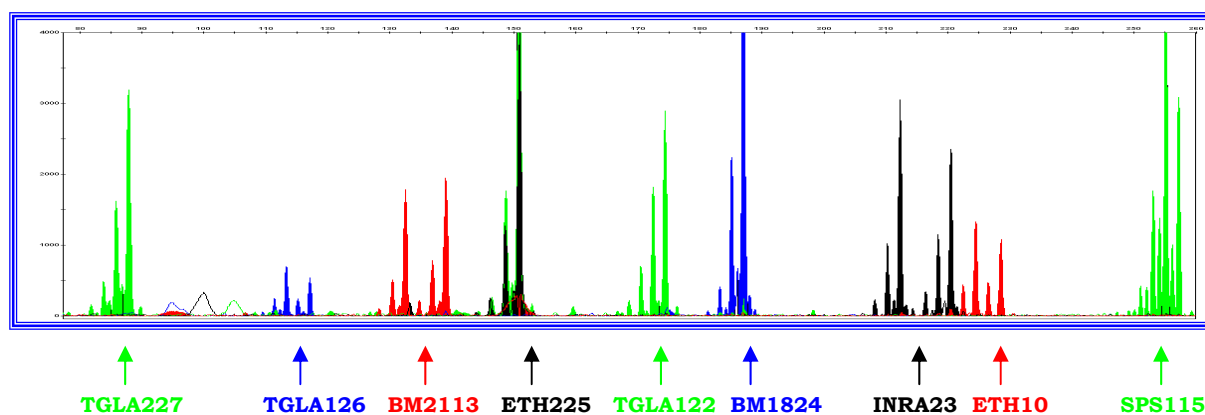


Figura 14: Electroferograma de una reacción de amplificación para los 9 microsatélites empleados en estudios de identificación.

Todos los loci analizados fueron polimórficos, detectándose un total de 73 alelos. El número de alelos por locus varió de 4, en el locus TGLA126, a 12 en el TGLA227 que fue el más polimórfico en todas las poblaciones analizadas (Tabla 12). Resultados similares fueron obtenidos en estudios realizados en IGP “Carne de Asturias” (Vazquez *et al.* 2004).

Tabla 12: Número de alelos por locus por razas y por explotaciones.

Nº. Alelos/locus									
Población	ETH225	ETH10	INRA23	BM2113	BM1824	SPS115	TGLA227	TGLA126	TGLA122
Parda de Montaña	7	5	9	8	5	7	10	4	10
Pirenaica I	5	5	6	6	4	4	8	4	6
Pirenaica II	5	5	8	9	5	6	10	4	9
Pirenaica I+II	5	5	8	9	5	6	10	4	9
Total alelos/locus	7	6	10	9	6	8	12	4	11

Las frecuencias alélicas fueron calculadas para las poblaciones de raza Parda de Montaña y Pirenaica, también se realizaron los cálculos teniendo en cuenta las explotaciones de las que procedían las muestras de raza Pirenaica (Pirenaica I y Pirenaica II). Los resultados se muestran en la Tabla 13.

A partir de las frecuencias alélicas se construyeron los histogramas para ilustrar gráficamente su distribución por locus y por población (Figura 15). Se puede observar para la mayoría de los loci una distribución variable de las frecuencias alélicas con comportamientos similares en las diferentes poblaciones. Algunos loci como el TGLA126, SPS115, BM1824 o ETH225 muestran alelos prevalentes, siendo el caso más notorio el del alelo 115 en el TGLA126 que muestra una frecuencia media superior de 0,6.

El equilibrio de Hardy-Weinberg ha sido testado utilizando el programa GENEPOP v3.1 (Raymond & Rousset 1995) y se ha visto que las poblaciones estudiadas están en equilibrio genético para la totalidad de los loci analizados.

Tabla 13: Frecuencias alélicas por locus para cada población.

Locus	Población	Alelos / Frecuencias											
ETH225		140	144	146	148	150	152	156					
	Parda de Montaña	0,150	0,189	0,128	0,044	0,444	0,006	0,039					
	Pirenaica I	0,288	0,063	0,025	0,100	0,525	-	-					
	Pirenaica II	0,094	0,044	0,275	0,125	0,463	-	-					
	Total	0,155	0,110	0,164	0,086	0,467	0,002	0,017					
ETH 10		213	215	217	219	221	223						
	Parda de Montaña	0,011	-	0,227	0,286	0,195	0,281						
	Pirenaica I	-	0,051	0,127	0,278	0,456	0,089						
	Pirenaica II	-	0,013	0,157	0,390	0,428	0,013						
	Total	0,005	0,014	0,182	0,324	0,331	0,144						
INRA23		198	200	206	208	210	212	214	216	218	222		
	Parda de Montaña	0,029	0,006	0,437	0,115	0,080	0,046	0,276	0,006	-	0,006		
	Pirenaica I	0,071	0,176	0,282	0,106	-	0,082	0,282	-	-	-		
	Pirenaica II	0,035	0,503	0,211	0,058	-	0,029	0,105	0,053	0,006	-		
	Total	0,040	0,237	0,316	0,091	0,033	0,047	0,209	0,023	0,002	0,002		
BM2113		125	127	131	133	135	137	139	141	143			
	Parda de Montaña	0,017	0,156	0,061	0,223	0,123	0,022	0,385	0,011	-			
	Pirenaica I	-	0,044	0,122	0,067	0,200	0,256	0,311	-	-			
	Pirenaica II	0,012	0,037	0,098	0,049	0,055	0,055	0,354	0,329	0,012			
	Total	0,012	0,088	0,088	0,125	0,113	0,083	0,358	0,129	0,005			
BM1824		178	180	182	186	188	190						
	Parda de Montaña	0,104	0,277	0,260	0,017	0,341	-						
	Pirenaica I	0,159	0,171	0,378	-	0,293	-						
	Pirenaica II	0,145	0,127	0,188	-	0,533	0,006						
	Total	0,131	0,198	0,255	0,007	0,407	0,002						
SPS115		248	250	252	254	256	258	260	262				
	Parda de Montaña	0,391	-	0,245	0,049	0,207	0,005	0,098	0,005				
	Pirenaica I	0,432	-	0,185		0,210	-	0,173	-				
	Pirenaica II	0,516	0,019	0,097	0,013	0,006	-	0,348	-				
	Total	0,445	0,007	0,179	0,026	0,133	0,002	0,205	0,002				
TGLA227		77	79	81	83	85	87	89	91	93	95	97	99
	Parda de Montaña	-	0,005	0,160	0,021	-	0,197	0,197	0,218	0,048	0,027	0,122	0,005
	Pirenaica I	0,011	-	0,172	0,023	-	0,046	0,287	0,184	0,172	-	0,103	-
	Pirenaica II	0,006	-	0,017	0,229	0,011	0,017	0,103	0,137	0,423	-	0,046	0,011
	Total	0,008	-	0,069	0,160	0,008	0,027	0,164	0,153	0,340	-	0,065	0,008
TGLA126		115	117	121	123								
	Parda de Montaña	0,470	0,341	0,073	0,116								
	Pirenaica I	0,657	0,045	0,104	0,194								
	Pirenaica II	0,814	0,068	0,068	0,051								
	Total	0,622	0,192	0,077	0,109								
TGLA122		141	143	147	149	151	153	161	167	175	177	181	
	Parda de Montaña	0,032	0,279	0,058	0,011	0,168	0,311	0,084	0,005	0,005	0,047	-	
	Pirenaica I	-	0,310	0,056	0,028	0,563	0,028	0,014	-	-	-	-	
	Pirenaica II	0,254	0,304	0,033	0,011	0,232	0,105	0,006	0,011	-	-	0,044	
	Total	0,183	0,306	0,040	0,016	0,325	0,083	0,008	0,008	-	-	0,032	

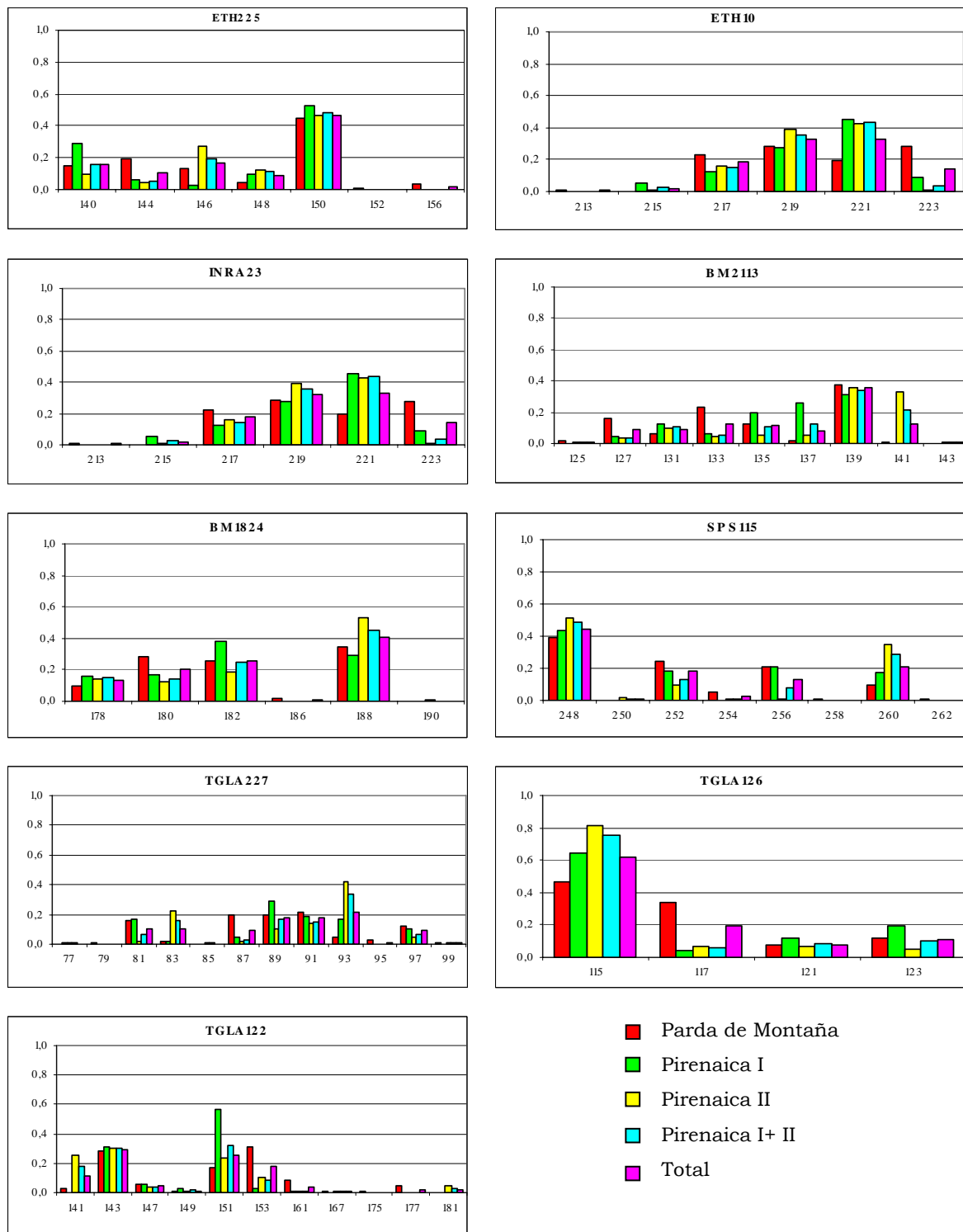


Figura 15: Distribución de las frecuencias alélicas por locus y por población. En el eje de las abcisas se muestran los alelos y en las ordenadas las frecuencias.

En la Tabla 14 se muestran los resultados del contenido de información polimórfica (PIC), heterocigosidad observada (H_o) y heterocigosidad esperada (H_e) por locus en las diferentes poblaciones analizadas.

Respecto a los valores de contenido de información polimórfica (PIC), podemos considerar todos los loci altamente informativos con valores superiores a 0,5. El locus TGLA126 fue el menos informativo en todos los grupos, mostrando valores por debajo de 0,5 en la subpoblación II de Pirenaica y en la Pirenaica I+II, lo que puede estar provocado por la distribución de sus frecuencias alélicas, que muestra una elevada frecuencia del alelo 115. El locus más informativo ha sido el TGLA227 con valores en el rango de 0,7287 en la Pirenaica II a 0,8431 en la población total.

Analizando los valores individuales de H_o por locus en los diferentes grupos encontramos los valores más bajos para el TGLA126 y los más altos para el TGLA227. En general se encuentran para la mayoría de los loci los valores más bajos en la subpoblación II de Pirenaica, lo que puede ser debido a la utilización de un número muy pequeño de machos.

Los resultados de las H_e por locus en los grupos sin considerar la población total, muestran un rango de variación con el mínimo para el TGLA126 en la subpoblación II de Pirenaica (0,3287) y el máximo para el TGLA227 (0,831) en la población de Parda de Montaña. En general podemos decir que las heterocigosidades esperadas fueron altas (superiores a 0,6) para todas las poblaciones y loci, excepto para el locus TGLA126 donde varió de 0,3287 en la subpoblación Pirenaica II a 0,6442 en la población de Parda de Montaña. Estos valores de diversidad genética son indicativos del alto grado de polimorfismo existente en estas poblaciones, lo que favorece la correcta identificación individual.

Tabla 14: Valores de contenido de información polimórfica (PIC), heterocigosidades observadas (Ho) y esperadas (He) por locus para las diferentes poblaciones.

Población	Parám.	ETH225	ETH10	INRA23	BM2113	BM1824	SPS115	TGLA227	TGLA126	TGLA122	Medias (d.e)
Parda de Montaña	PIC	0,7208	0,7287	0,7048	0,7549	0,7157	0,7136	0,8240	0,5914	0,7760	0,7256 (0,06)
	Ho	0,7476	0,7961	0,6893	0,7379	0,6796	0,7864	0,8252	0,5922	0,8447	0,7443 (0,08)
	He	0,7244	0,7494	0,7104	0,7572	0,7279	0,7324	0,831	0,6442	0,7835	0,74 (0,05)
Pirenaica I	PIC	0,6210	0,6503	0,7781	0,7581	0,6828	0,6866	0,8021	0,5156	0,5800	0,675 (0,09)
	Ho	0,7021	0,6809	0,8085	0,9149	0,7447	0,7234	0,8511	0,4255	0,5106	0,7069 (0,16)
	He	0,6272	0,6884	0,7864	0,7765	0,7171	0,7051	0,8107	0,5182	0,5816	0,6901 (0,1)
Pirenaica II	PIC	0,6748	0,5768	0,6599	0,7189	0,6209	0,6021	0,7287	0,3202	0,7616	0,6293 (0,13)
	Ho	0,6598	0,6495	0,7629	0,6907	0,7216	0,6082	0,8041	0,2371	0,8763	0,6678 (0,18)
	He	0,6835	0,6412	0,679	0,7425	0,6402	0,6032	0,738	0,3287	0,7729	0,6477 (0,13)
Pirenaica I+II	PIC	0,6808	0,6073	0,7327	0,7824	0,6595	0,6563	0,7910	0,4024	0,7495	0,6735 (0,11)
	Ho	0,6759	0,6621	0,7793	0,7655	0,7241	0,6483	0,8207	0,2966	0,7586	0,6812 (0,16)
	He	0,6885	0,6608	0,7513	0,797	0,6888	0,6573	0,7988	0,4067	0,7577	0,6897 (0,12)
Total	PIC	0,7068	0,6969	0,7730	0,7973	0,7066	0,7093	0,8431	0,5296	0,7882	0,7279 (0,09)
	Ho	0,7056	0,7177	0,7419	0,754	0,7056	0,7056	0,8226	0,4194	0,7944	0,7074 (0,12)
	He	0,7117	0,7314	0,7863	0,8043	0,7131	0,7094	0,8491	0,5587	0,7956	0,74 (0,08)

5.2.1.2. Probabilidades de exclusión

Las probabilidades de exclusión (PE) han sido calculadas tanto por locus como combinadas (Tabla 15). Se ha visto que el locus TGLA126 es el que ha mostrado menor PE en todos los grupos con valores de 0,1811 en Pirenaica II a 0,3778 en Parda de Montaña, y el TGLA227 ha mostrado los mayores valores en todos los grupos.

Las PE combinadas analizando los 9 microsatélites en todos los casos han sido elevadas y superiores a 0,99 por lo que el panel puede ser utilizado para análisis de paternidad (Gomez-Raya *et al.* 2008). Los valores de PE para la población de Pirenaica I+II y para sus subpoblaciones han sido menores comparados con los de Parda de Montaña, debido al bajo valor de heterocigosidad del locus TGLA126 en la raza Pirenaica, hecho que depende de la distribución de alelos en la población (Sherman *et al.* 2004). Además de

la heterocigosidad los valores de PE pueden estar influenciados por el número de loci microsatélites analizados.

Tabla 15: Probabilidades de exclusión por locus y combinada.

Población	ETH225	ETH10	INRA23	BM2113	BM1824	SPS115	TGLA227	TGLA126	TGLA122	PEcomb.
Parda de Montaña	0,509	0,5071	0,4843	0,5483	0,4785	0,5003	0,6613	0,3778	0,5863	0,9987
Pirenaica I	0,3762	0,4477	0,5844	0,5664	0,4652	0,4576	0,6268	0,2932	0,3251	0,9968
Pirenaica II	0,4428	0,3591	0,4699	0,5358	0,4045	0,3308	0,5256	0,1811	0,5644	0,9941
Pirenaica I+II	0,4591	0,3955	0,5425	0,6135	0,4403	0,4067	0,6161	0,2307	0,5406	0,9974
Total	0,4962	0,4873	0,5892	0,6333	0,4656	0,4769	0,6966	0,3294	0,6042	0,9991

A partir de la información de genealogías aportada por los ganaderos, se ha estimado la Probabilidad de Exclusión Parental (Tabla 16), asumiendo tres situaciones diferentes:

- PEP1: excluir un progenitor, conocidos los genotipos de los dos progenitores y el descendiente.
- PEP2: conocidos los genotipos de un progenitor y un descendiente, excluir la relación entre ellos.
- PEP3: excluir ambos progenitores, conocidos los genotipos de los dos progenitores y el descendiente.

Se puede apreciar que para ambas poblaciones las probabilidades de exclusión parental totales para el panel de 9 microsatélites es menor para el caso dos (PEP.2) donde se evalúa la probabilidad de exclusión parental pero teniendo en cuenta solamente los genotipos de uno de los progenitores y del descendiente. Le sigue el valor de PEP.1 donde se evalúa una probabilidad de exclusión familiar contando con los genotipos de los dos progenitores y del descendiente. Finalmente se puede apreciar que el mayor valor de probabilidad de exclusión parental se obtiene para el caso tres (PEP.3) donde ambos progenitores pueden ser excluidos.

Tabla 16: Estimación de la probabilidad de exclusión parental asumiendo tres situaciones diferentes en tríos de las poblaciones de raza Parda de Montaña y de Pirenaica.

	Parda de Montaña	Pirenaica
PEP.1	0,997874247	0,99560148
PEP.2	0,970110007	0,949535599
PEP.3	0,999966762	0,999892308

Estos resultados son esperables y han sido similares a los encontrados para un caso práctico de estimación de probabilidades de exclusión de paternidades en equino (Jamieson & Taylor 1997) y en una población bovina de raza Galloway (Herraez *et al.* 2005).

5.2.1.3. Probabilidades de identidad teóricas y empíricas

5.2.1.3.1. Probabilidad de identidad teórica (PIt)

Una medida de la eficiencia del sistema genético utilizado es la probabilidad de identidad (PI), que se define como la probabilidad de que al escoger dos individuos al azar estos tengan genotipos idénticos. Las PI teóricas por locus se muestran en la Tabla 17, donde han sido dispuestas en orden decreciente dependiendo de su contenido de informatividad, pudiéndose apreciar que el comportamiento varía en las diferentes poblaciones analizadas.

El locus TGLA 126, que según se ha dicho anteriormente es el que tiene un menor número de alelos así como una menor heterocigosidad, es el que ha mostrado en todos los casos un valor más alto de probabilidad de identidad. El comportamiento de los loci más informativos que han mostrado los menores valores de PI ha variado en las diferentes poblaciones tal y como puede observarse en la Tabla 17. Cuando se realiza el cálculo teniendo en cuenta todos los individuos como una única población el orden de informatividad de los locus fue de menor a mayor TGLA227, BM2113, TGLA122 e INRA23. Cabe destacar que estos loci aunque en distinto orden

son los que presentan un menor valor en las poblaciones consideradas por separado.

Tabla 17: Probabilidad de identidad teórica (PI) por locus.

Parda de Montaña		Pirenaica I		Pirenaica II		Pirenaica I+II		Total	
TGLA227	0,051	TGLA227	0,0618	TGLA122	0,087	TGLA227	0,0656	TGLA227	0,0409
TGLA122	0,0775	INRA23	0,0773	TGLA227	0,1025	BM2113	0,0667	BM2113	0,0593
BM2113	0,0925	BM2113	0,0844	BM2113	0,1037	INRA23	0,0956	TGLA122	0,0704
ETH10	0,1085	BM1824	0,1308	INRA23	0,1371	TGLA122	0,0963	INRA23	0,0761
ETH225	0,1108	SPS115	0,1351	ETH225	0,1484	ETH225	0,1374	ETH225	0,1174
SPS115	0,1143	ETH10	0,1448	BM1824	0,174	BM1824	0,1465	ETH10	0,1198
BM1824	0,1239	ETH225	0,1951	ETH10	0,201	ETH10	0,1713	SPS115	0,1265
INRA23	0,1253	TGLA122	0,2401	SPS115	0,2294	SPS115	0,1777	BM1824	0,1311
TGLA126	0,1908	TGLA126	0,2768	TGLA126	0,4665	TGLA126	0,276	TGLA126	0,2385

Estos resultados muestran que los microsatélites no comparten el mismo contenido de polimorfismo entre poblaciones. Estas diferencias vienen dadas por la estructura genética de cada población donde se pueden encontrar diferentes patrones de frecuencias (Orru *et al.* 2006).

Los valores de PI calculados teniendo en cuenta la población total muestran mejores resultados en los loci más informativos, comparados con el mismo valor calculado en cada población. Esto significa que la evaluación de los microsatélites sin tener en cuenta la población puede ocultar el valor real de informatividad de un locus, lo cual puede afectar la fiabilidad del ensayo en animales de determinadas razas cuando se disminuye el número de loci analizados (Orru *et al.* 2006).

Las frecuencias alélicas y el número de alelos fijados determinan el grado de informatividad de cada locus, por lo que la estructura genética de cada población puede influir la fiabilidad de los análisis cuando se reduce el número de loci genotipados, debido a que un mismo locus puede tener diferente contenido de informatividad en distintas poblaciones como consecuencia de las distribuciones no uniformes de los alelos entre poblaciones (Orru *et al.* 2006). Por esta razón, en la evaluación de los microsatélites, hay que tener en cuenta las diferencias en los alelos fijados

entre las poblaciones con la intención de minimizar el efecto de la pérdida de informatividad.

En la Tabla 18 se pueden apreciar los valores de PI teóricos calculados usando combinaciones de 4 a 9 microsatélites que muestren los mejores valores de PI. En la columna A las probabilidades combinadas se han calculado considerando las PI por locus de cada población y en la B a partir de las PI por locus en la población total. En general para las combinaciones de 4 a 7 marcadores se observan menores valores de PI cuando se tiene en cuenta la estructura de la población, a partir de 8 marcadores no existen diferencias entre los dos valores calculados. Una vez más se observa que al disminuir el número de loci analizados es importante tener en cuenta el comportamiento del marcador en la población analizada, ya que el no hacerlo podría afectar la fiabilidad del ensayo (Orru *et al.* 2006).

Tabla 18: Valores de PI teóricos calculados usando combinaciones de 4 a 9 microsatélites que muestren los mejores valores de PI.

Nº locus	Parda de Montaña		Pirenaica I		Pirenaica II		Pirenaica I+II	
	A	B	A	B	A	B	A	B
4	3,97E-05	4,58E-05	5,27E-05	9,68E-05	1,27E-04	1,27E-04	4,03E-05	4,03E-05
5	4,40E-06	5,08E-06	7,12E-06	1,89E-05	1,88E-05	1,88E-05	5,53E-06	5,53E-06
6	5,02E-07	5,51E-07	1,03E-06	2,73E-06	3,27E-06	3,78E-06	8,11E-07	9,48E-07
7	6,22E-08	6,29E-08	2,01E-07	3,69E-07	6,58E-07	8,68E-07	1,39E-07	1,68E-07
8	7,80E-09	7,80E-09	4,83E-08	4,83E-08	1,51E-07	1,51E-07	2,47E-08	2,47E-08
9	1,49E-09	1,49E-09	1,34E-08	1,34E-08	7,04E-08	7,04E-08	9,29E-09	9,29E-09

En la Tabla 19 se resumen las probabilidades de identidad teóricas calculadas para el panel de 9 microsatélites. Para el caso de la población de Parda de Montaña así como de la Pirenaica I+II se obtienen valores del orden de 10^{-9} . Resultados similares fueron encontrados por otros autores (San Cristóbal-Gaudy *et al.* 2000; Cunningham & Meghan 2001) en poblaciones de individuos emparentados. También se encontraron valores similares en la raza bovina Galloway (Herraez *et al.* 2005) donde se obtuvo un valor del orden de 10^{-8} trabajando con un panel de 10 microsatélites. Si tenemos en cuenta la población como un todo se observa que el valor de PI disminuye hasta 10^{-10} ya que al incluir todos los individuos en una única

población la variabilidad genética será mayor y por tanto disminuirá la probabilidad de encontrar individuos iguales.

Tabla 19: Probabilidades de identidad teórica (PI) con el panel de 9 microsatélites.

Población	Probabilidad de identidad teórica
Parda de Montaña	$1,49 \times 10^{-9}$
Pirenaica I	$1,34 \times 10^{-8}$
Pirenaica II	$7,04 \times 10^{-8}$
Pirenaica I+II	$9,29 \times 10^{-9}$
Total	$7,23 \times 10^{-10}$

Teniendo en cuenta condiciones de comercialización que incluya un sistema de identificación que permita enlazar una pieza de carne con el animal de origen, la probabilidad de que exista un fraude y no sea detectado por el sistema será la probabilidad de que dos muestras al azar compartan el mismo perfil genético (Vazquez *et al.* 2004), por lo que en nuestras condiciones se obtiene una probabilidad de error del orden de 10^{-9} , dato muy importante a tener en cuenta ya que da una gran fiabilidad al estudio.

5.2.1.3.2. Probabilidad de identidad empírica (PIe)

Los valores de probabilidad de identidad teórica se obtienen asumiendo independencia de las frecuencias alélicas, por lo que esta condición puede infraestimar las probabilidades calculadas (Arana *et al.* 2002). Se ha realizado el cálculo de las probabilidades de identidad empíricas usando el algoritmo diseñado que se ha descrito en el apartado 5.1.3. Los resultados por locus se muestran en la Tabla 20 donde han sido dispuestas en orden ascendente.

Se puede apreciar que los valores obtenidos para las probabilidades de identidad empírica son ligeramente diferentes a los teóricos que se muestran en la Tabla 17. En todas las poblaciones estudiadas el valor más elevado es el del TGLA126 y los menores valores se presentan en general para el

TGLA227, BM2113, TGLA122 e INRA23. Los valores empíricos por locus siguen un comportamiento similar a los valores teóricos. Por tanto la distribución de los alelos es un hecho a tener en cuenta cuando se disminuye el número de loci genotipados para realizar estudios de identificación y verificación de la trazabilidad.

Tabla 20: Probabilidad de identidad empírica por locus.

Parda de Montaña		Pirenaica I		Pirenaica II		Pirenaica I+II		Total	
TGLA227	0,0508	TGLA227	0,0694	BM2113	0,0994	BM2113	0,0693	TGLA227	0,0436
TGLA122	0,0906	INRA23	0,0953	TGLA122	0,1078	TGLA227	0,0716	BM2113	0,0625
BM2113	0,0908	BM2113	0,1036	TGLA227	0,1132	INRA23	0,1106	TGLA122	0,0800
ETH10	0,1093	BM1824	0,1332	INRA23	0,1589	TGLA122	0,1114	INRA23	0,0801
BM1824	0,1279	SPS115	0,1517	ETH225	0,1813	BM1824	0,162	ETH10	0,128
ETH225	0,1289	ETH10	0,1628	BM1824	0,2036	ETH225	0,1659	BM1824	0,1373
SPS115	0,138	ETH225	0,2424	ETH10	0,2128	ETH10	0,196	ETH225	0,1388
INRA23	0,1411	TGLA122	0,2618	SPS115	0,2537	SPS115	0,206	SPS115	0,1509
TGLA126	0,2111	TGLA126	0,3404	TGLA126	0,5971	TGLA126	0,5011	TGLA126	0,3212

Al calcular las probabilidades de identidad empíricas combinadas incrementando el número de loci en función de su grado de informatividad (Tabla 21), se observó que para la población de Parda de Montaña la combinación de los microsátélites TGLA227, TGLA122, BM2113 y ETH10 era suficiente para alcanzar la probabilidad 0. Para las subpoblaciones de raza Pirenaica no fue posible alcanzar el valor de probabilidad 0, debido a la existencia de muestras repetidas, detectándose errores en el sistema de trazabilidad, lo cual será comentado en próximos capítulos.

Teniendo en cuenta las diferencias encontradas entre los valores teóricos y empíricos se ha calculado la relación entre estos (Tabla 22) y se ha observado que de forma general el valor empírico es mayor que el teórico, por tanto al realizar el cociente obtenemos valores menores de 1, resultados similares fueron encontrados por Arana y colaboradores (2002). Esto significa que el valor teórico sobrevalora la capacidad del marcador respecto al empírico que se obtiene a partir de los genotipos de los individuos directamente teniendo en cuenta la estructura de la población y por tanto el

número de progenitores empleados, tomando en consideración la heterocigocidad existente en cada una de las poblaciones.

Tabla 21: Valores de P_{Ie} combinados incrementando el número de locus.

Parda de Montaña

TGLA227	0,0508								
TGLA227	TGLA122	3,99E-03							
TGLA227	TGLA122	BM2113	1,90E-04						
TGLA227	TGLA122	BM2113	ETH10	0					
TGLA227	TGLA122	BM2113	ETH10	BM1824	0				
TGLA227	TGLA122	BM2113	ETH10	BM1824	ETH225	0			
TGLA227	TGLA122	BM2113	ETH10	BM1824	ETH225	SPS115	0		
TGLA227	TGLA122	BM2113	ETH10	BM1824	ETH225	SPS115	INRA23	0	
TGLA227	TGLA122	BM2113	ETH10	BM1824	ETH225	SPS115	INRA23	TGLA126	0

Pirenaica I

TGLA227	0,0694								
TGLA227	IINRA23	8,33E-03							
TGLA227	INRA23	BM2113	1,85E-03						
TGLA227	INRA23	BM2113	BM1824	9,25E-04					
TGLA227	INRA23	BM2113	BM1824	SPS115	9,25E-04				
TGLA227	INRA23	BM2113	BM1824	SPS115	ETH10	9,25E-04			
TGLA227	INRA23	BM2113	BM1824	SPS115	ETH10	ETH225	9,25E-04		
TGLA227	INRA23	BM2113	BM1824	SPS115	ETH10	ETH225	TGLA122	9,25E-04	
TGLA227	INRA23	BM2113	BM1824	SPS115	ETH10	ETH225	TGLA122	TGLA126	9,25E-04

Pirenaica II

BM2113	0,0994								
BM2113	TGLA122	1,01E-02							
BM2113	TGLA122	TGLA227	2,15E-03						
BM2113	TGLA122	TGLA227	INRA23	4,29E-04					
BM2113	TGLA122	TGLA227	INRA23	ETH225	4,29E-04				
BM2113	TGLA122	TGLA227	INRA23	ETH225	BM1824	4,29E-04			
BM2113	TGLA122	TGLA227	INRA23	ETH225	BM1824	ETH10	2,15E-04		
BM2113	TGLA122	TGLA227	INRA23	ETH225	BM1824	ETH10	SPS115	2,15E-04	
BM2113	TGLA122	TGLA227	INRA23	ETH225	BM1824	ETH10	SPS115	TGLA126	2,15E-04

Pirenaica I+II

BM2113	0,0693								
BM2113	TGLA227	7,85E-03							
BM2113	TGLA227	INRA23	1,34E-03						
BM2113	TGLA227	INRA23	TGLA122	3,83E-04					
BM2113	TGLA227	INRA23	TGLA122	BM1824	2,87E-04				
BM2113	TGLA227	INRA23	TGLA122	BM1824	ETH225	2,87E-04			
BM2113	TGLA227	INRA23	TGLA122	BM1824	ETH225	ETH10	1,92E-04		
BM2113	TGLA227	INRA23	TGLA122	BM1824	ETH225	ETH10	SPS115	1,92E-04	
BM2113	TGLA227	INRA23	TGLA122	BM1824	ETH225	ETH10	SPS115	TGLA126	1,92E-04

Total

TGLA227	0,0436								
TGLA227	BM2113	4,11E-03							
TGLA227	BM2113	TGLA122	5,22E-04						
TGLA227	BM2113	TGLA122	INRA23	1,63E-04					
TGLA227	BM2113	TGLA122	INRA23	ETH10	1,60E-04				
TGLA227	BM2113	TGLA122	INRA23	ETH10	BM1824	9,79E-05			
TGLA227	BM2113	TGLA122	INRA23	ETH10	BM1824	ETH225	9,79E-05		
TGLA227	BM2113	TGLA122	INRA23	ETH10	BM1824	ETH225	SPS115	9,79E-05	
TGLA227	BM2113	TGLA122	INRA23	ETH10	BM1824	ETH225	SPS115	TGLA126	9,79E-05

Tabla 22: Relación entre los valores teóricos y empíricos de las probabilidades de identidad por locus para las diferentes poblaciones analizadas.

Parda de Montaña				Pirenaica I				Pirenaica II				Pirenaica I+II				Total			
Loci	PIt	PIe	PIt/PIe	Loci	PIt	PIe	PIt/PIe	Loci	PIt	PIe	PIt/PIe	Loci	PIt	PIe	PIt/PIe	Loci	PIt	PIe	PIt/PIe
TGLA227	0,051	0,0508	1,0039	TGLA227	0,0618	0,0694	0,8905	TGLA122	0,087	0,1078	0,8071	TGLA227	0,0656	0,0716	0,9162	TGLA227	0,0409	0,0436	0,9381
TGLA122	0,0775	0,0906	0,8554	INRA23	0,0773	0,0953	0,8111	TGLA227	0,1025	0,1132	0,9055	BM2113	0,0667	0,0693	0,9625	BM2113	0,0593	0,0625	0,9488
BM2113	0,0925	0,0908	1,0187	BM2113	0,0844	0,1036	0,8147	BM2113	0,1037	0,0994	1,0433	INRA23	0,0956	0,1106	0,8644	TGLA122	0,0704	0,08	0,8800
ETH10	0,1085	0,1093	0,9927	BM1824	0,1308	0,1332	0,9820	INRA23	0,1371	0,1589	0,8628	TGLA122	0,0963	0,1114	0,8645	INRA23	0,0761	0,0801	0,9501
ETH225	0,1108	0,1289	0,8596	SPS115	0,1351	0,1517	0,8906	ETH225	0,1484	0,1813	0,8185	ETH225	0,1374	0,1659	0,8282	ETH225	0,1174	0,1388	0,8458
SPS115	0,1143	0,138	0,8283	ETH10	0,1448	0,1628	0,8894	BM1824	0,174	0,2036	0,8546	BM1824	0,1465	0,162	0,9043	ETH10	0,1198	0,128	0,9359
BM1824	0,1239	0,1279	0,9687	ETH225	0,1951	0,2424	0,8049	ETH10	0,2008	0,2128	0,9436	SPS115	0,1713	0,196	0,8740	SPS115	0,1265	0,1509	0,8383
INRA23	0,1253	0,1411	0,8880	TGLA122	0,2401	0,2618	0,9171	SPS115	0,2294	0,2537	0,9042	ETH10	0,1777	0,206	0,8626	BM1824	0,1311	0,1373	0,9548
TGLA126	0,1908	0,2111	0,9038	TGLA126	0,2768	0,3404	0,8132	TGLA126	0,4665	0,5971	0,7813	TGLA126	0,376	0,5011	0,7503	TGLA126	0,2385	0,3212	0,7425

PIt: Probabilidad de identidad teórica

PIe: Probabilidad de identidad empírica

PIt/PIe: relación entre el valor de PI teórico y el empírico

5.2.1.4. Estimación de parámetros genéticos en la descendencia a partir de parámetros genéticos de los progenitores utilizando microsatélites

En la valoración de la eficacia de los diferentes paneles de marcadores moleculares en estudios de identificación y verificación de trazabilidad, uno de los puntos a tener en cuenta es la posibilidad de estimar parámetros genéticos en los descendientes a partir de parámetros de los progenitores. Para llevar a cabo estos análisis se ha utilizado el paquete estadístico SPSS 14.0.

Primeramente se ha llevado a cabo el estudio de correlación entre las diferentes variables genéticas de los progenitores y los descendientes de las poblaciones analizadas. Se han calculado los coeficientes de correlación de Pearson (r), que cuantifican la fuerza de la relación lineal entre las diferentes variables. En la Tabla 23 se pueden encontrar los valores de estos coeficientes, mostrándose que existe correlación estadísticamente significativa entre las diferentes variables.

En la Figura 16 se puede observar que existe correlación positiva entre las probabilidades de identidad de los descendientes con respecto a las probabilidades de identidad de los progenitores (machos y hembras) con valores de coeficiente de correlación altos, de manera que el aumento de uno supone el aumento del otro aunque no en la misma proporción. Por otra parte la influencia de las heterocigosidades de los progenitores sobre las probabilidades de identidad tanto empíricas como teóricas de los descendientes muestra una correlación alta y negativa, por lo que cuanto mayor sean las heterocigosidades de los progenitores menores serán las PI de los descendientes, es decir menor probabilidad de encontrar individuos con los mismos genotipos entre los descendientes.

Las heterocigosidades en los descendientes también muestran correlación con los parámetros de los progenitores, siendo negativa con las probabilidades de identidad de los progenitores y positivas con las heterocigosidades.

Tabla 23: Coeficientes de correlación de Pearson entre parámetros genéticos de progenitores y descendientes (** $p < 0,01$, * $p < 0,05$)

		PIt(p)	PIt(m)	PIt(d)	PIe(p)	PIe(m)	PIe(d)	He(p)	He(m)	He(d)	Ho(p)	Ho(m)	Ho(d)
Microsatélites	PIt(p)	1	0,768(**)	0,755(**)	0,775(**)	0,788(**)	0,748(**)	-0,984(**)	-0,771(**)	-0,758(**)	-0,457(**)	-0,791(**)	-0,704(**)
	PIt(m)	0,768(**)	1	0,892(**)	0,824(**)	0,983(**)	0,899(**)	-0,805(**)	-0,994(**)	-0,908(**)	-0,447(**)	-0,914(**)	-0,871(**)
	PIt(d)	0,755(**)	0,892(**)	1	0,809(**)	0,910(**)	0,989(**)	-0,802(**)	-0,886(**)	-0,984(**)	-0,516(**)	-0,855(**)	-0,931(**)
	PIe(p)	0,775(**)	0,824(**)	0,809(**)	1	0,805(**)	0,812(**)	-0,816(**)	-0,806(**)	-0,810(**)	-0,223	-0,704(**)	-0,743(**)
	PIe(m)	0,788(**)	0,983(**)	0,910(**)	0,805(**)	1	0,923(**)	-0,813(**)	-0,989(**)	-0,917(**)	-0,500(**)	-0,938(**)	-0,889(**)
	PIe(d)	0,748(**)	0,899(**)	0,989(**)	0,812(**)	0,923(**)	1	-0,799(**)	-0,896(**)	-0,988(**)	-0,518(**)	-0,870(**)	-0,929(**)
	He(p)	-0,984(**)	-0,805(**)	-0,802(**)	-0,816(**)	-0,813(**)	-0,799(**)	1	0,804(**)	0,809(**)	0,454(**)	0,810(**)	0,747(**)
	He(m)	-0,771(**)	-0,994(**)	-0,886(**)	-0,806(**)	-0,989(**)	-0,896(**)	0,804(**)	1	0,903(**)	0,463(**)	0,932(**)	0,866(**)
	He(d)	-0,758(**)	-0,908(**)	-0,984(**)	-0,810(**)	-0,917(**)	-0,988(**)	0,809(**)	0,903(**)	1	0,500(**)	0,862(**)	0,934(**)
	Ho(p)	-0,457(**)	-0,447(**)	-0,516(**)	-0,223	-0,500(**)	-0,518(**)	0,454(**)	0,463(**)	0,500(**)	1	0,583(**)	0,622(**)
	Ho(m)	-0,791(**)	-0,914(**)	-0,855(**)	-0,704(**)	-0,938(**)	-0,870(**)	0,810(**)	0,932(**)	0,862(**)	0,583(**)	1	0,866(**)
Ho(d)	-0,704(**)	-0,871(**)	-0,931(**)	-0,743(**)	-0,889(**)	-0,929(**)	0,747(**)	0,866(**)	0,934(**)	0,622(**)	0,866(**)	1	

PIt: Probabilidades de identidad teóricas
 PIE: Probabilidades de identidad empíricas
 He: Heterocigosidades esperadas
 Ho: Heterocigosidades observadas
 (p): padres
 (m): madres
 (d): descendientes

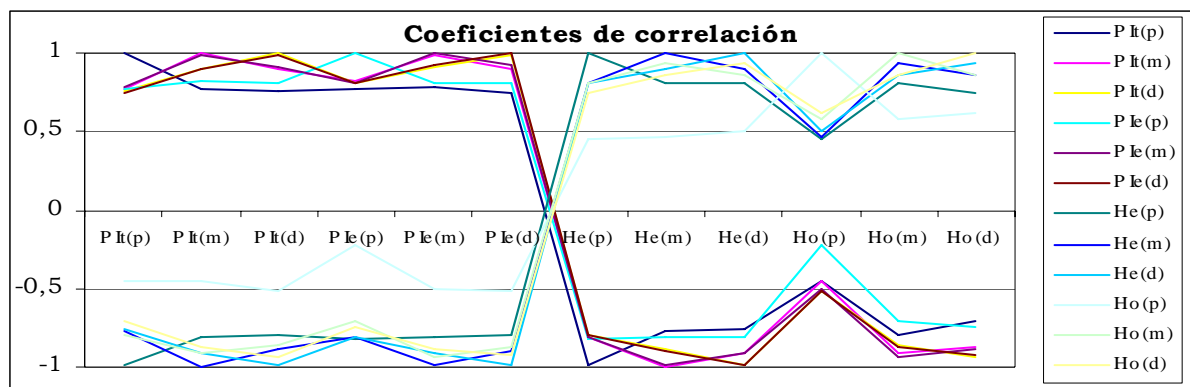


Figura 16: Representación de los coeficientes de correlación de Pearson entre las diferentes variables.

Debido a la existencia de una correlación clara entre las probabilidades de identidad empíricas y las heterocigosidades observadas de los descendientes con las mismas variables en los progenitores se han llevado a cabo análisis de regresión para intentar predecir el valor de las variables en los descendientes a partir de los valores de las variables teóricas de los progenitores.

5.2.1.4.1. Análisis de regresión para las probabilidades de identidad empíricas de los descendientes

Como se ha visto anteriormente existe una clara correlación entre la probabilidad de identidad empírica de los descendientes y variables de los progenitores, por lo que en este apartado se ha llevado a cabo el diseño de modelos de regresión lineal que permite predecir el comportamiento de las probabilidades de identidad empírica de los descendientes a partir de variables de los progenitores.

Se trata de explicar el comportamiento de la variable de respuesta lo mejor posible, en este caso la probabilidad de identidad empírica de los descendientes, mediante el uso del menor número de variables predictoras posibles. Para esto se ha utilizado el método "stepwise". La Tabla 24 muestra que la variable más importante en la obtención de modelos de regresión ha

sido en todos los grupos analizados la Probabilidad de Identidad teórica de las madres (PIt(m)).

Tabla 24: Variables incluidas en los modelos de regresión mediante el método stepwise.

Marcador	Grupo	Modelo	Variables incluidas
Microsatélites	Total	1	PIt(m)
	Parda de Montaña	1	PIt(m)
	Pirenaica I	1	PIt(m)
	Pirenaica I+II	1	PIt(m)

Para verificar el ajuste de los modelos obtenidos a los datos observados en los diferentes grupos poblacionales, se han calculado los coeficientes de determinación (R^2) que nos medirán el grado de dependencia de las diferentes variables. En la Tabla 25 aparecen los valores de R^2 para cada uno de los modelos en los diferentes grupos, el cual nos permite interpretar qué porcentaje de la variación de la probabilidad de identidad empírica de los descendientes es debido a las restantes variables correlacionadas en los modelos obtenidos por grupos para los marcadores microsatélites.

Tabla 25: Coeficiente de determinación ajustado para cada modelo.

Marcador	Grupo	Modelo	R^2 ajustado
Microsatélites	Total	1	0,956934522
	Parda de Montaña	1	0,978326288
	Pirenaica I	1	0,985375617
	Pirenaica I+ II	1	0,952805329

De la Tabla 25 se puede interpretar que los modelos obtenidos permiten explicar en todos los grupos poblacionales más del 95% de la variación de la probabilidad de identidad empírica de los descendientes en función de la probabilidad de identidad teórica de las madres. Siendo los modelos que más se ajustan los obtenidos para las poblaciones de Parda de Montaña (97,8%) y de Pirenaica I (98,54%). Los modelos que menos se ajustan son los obtenidos para Pirenaica I+II (95,28%) que incluye dos subpoblaciones de la raza, o cuando consideramos todos los individuos

como una única población (95,69%). A partir de estos resultados podemos concluir que en estas poblaciones al realizar la estimación de la PI empírica de los descendientes ($PI_{est}(d)$) una de las variables más importantes es la PI teórica de las madres.

En la Tabla 26 se muestran los coeficientes de las rectas de regresión para estimar la variación de las PI empíricas de los descendientes en función de la PI teórica de las madres.

Tabla 26: Coeficientes de las rectas de regresión obtenidas.

Marcador	Grupo	Modelo	Variables	Coeficientes estandarizados	
				B	Std. Error
Microsatélites	Total	1	PIt(m)	1,302439212	0,091870614
	Parda de Montaña	1	PIt(m)	1,14753947	0,056863987
	Pirenaica I	1	PIt(m)	1,16868838	0,047419564
	Pirenaica I+II	1	PIt(m)	1,507302883	0,111514632

De la tabla anterior podemos inferir los modelos matemáticos que permiten predecir el comportamiento de la PI empírica de los descendientes en función de las PI teóricas de las madres en los diferentes grupos, los cuales se muestran a continuación por poblaciones.

Población total (Parda de Montaña y Pirenaica I+II):

$$PI_{est}(d) = 1,302 * PI_{t(m)}$$

Parda de Montaña:

$$PI_{est}(d) = 1,147 * PI_{t(m)}$$

Pirenaica I:

$$PI_{est}(d) = 1,168 * PI_{t(m)}$$

Pirenaica I + II:

$$PI_{est}(d) = 1,507 * PI_{t(m)}$$

Mediante dichos modelos se puede determinar la probabilidad de identidad de una población teniendo genotipados solamente los progenitores de dicha población.

5.2.1.4.1.1. Comparción entre las PI teóricas y las PI estimadas por regresión en los descendientes

Con cada uno de los modelos anteriores obtenidos por regresión se han calculado las Probabilidades de identidad estimadas (PIest) en los descendientes y se han comparado mediante gráficos con los valores teóricos y empíricos. Se han realizado las pruebas *t* para muestras pareadas para ver si existen diferencias significativas entre los valores teóricos y los estimados utilizando el modelo de regresión.

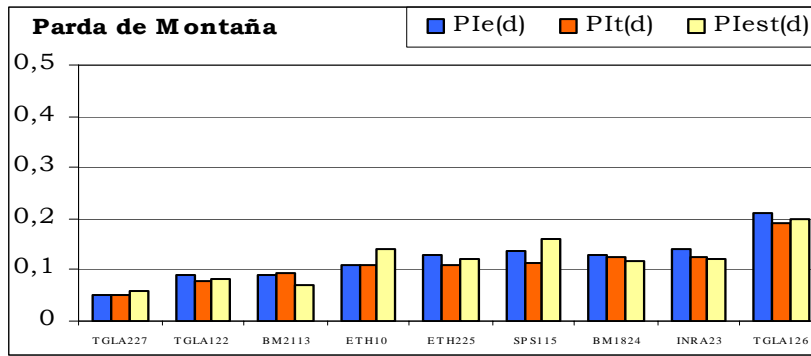
En la Tabla 27 y la Figura 17 se puede observar respectivamente los valores de PIE, PI_t y PIest, así como la representación gráfica de las mismas. Los resultados de las pruebas “*t*” para muestras pareadas han mostrado que existen diferencias muy significativas entre los valores teóricos y estimados por regresión para los grupos Pirenaica I+II y total.

Para la población de Parda de Montaña y Pirenaica I no existen diferencias significativas entre los valores teóricos y los estimados, ya que en el cálculo de este valor se tiene en cuenta la estructura de cada una de las poblaciones, es decir su contenido de información polimórfica. Por lo que si se analizaran estas poblaciones sería posible realizar la estimación de las PIE de los descendientes a partir de los progenitores.

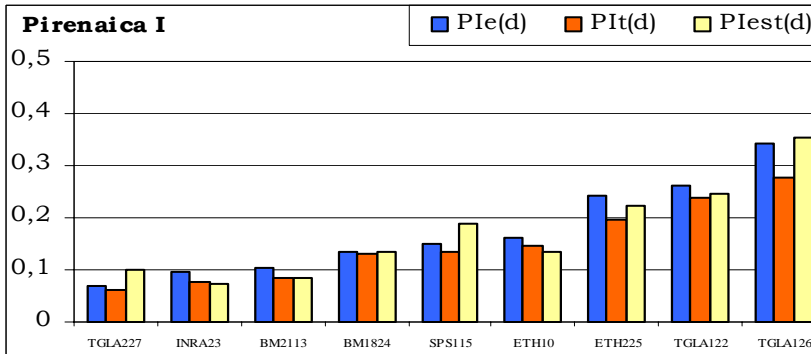
Las diferencias que existen entre los valores teóricos y estimados para el grupo Pirenaica I+II y el total, pueden ser debidas a que en ambos grupos se han mezclado individuos de diferentes procedencias. En el caso de Pirenaica I+II se han tomado individuos de dos subpoblaciones y en el grupo total se han mezclado individuos de dos poblaciones diferentes genéticamente (Parda de Montaña y Pirenaica) con diferentes contenidos de información polimórfica por locus. Si no se tiene en cuenta la estructura genética de la población, la estimación de los valores de PI puede no ajustarse a los valores reales al ser enmascarado el contenido real de polimorfismo que existe en cada una de las poblaciones incluidas en el grupo.

Tabla 27: Comparación de las Probabilidades de identidad empírica [PIe(d)], teórica [PIt(d)] y estimada por regresión [PIest(d)] de los descendientes por subpoblaciones.

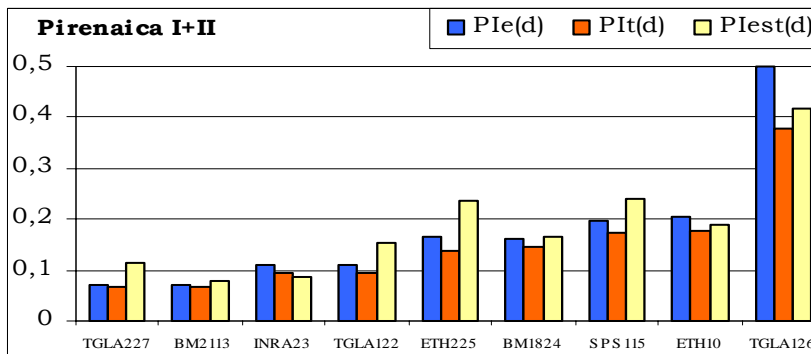
Parda de Montaña				Pirenaica I				Pirenaica I+II				Total			
Loci	PIe(d)	PIt(d)	PIest(d)	Loci	PIe(d)	PIt(d)	PIest(d)	Loci	PIe(d)	PIt(d)	PIest(d)	Loci	PIe(d)	PIt(d)	PIest(d)
TGLA227	0,0508	0,051	0,0604	TGLA227	0,0694	0,0618	0,0986	TGLA227	0,0716	0,0656	0,1154	TGLA227	0,0436	0,0409	0,0691
TGLA122	0,0906	0,0775	0,0820	INRA23	0,0953	0,0773	0,0749	BM2113	0,0693	0,0667	0,0791	BM2113	0,0625	0,0593	0,0633
BM2113	0,0908	0,0925	0,0695	BM2113	0,1036	0,0844	0,0864	INRA23	0,1106	0,0956	0,0857	TGLA122	0,08	0,0704	0,0983
ETH10	0,1093	0,1085	0,1419	BM1824	0,1332	0,1308	0,1362	TGLA122	0,1114	0,0963	0,1546	INRA23	0,0801	0,0761	0,0837
ETH225	0,1289	0,1108	0,1215	SPS115	0,1517	0,1351	0,1871	ETH225	0,1659	0,1374	0,2374	ETH225	0,1388	0,1174	0,1603
SPS115	0,138	0,1143	0,1600	ETH10	0,1628	0,1448	0,1346	BM1824	0,162	0,1465	0,1658	ETH10	0,128	0,1198	0,1564
BM1824	0,1279	0,1239	0,1180	ETH225	0,2424	0,1951	0,2244	SPS115	0,196	0,1713	0,2390	SPS115	0,1509	0,1265	0,1807
INRA23	0,1411	0,1253	0,1207	TGLA122	0,2618	0,2401	0,2463	ETH10	0,206	0,1777	0,1893	BM1824	0,1373	0,1311	0,1354
TGLA126	0,2111	0,1908	0,2005	TGLA126	0,3404	0,2768	0,3533	TGLA126	0,5011	0,376	0,4182	TGLA126	0,3212	0,2385	0,2544



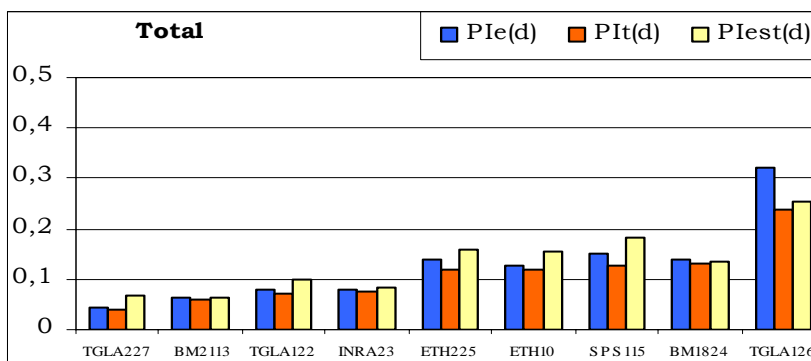
prueba *t*
 PIIt(d): PIest(d)
0,2322



prueba *t*
 PIIt(d): PIest(d)
0,0544



prueba *t*
 PIIt(d): PIest(d)
0,0088**



prueba *t*
 PIIt(d): PIest(d)
0,0034**

Figura 17: Representación gráfica de las PI empíricas, teóricas y estimadas de los descendientes. Valores de las pruebas “*t*” de student para muestras pareadas para las PIIt(d) y PIest(d).

5.2.1.4.2. Análisis de regresión para las heterocigosidades de los descendientes

De la misma manera que se ha procedido para el análisis de regresión en el caso de la probabilidad de identidad empírica de los descendientes se ha hecho para las heterocigosidades. Se trata en este caso de poder predecir el valor de la heterocigosidad observada de los descendientes ($H_{est}(d)$) a través de los diferentes parámetros de los progenitores, mediante un análisis de regresión. En la Tabla 28 se muestran las variables que se han tenido en cuenta en cada uno de los modelos obtenidos por regresión utilizando el método “Stepwise”.

Tabla 28: Variables incluidas en los modelos de regresión mediante el método stepwise.

Marcador	Grupo	Modelo	Variabes incluidas
Microsatélites	Total	1	$He(m)$
	Parda de Montaña	1	$He(m)$
	Pirenaica I	1	$He(p)$
	Pirenaica I+II	1	$He(m)$
		2	$PIt(m)$

En general se observa que la heterocigosidad esperada de las progenitoras de los diferentes grupos es una variable muy importante en la estimación de las heterocigosidades en los descendientes. En la subpoblación I de Pirenaica, la variable que permite estimar mejor la heterocigosidad en los descendientes es la heterocigosidad esperada de los padres. En el grupo Pirenaica I+II obtenemos dos modelos, el primero involucra solamente la heterocigosidad esperada de las madres y un segundo modelo implica además la probabilidad de identidad teórica de las madres.

La medida del ajuste de los modelos obtenidos por regresión se midió mediante el cálculo del coeficiente de determinación, cuyos valores aparecen en la Tabla 29. Se puede apreciar que la capacidad de estimar las heterocigosidades de los descendientes con los modelos obtenidos en los diferentes grupos poblacionales es muy alta, siempre superiores al 98%.

En todos los grupos los mejores modelos dan un ajuste superior al 99%. En la población Pirenaica I+II se obtienen dos modelos de regresión, uno con una capacidad de predicción del 98,77% y el segundo de 99,31%. El primero solo depende de las heterocigosidades esperadas de las madres y el segundo depende además de las probabilidades de identidad.

Tabla 29: Coeficientes de determinación ajustados para cada modelo.

Marcador	Grupo	Modelo	R ² ajustado
Microsatélites	Total	1	0,9910579
	Parda de Montaña	1	0,99231665
	Pirenaica I	1	0,99802047
	Pirenaica I+II	1	0,98774947
		2	0,99308424

La Tabla 30 muestra los coeficientes de las rectas de regresión que permite establecer un modelo matemático para estimar las heterocigosidades en los descendientes a partir de las variables de los progenitores.

Tabla 30: Coeficientes de las rectas de regresión pasando por el origen de coordenadas.

Marcador	Grupo	Modelo	Variables	Coeficientes estandarizados	
				B	Std. Error
Microsatélites	Total	1	He(m)	0,950130205	0,030068698
	Parda de Montaña	1	He(m)	0,99708564	0,029233075
	Pirenaica I	1	He(p)	1,115371683	0,016556247
	Pirenaica I+II	1	He(m)	0,954907084	0,035423783
			He(m)	1,056743761	0,046417311
			PIt(m)	-0,642856013	0,240059859

Los modelos matemáticos que permiten estimar el comportamiento de las heterocigosidades en los descendientes en función de variables de los progenitores en distintas poblaciones se muestran a continuación. Se puede observar que en todos los casos (excepto el modelo 2 de Pirenaica I+II), los modelos siguen una función del tipo $y = mx + n$; al no existir término independiente la función se representaría como una recta que pasa por el origen de coordenadas. El modelo 2 de Pirenaica I+II no se puede expresar como una recta ya que es una función con dos variables.

Población total (Parda de Montaña y Pirenaica I+II):

$$Hest(d) = 0,95 * He(m)$$

Parda de Montaña:

$$Hest(d) = 0,997 * He(m)$$

Pirenaica I:

$$Hest(d) = 1,115 * He(p)$$

Pirenaica I + II:

$$1) Hest(d) = 0,955 * He(m)$$

$$2) Hest(d) = 1,057 * He(m) - 0,643 Pit(m)$$

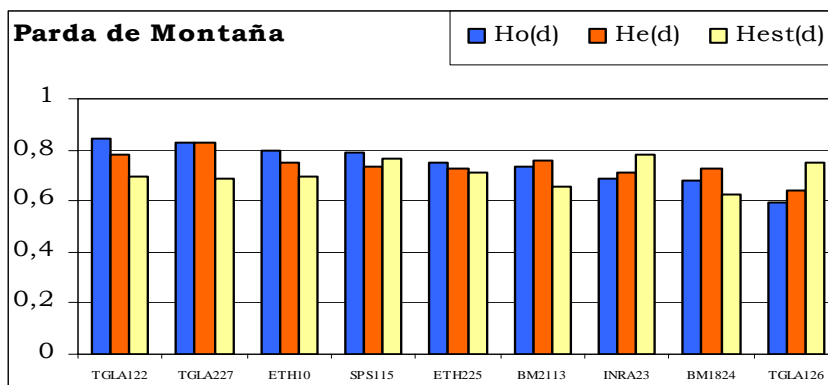
5.2.1.4.2.1. Comparación entre las heterocigosidades esperadas y estimadas de los descendientes

Utilizando los modelos matemáticos anteriores se han estimado las heterocigosidades de los descendientes y se han comparado con los valores esperados y observados. Mediante la prueba *t* para muestras pareadas se han comparado los valores estimados con los esperados (Tabla 31 y Figura 18).

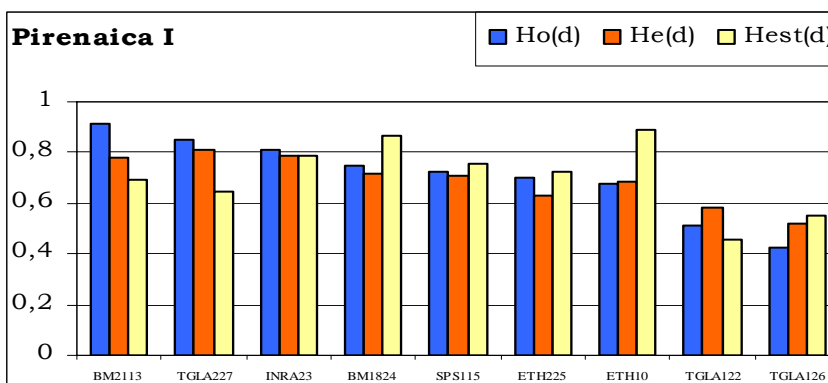
Los valores de las heterocigosidades observadas, esperadas y estimadas por regresión, así como las pruebas *t* (Tabla 31 y Figura 18), nos permiten concluir que con los modelos obtenidos por regresión se pueden estimar las heterocigosidades en los descendientes a partir de los progenitores con una elevada fiabilidad en las poblaciones analizadas. En la población total se puede apreciar que los valores estimados por regresión difieren de los teóricos ($p < 0,001$) por lo que cuando se trabaja con poblaciones genéticamente diferentes mezcladas, no se pueden aplicar este tipo de aproximaciones.

Tabla 31: Comparación de las heterocigosidades observadas [Ho(d)], esperadas [He(d)] y estimadas por regresión [Hest(d)] en los descendientes.

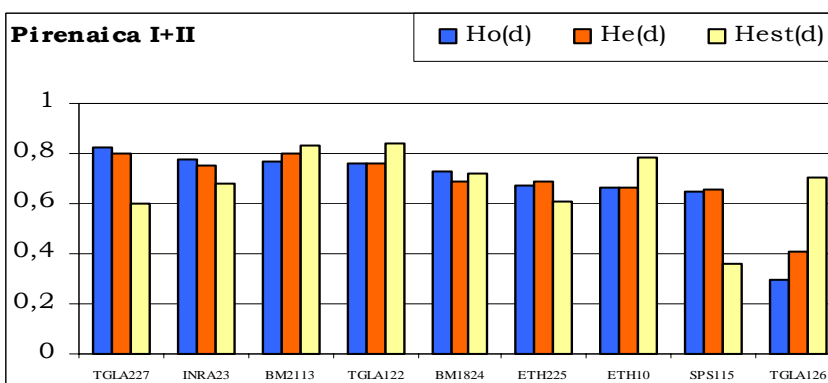
Parda de Montaña				Pirenaica I				Pirenaica I+II				Total			
Locus	Ho(d)	He(d)	Hest(d)	Locus	Ho(d)	He(d)	Hest(d)	Locus	Ho(d)	He(d)	Hest(d)	Locus	Ho(d)	He(d)	Hest(d)
TGLA122	0,8447	0,7835	0,69236	BM2113	0,9149	0,7765	0,6913	TGLA227	0,8207	0,7988	0,5981444	TGLA227	0,8226	0,8491	0,806645
TGLA227	0,8252	0,831	0,69065	TGLA227	0,8511	0,8107	0,6467	INRA23	0,7793	0,7513	0,6835559	TGLA122	0,7944	0,7956	0,75582
ETH10	0,7961	0,7494	0,6992	INRA23	0,8085	0,7864	0,7897545	BM2113	0,7655	0,797	0,8295191	BM2113	0,754	0,8043	0,764085
SPS115	0,7864	0,7324	0,76589	BM1824	0,7447	0,7171	0,8662435	TGLA122	0,7586	0,7577	0,8364706	INRA23	0,7419	0,7863	0,746985
ETH225	0,7476	0,7244	0,714305	SPS115	0,7234	0,7051	0,758869	BM1824	0,7241	0,6888	0,7200117	ETH10	0,7177	0,7314	0,69483
BM2113	0,7379	0,7572	0,653315	ETH225	0,7021	0,6272	0,727872	ETH225	0,6759	0,6885	0,6111781	SPS115	0,7056	0,7094	0,67393
INRA23	0,6893	0,7104	0,783655	ETH10	0,6809	0,6884	0,892	ETH10	0,6621	0,6608	0,7841907	ETH225	0,7056	0,7117	0,676115
BM1824	0,6796	0,7279	0,627475	TGLA122	0,5106	0,5816	0,4530245	SPS115	0,6483	0,6573	0,3600033	BM1824	0,7056	0,7131	0,677445
TGLA126	0,5922	0,6442	0,75183	TGLA126	0,4255	0,5182	0,550587	TGLA126	0,2966	0,4067	0,7066952	TGLA126	0,4194	0,5587	0,530765



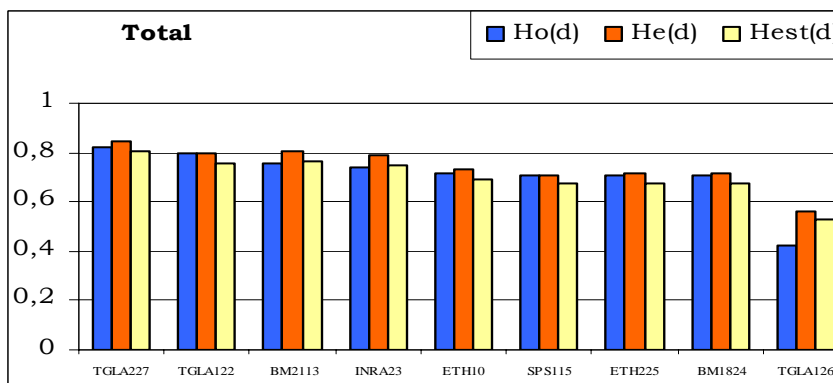
prueba *t*
He(d): Hest(d)
0,3126



prueba *t*
He(d): Hest(d)
0,6718



prueba *t*
He(d): Hest(d)
0,8884



prueba *t*
He(d): Hest(d)
4,59x10⁻⁹ ***

Figura 18: Representación gráfica de las Heterocigosidades observadas, esperadas y estimadas de los descendientes. Valores de las pruebas “*t*” de Student para muestras pareadas para las He(d) y Hest(d).

5.2.2. SNPs

5.2.2.1 Estimación de la variabilidad genética

Tal como se ha descrito en el capítulo “Material y Métodos” se han analizado un total de 20 SNPs para lo que se han puesto a punto dos reacciones de amplificación, cada una de las cuales incluía 10 SNPs. Posteriormente se llevaron a cabo las 2 reacciones múltiplex de minisequenciación. En la Figura 19 se pueden apreciar dos electroferogramas de las reacciones de minisequenciación, el primero muestra la reacción del Múltiplex 1 y el segundo el Múltiplex 2.

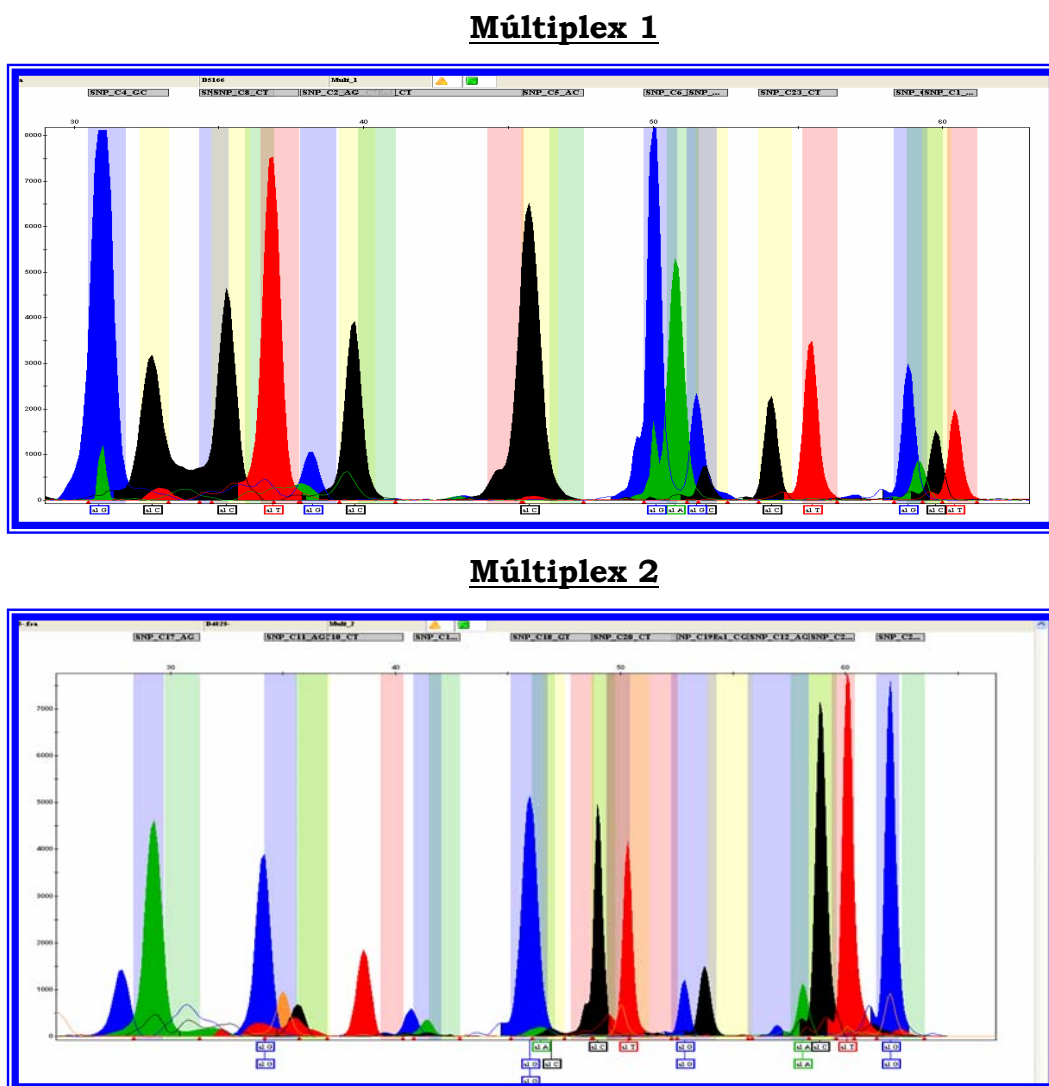


Figura 19: Electroferogramas que muestran las reacciones de minisequenciación para los múltiplex 1 y 2.

La selección del panel de SNPs empleado en este trabajo se realizó a partir de polimorfismos descritos teniendo en cuenta su ubicación cromosómica, incluyéndose además un nuevo SNP g.111A>G (AF465160) en el cromosoma 11 detectado por secuenciación por primera vez en este trabajo. Para la elección de los 20 SNPs se tuvo en cuenta las frecuencias alélicas descritas previamente por otros autores, así como las encontradas en razas españolas, intentando que la frecuencia alélica mínima en el grupo total fuera al menos del 20% (Orrù *et al.* 2009). Aunque el SNP ubicado en el cromosoma 11 detectado por secuenciación no cumpliera esta condición se incluyó en el trabajo por ser nuevo y para estudiar su comportamiento en diferentes razas españolas.

La informatividad aportada por los SNPs depende de sus frecuencias, por lo que resulta imprescindible el estudio de los mismos en diferentes razas, ya que esta puede variar significativamente entre poblaciones (Krawczak 1999). En el presente trabajo se han analizado los 20 SNPs distribuidos por todo en las siguientes razas explotadas en España: Asturiana, Frisona, Limousine, Menorquina, Monchina, Serrana de Teruel, Pirenaica y Parda de Montaña. Los resultados de las frecuencias alélicas observadas se muestran en la Tabla 32.

A partir de los valores de frecuencias alélicas totales se realizó el gráfico de la distribución de estas (Figura 20), mostrándose que para la mayoría de los SNPs se cumplen los requerimientos mínimos para su empleo en identificación.

El SNP g.111A>G (AF465160) en el cromosoma 11 presenta una distribución de frecuencias muy desigual en las diferentes razas. En ellas está fijado alelo G, no siendo así en la raza Pirenaica que es una de las empleadas en el estudio de trazabilidad.

Tabla 32: Frecuencias alélicas de los 20 SNPs en razas españolas.

Locus	Alelos	Asturianas	Frisonas	Limousine	Menorquina	Monchina	SerranaT	Pardas	Pirenaicas	Global
SNP_C1	C	0,344	0,382	0,367	0,375	0,393	0,711	0,542	0,354	0,443
	T	0,656	0,618	0,633	0,625	0,607	0,289	0,458	0,646	0,557
SNP_C4	C	0,217	0,214	0,341	0,565	0,304	0,261	0,25	0,438	0,325
	G	0,783	0,786	0,659	0,435	0,696	0,739	0,75	0,562	0,675
SNP_C5	A	0,043	0,214	0,105	0,152	0,091	0,065	0,104	0,042	0,101
	C	0,957	0,786	0,895	0,848	0,909	0,935	0,896	0,958	0,899
SNP_C6	A	0,478	0,587	0,587	0,63	0,587	0,63	0,812	0,812	0,642
	G	0,522	0,413	0,413	0,37	0,413	0,37	0,188	0,188	0,358
SNP_C7Ex3	C	0,152	0,262	0,125	0,674	0,227	0,391	0,229	0,146	0,278
	T	0,848	0,738	0,875	0,326	0,773	0,609	0,771	0,854	0,722
SNP_C8	C	0,304	0,119	0,5	0,239	0,19	0,348	0,375	0,479	0,32
	T	0,696	0,881	0,5	0,761	0,81	0,652	0,625	0,521	0,68
SNP_C21	C	0,304	0,375	0,174	0,152	0,239	0,409	0,354	0,354	0,294
	G	0,696	0,625	0,826	0,848	0,761	0,591	0,646	0,646	0,706
SNP_C23	C	0,609	0,619	0,789	0,761	0,591	0,609	0,708	0,479	0,642
	T	0,391	0,381	0,211	0,239	0,409	0,391	0,292	0,521	0,358
SNP_C24	A	0,543	0,174	0,237	0,565	0,545	0,478	0,312	0,521	0,425
	G	0,457	0,826	0,763	0,435	0,455	0,522	0,688	0,479	0,575
SNP_C2	A	0,261	0,548	0,447	0,5	0,357	0,326	0,292	0,229	0,365
	G	0,739	0,452	0,553	0,5	0,643	0,674	0,708	0,771	0,635
SNP_C10	C	0,848	0,891	0,711	0,727	0,848	0,891	0,792	0,833	0,818
	T	0,152	0,109	0,289	0,273	0,152	0,109	0,208	0,167	0,182
SNP_C11	A	0	0	0	0	0	0,022	0	0,229	0,032
	G	1	1	1	1	1	0,978	1	0,771	0,968
SNP_C13	A	0,543	0,5	0,543	0,523	0,587	0,565	0,429	0,625	0,541
	G	0,457	0,5	0,457	0,477	0,413	0,435	0,571	0,375	0,459
SNP_C17	A	0,348	0	0,571	0,739	0,522	0,522	0,604	0,604	0,489
	G	0,652	1	0,429	0,261	0,478	0,478	0,396	0,396	0,511
SNP_C18	G	0,457	0,674	0,523	0,295	0,587	0,435	0,396	0,479	0,481
	T	0,543	0,326	0,477	0,705	0,413	0,565	0,604	0,521	0,519
SNP_C19Ex	C	0,239	0,413	0,587	0,409	0,217	0,283	0,104	0,104	0,292
	G	0,761	0,587	0,413	0,591	0,783	0,717	0,896	0,896	0,708
SNP_C20	C	0,5	0,696	0,174	0,045	0,304	0,152	0,083	0,146	0,262
	T	0,5	0,304	0,826	0,955	0,696	0,848	0,917	0,854	0,738
SNP_C12	A	0,739	0,804	0,304	0,341	0,739	0,565	0,771	0,646	0,616
	G	0,261	0,196	0,696	0,659	0,261	0,435	0,229	0,354	0,384
SNP_C26	C	0,5	0,5	0,5	0,5	0,522	0,5	0,5	0,5	0,503
	T	0,5	0,5	0,5	0,5	0,478	0,5	0,5	0,5	0,497
SNP_C29	A	0,391	0,152	0,452	0	0,261	0,261	0,479	0,521	0,32
	G	0,609	0,848	0,548	1	0,739	0,739	0,521	0,479	0,68

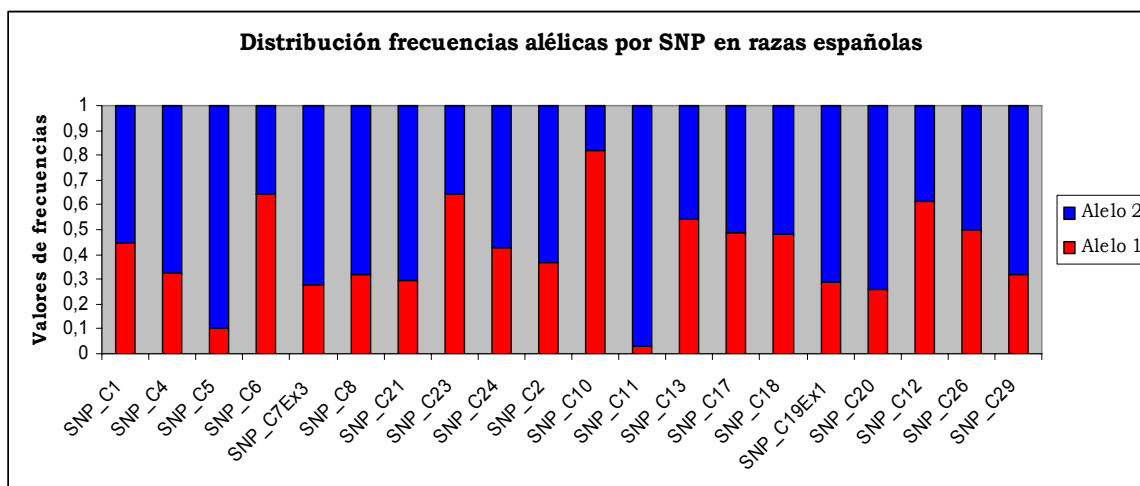


Figura 20: Distribución de las frecuencias alélicas por SNPs en las razas españolas estudiadas.

Una vez elegido el panel de polimorfismos para llevar a cabo el estudio de verificación de trazabilidad, se realizó el genotipado de todas las muestras que forman parte de las poblaciones incluidas en el estudio de trazabilidad (Parda de Montaña, Pirenaica I, Pirenaica II, Pirenaica I+II y población total). A partir de los resultados de los genotipos se calcularon las frecuencias alélicas por SNP en las poblaciones estudiadas (Tabla 33).

Se detectó equilibrio de Hardy-Weinberg para todos los loci excepto para el SNP ubicado en el cromosoma 26 que mostró un desequilibrio altamente significativo. Las frecuencias alélicas para dicho SNP eran las más cercanas a las recomendadas, sin embargo había un exceso de heterocigotos, por lo que este polimorfismo fue eliminado del estudio. Los resultados del desequilibrio serán comentados en el apartado 5.2.2.1.1.

Tabla 33: Frecuencias alélicas de SNPs por poblaciones.

	Alelos	Parda de Montaña	Pirenaica I	Pirenaica II	Pirenaica I+II	Total
SNP_C1	C	0,4238	0,4648	0,3088	0,3623	0,3883
	T	0,5762	0,5352	0,6912	0,6377	0,6117
SNP_C4	C	0,4605	0,3125	0,4362	0,3991	0,4247
	G	0,5395	0,6875	0,5638	0,6009	0,5753
SNP_C5	A	0,1694	0,0408	0,0204	0,0272	0,0923
	C	0,8307	0,9592	0,9796	0,9728	0,9078
SNP_C6	A	0,6424	0,7077	0,7231	0,7180	0,6850
	G	0,3576	0,2923	0,2769	0,2821	0,3150
SNP_C7	C	0,4937	0,5070	0,1441	0,2857	0,3824
	T	0,5063	0,4930	0,8559	0,7143	0,6177
SNP_C8	C	0,3624	0,5139	0,4745	0,4880	0,4358
	T	0,6376	0,4861	0,5256	0,5120	0,5643
SNP_C21	C	0,2867	0,4028	0,3504	0,3684	0,3352
	G	0,7133	0,5972	0,6496	0,6316	0,6648
SNP_C23	C	0,5520	0,6349	0,5845	0,6000	0,5794
	T	0,4481	0,3651	0,4155	0,4000	0,4206
SNP_C24	A	0,5278	0,6267	0,4962	0,5433	0,5369
	G	0,4722	0,3733	0,5038	0,4567	0,4631
SNP_C2	A	0,3630	0,2879	0,2149	0,2406	0,2943
	G	0,6370	0,7121	0,7851	0,7594	0,7057
SNP_C10	C	0,7445	0,8704	0,9314	0,9103	0,8328
	T	0,2555	0,1296	0,0686	0,0897	0,1672
SNP_C11	A	0,0721	0,4559	0,1810	0,2826	0,2034
	G	0,9279	0,5441	0,8190	0,7174	0,7966
SNP_C13	A	0,4408	0,6716	0,5480	0,5869	0,5260
	G	0,5592	0,3284	0,4521	0,4132	0,4740
SNP_C17	A	0,4468	0,3623	0,4275	0,4058	0,4224
	G	0,5532	0,6377	0,5725	0,5942	0,5776
SNP_C18	G	0,4813	0,4412	0,2955	0,3450	0,4056
	T	0,5188	0,5588	0,7046	0,6550	0,5944
SNP_C19	C	0,1345	0,4189	0,2439	0,3096	0,2437
	G	0,8655	0,5811	0,7561	0,6904	0,7563
SNP_C20	C	0,3517	0,2581	0,0680	0,1394	0,2387
	T	0,6483	0,7419	0,9320	0,8606	0,7613
SNP_C12	A	0,6689	0,6286	0,7724	0,7202	0,6977
	G	0,3311	0,3714	0,2276	0,2798	0,3023
SNP_C26	C	0,5024	0,5000	0,5000	0,5000	0,5010
	T	0,4976	0,5000	0,5000	0,5000	0,4990
SNP_C29	A	0,3237	0,5571	0,5177	0,5308	0,4486
	G	0,6763	0,4429	0,4823	0,4692	0,5514

Los valores de heterocigosidades esperadas y observadas así como el contenido de información polimórfica se muestran en la Tabla 34.

Tabla 34: Valores de heterocigosidades y contenido de información polimórfica.

Locus	Parda de Montaña			Pirenaica I			Pirenaica II			Pirenaica I+II			Total		
	PIC	Ho	He	PIC	Ho	He	PIC	Ho	He	PIC	Ho	He	PIC	Ho	He
SNP_C1	0,369	0,466	0,488	0,374	0,511	0,498	0,336	0,412	0,425	0,355	0,448	0,462	0,362	0,456	0,475
SNP_C4	0,373	0,476	0,497	0,337	0,362	0,430	0,371	0,557	0,493	0,365	0,490	0,480	0,369	0,484	0,489
SNP_C5	0,242	0,204	0,281	0,075	0,043	0,078	0,039	0,021	0,040	0,052	0,028	0,053	0,153	0,101	0,168
SNP_C6	0,354	0,466	0,460	0,328	0,383	0,414	0,320	0,351	0,397	0,323	0,366	0,405	0,338	0,407	0,432
SNP_C7	0,375	0,534	0,500	0,375	0,511	0,500	0,216	0,155	0,249	0,325	0,269	0,408	0,361	0,379	0,472
SNP_C8	0,355	0,447	0,462	0,375	0,532	0,500	0,374	0,433	0,499	0,375	0,469	0,500	0,371	0,460	0,492
SNP_C21	0,325	0,388	0,409	0,365	0,532	0,481	0,352	0,448	0,454	0,357	0,479	0,465	0,346	0,441	0,446
SNP_C23	0,372	0,495	0,495	0,356	0,340	0,464	0,368	0,474	0,487	0,365	0,428	0,480	0,369	0,456	0,487
SNP_C24	0,374	0,398	0,499	0,358	0,596	0,468	0,375	0,478	0,500	0,373	0,522	0,496	0,374	0,469	0,497
SNP_C2	0,356	0,418	0,463	0,326	0,404	0,410	0,280	0,271	0,339	0,299	0,313	0,365	0,329	0,356	0,415
SNP_C10	0,308	0,330	0,380	0,200	0,149	0,226	0,120	0,073	0,129	0,150	0,097	0,163	0,240	0,194	0,279
SNP_C11	0,125	0,078	0,134	0,373	0,447	0,496	0,253	0,219	0,299	0,323	0,292	0,405	0,272	0,202	0,324
SNP_C13	0,371	0,490	0,493	0,344	0,426	0,441	0,373	0,537	0,495	0,367	0,504	0,485	0,374	0,498	0,499
SNP_C17	0,372	0,369	0,494	0,355	0,468	0,462	0,370	0,474	0,490	0,366	0,469	0,482	0,369	0,427	0,488
SNP_C18	0,375	0,553	0,499	0,372	0,478	0,493	0,330	0,385	0,418	0,350	0,413	0,452	0,366	0,472	0,482
SNP_C19	0,206	0,155	0,233	0,368	0,575	0,487	0,301	0,281	0,365	0,336	0,382	0,428	0,301	0,287	0,369
SNP_C20	0,352	0,408	0,456	0,310	0,319	0,383	0,119	0,072	0,128	0,211	0,152	0,240	0,297	0,258	0,364
SNP_C12	0,345	0,466	0,443	0,358	0,489	0,467	0,290	0,289	0,354	0,322	0,352	0,403	0,333	0,399	0,422
SNP_C29	0,342	0,350	0,438	0,372	0,489	0,494	0,375	0,490	0,499	0,374	0,493	0,498	0,372	0,433	0,495
Media	0,331	0,394	0,428	0,333	0,424	0,431	0,293	0,338	0,372	0,315	0,366	0,404	0,331	0,378	0,426
(SD)	(0,07)	(0,13)	(0,103)	(0,07)	(0,139)	(0,107)	(0,10)	(0,166)	(0,143)	(0,09)	(0,144)	(0,123)	(0,06)	(0,115)	(0,09)

En general se han obtenido para la mayoría de los loci valores de heterocigosidades superiores a 0,3 que es el valor recomendado para considerar un SNP como altamente informativo (Goffaux *et al.* 2005). Los SNPs genotipados del cromosoma 5, 10, 11, 19 y 20 en el cálculo total de la heterocigosidad han mostrado valores por debajo del recomendado. En estos casos se observa un comportamiento diferente de los SNPs en las poblaciones analizadas, mostrando valores mayores de 0,3 en algunas poblaciones o menores en otras.

Las heterocigosidades medias fueron en general altas (0,33 - 0,42) y superiores a las descritas recientemente en 13 razas europeas donde se encontraron valores en el rango de 0,22 a 0,31 (Negrini *et al.* 2009). Los valores más bajos se encontraron en la subpoblación II de Pirenaica (0,33), pudiendo ser debido al bajo número de reproductores que existe en este grupo.

Un comportamiento similar han tenido los valores de PIC, con valores en general superiores a 0,3. Resultados similares han sido encontrados en un estudio con 36 SNPs en animales de raza Holstein-Friesian, donde han encontrado un valor medio de PIC de 0,28 (Schopen *et al.* 2008). Por lo que se pueden considerar altamente informativos y elegibles para formar parte de un panel utilizable en estudios de identificación y trazabilidad.

5.2.2.1.1. Desequilibrio H-W en SNP del cromosoma 26

Al testar la totalidad de las muestras para este SNP se observó que las frecuencias alélicas eran intermedias en todas las razas y poblaciones analizadas. Valores similares fueron encontrados por el autor que describió dicho polimorfismo en las razas German Holstein, German Fleckvieh y German Braunvieh (Werner *et al.* 2004). Sin embargo encontrábamos un 99% de animales heterocigóticos, mostrándose un desequilibrio altamente significativo ($p < 0.001$) al realizar la prueba de χ^2 usando el programa Genepop (Raymond & Rousset 1995). Este hecho resulta sorprendente teniendo en cuenta que trabajábamos con poblaciones en equilibrio.

Con el objetivo de buscar una posible explicación a este hecho se llevó a cabo un BLAST (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) entre la secuencia del GenBank que contenía el SNP g.329C>T (AJ496781) del cromosoma 26 y las bases de datos de secuencias de *Bos Taurus*. Como resultado de este alineamiento se encontraron dos identidades del 99% en el contig NW_001503883, cada una contenía una C o una T en la posición correspondiente al SNP (Figura 21 y 22).

Estas dos identidades se identificaban con los genes LOC784256 (6164pb) y LOC784299 (1161pb) derivados mediante análisis computacional automatizado usando el método de predicción GNOMON localizados en la cadena inversa del cromosoma 26, correspondiendo a las secuencias de mRNA XM_001250368.1 y XM_001250420.1 (Figura 23) ambas similares al Citocromo P450 17A1 (CYPXVII) (P450-C17) (P450c17) (Steroid 17-alpha-hydroxylase/17,20 lyase).

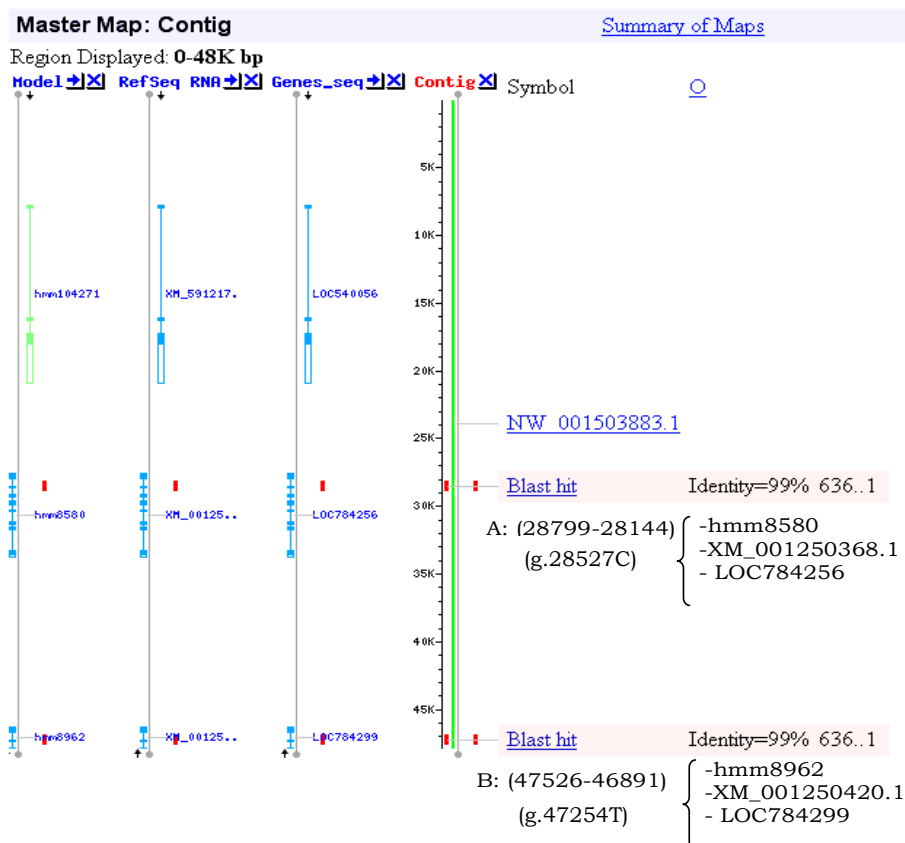


Figura 21: Identities en el contig NW_001503883.

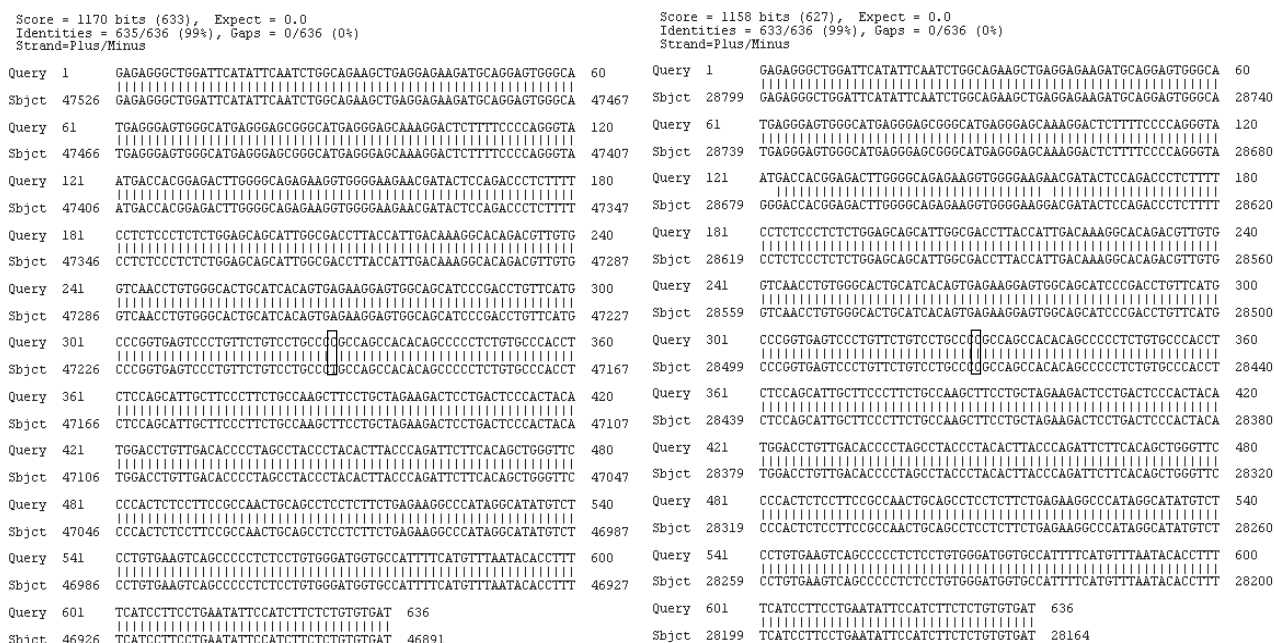


Figura 22: Alineamiento entre la secuencia que contiene el SNP (AJ496781) y la secuencia genómica de Bos taurus NW_001503883

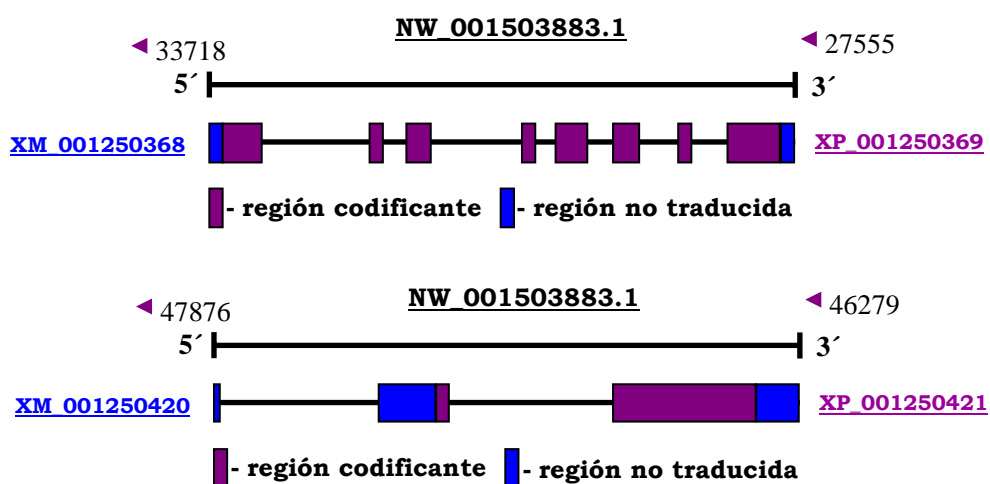


Figura 23: Estructura de los genes que comprendían las identidades.

El citocromo P-450 (P-450) es el principal responsable del metabolismo oxidativo de los xenobióticos, se trata de una familia de hemoproteínas presentes en numerosas especies, desde bacterias a mamíferos, y de las que ya se han identificado más de 2000 isoformas diferentes, diversos estudios en genomas de vertebrados indican que cada genoma contiene de 50-80 genes que codifican citocromos P450 (Thomas 2007). Todos los P-450s conocidos se nombran siguiendo un criterio común y se agrupan en familias y subfamilias en función de la similitud en la secuencia del ADN que los codifica. Las familias 1, 2 y 3 están constituidas por enzimas encargados de la biotransformación de xenobióticos, mientras que el resto de familias incluyen P-450s que intervienen en la biosíntesis y el metabolismo de compuestos endógenos. El extremo C-terminal de la molécula presenta una conservación mayor de las secuencias de aminoácidos entre los distintos P-450s que la región N-terminal. CYP17 (citocromo P450c17a), codifica la enzima P450c17a con actividad 17-a hidroxilasa y 17, 20-liasa y está involucrada en la conversión de pregnenolona y progesterona a sus respectivos 17-hidroxi metabolitos, así como la subsecuente conversión de estos intermediarios a dehidroepiandrosterona o androstenediona, los precursores de estrona y testosterona.

Debido a que se han descrito numerosas duplicaciones de genes pertenecientes a esta superfamilia (Lundqvist *et al.* 1999; Thomas 2007) y

debido a la importante función metabólica que desempeña este gen en el metabolismo, nuestros resultados permiten plantear la hipótesis de que este gen esté repetido en el genoma de *Bos taurus*.

Con el objetivo de confirmar los resultados obtenidos por BLAST se secuenció primeramente un fragmento que comprendía parte de la región 3' adyacente a los loci LOC784256 y LOC784299, así como parte de la región exónica. Para ello se diseñaron dos oligonucleótidos específicos de cadena directa (Primers 1 y 2 en la Figura 24) y un reverso común en la región homóloga (Primer 3 en la Figura 24). Los resultados del alineamiento entre las secuencias obtenidas para los fragmentos A y B se muestran en la Figura 25, apreciándose que solo muestran identidad del 100% en la región 5' que forma parte del exón, de manera que podemos sostener la idea que son dos fragmentos diferentes.

Posteriormente se secuenció un segundo fragmento que comprendía desde la región 3' adyacente a los loci LOC784256 y LOC784299 hasta el intrón 6 del gen incluyendo la secuencia AJ496781. Se utilizaron oligonucleótidos específicos diseñados en la cadena directa aproximadamente 200pb desde el final de los dos loci (Primers 4 y 5 en la Figura 24) y un reverso común en la cadena inversa (Primer 6 en la Figura 24). Como resultado se obtuvieron dos fragmentos diferentes que nos permitió verificar la identidad de las regiones homólogas correspondientes a los loci LOC784256 y LOC784299, pero no en la región 3' adyacente a dichos loci. Además se encontró como se esperaba una C o una T en las posiciones g.28527 y g.47254 referidas al contig NW_001503883. Por lo que estos resultados sugieren la existencia de los dos loci altamente homólogos conteniendo cada uno una C o una T en las posiciones respectivas donde se había descrito un SNP (Figura 26).

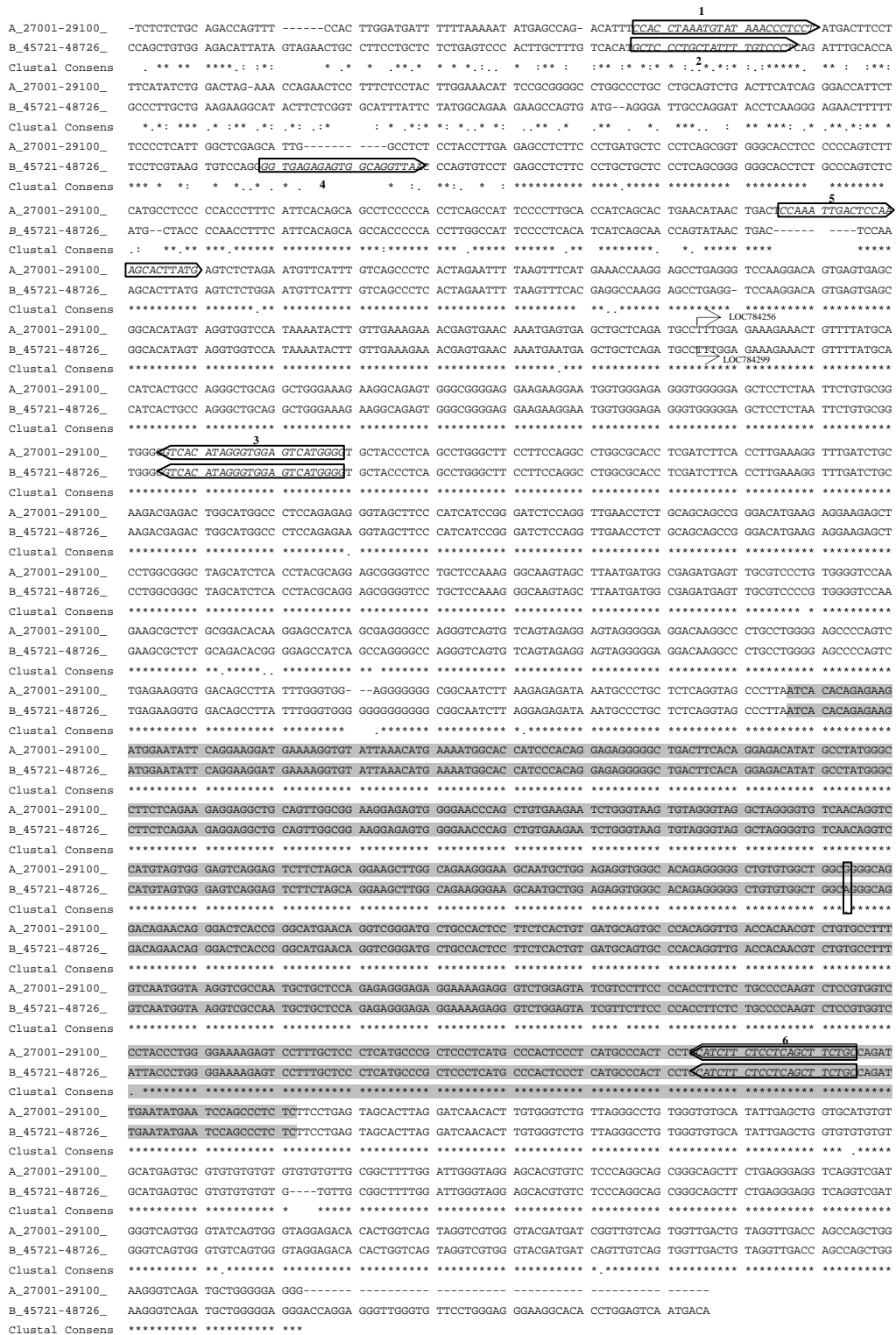


Figura 24: Alineamiento entre las dos regiones A y B del contig NW_001503883. Los primers aparecen numerados, la región sombreada se corresponde con la secuencia AJ496781. El SNP g.329C>T (AJ496781) aparece enmarcado.

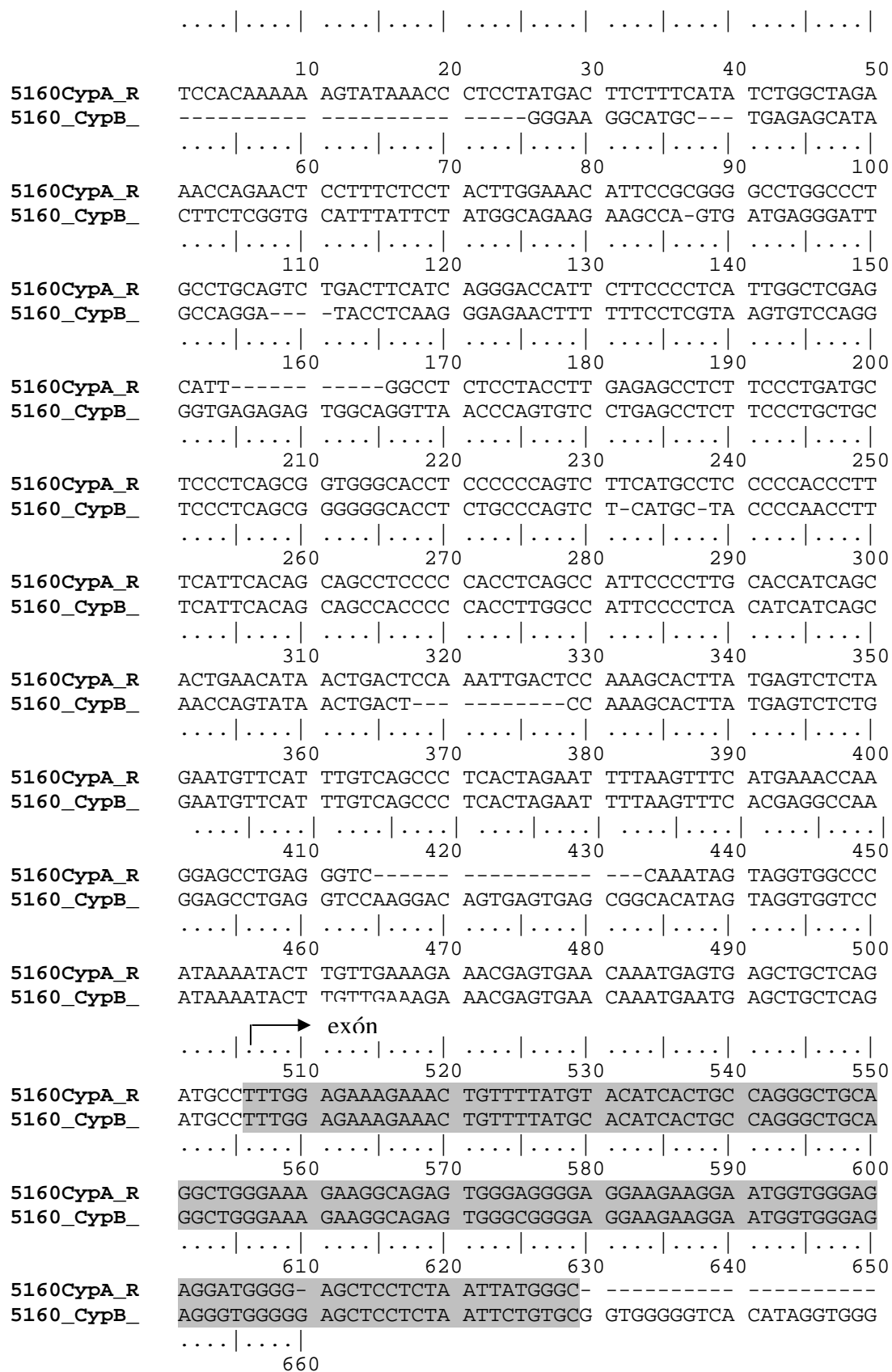


Figura 25: Alineamiento entre los fragmentos A y B obtenidos con los primers 1, 2 y 3.

```

A-NW TCTCTCTGCA GACCAAGTTT CACTTGGATG ATTTTTTAAA AATATGAGCC AGACATTTCC ACCTAAATG TATAAACCCCT
CAP-A -CCAGCTGTG GAG ACATTA TAGTAGAACT GCGTTCCCTGC TCTCTGAGTC CCACTTGCTT TGTCACATGC TOCCTGCTAT
B-NW
CAP-B

A-NW CCTATGACTT CTTTTCATAT CTGGACTAG AAACCAGAA CTCCTTTTCTC CTACTTGGAA ACATT CCG GGGGGCCTGG
CAP-A TTTTGTCCCTC AGATTTGCAC CAGCCCTTGC TGAAGAAAGGC ATACTTCTCG GTGCATTTAT TCTATGGCAG AAGAAAGCCAG
B-NW
CAP-B

A-NW CCCTGC --- CTGCACTCTG ACTTCATCAG GGACCATTCT TCCCCTCATT GGCTCGAGCA TT --- --- GGCCCTCT
CAP-A TGATGAGGGA TTGCCAGGAT ACCTCAAGGG AGAACTTTTT TCCTCGTAAAG TGTCCAAGGGG TGAAGAGATG GCAGGGTTAAC
B-NW
CAP-B

A-NW CCTACCTTGA GAGCCTCTTC CCTGATGCTC CCT CAGCGG TGGGCACCTC CCCCCAGTCT TCATGCTCTC CCCACCCCTT
CAP-A CCAAGTGTCTT GAGCCTCTTC CCTGCTGCTC CCT CAGCGG GGGGCACCTC TGCCAGTCTT CATGC TAC CCCAACCTTT
B-NW GCGCCGCCA AACGTTGTGA GCTTCCAGCGA GTCTTGCCTC TGCCAGTCTT CATGC TCC GCGAACCTTT
CAP-B

A-NW CATTCCACAGC AGCCTCCCCC ACCTCAGCCA TTCCCTTGC ACCATCAGCA CTGAACATAA CTGACTCCAA ATTGACTCCA
CAP-A CATTCCACAGC AGCCACCCCC ACCTTGCCCA TTCCCTTGC ATCATCAGCA ACCAGTATAA CTGACTCCAA A --- ---
B-NW TTTCCACAGC AACCCACCCC ACCTTGCCCA TTCCCTTGC ATCATCAGCA ACCAGTATAA CTGACTCCAA A --- ---
CAP-B

A-NW AAGCACTTAT GAGTCTCTAG AATGTTCAAT TGTCAAGCCTC CACTAGAATT TTAAGTTTCA TGAAACCAAAG GAGCCTGAGG
CAP-A --- --- CAAAACTCTG G --- TAATA CGCAACCCCT CACTAGAATT TTAAGTTTCA TGAAACCAAAG GAGCCTGAGG
B-NW -GCACTTAT GAGTCTCTAG AATGTTCAAT TGTCAAGCCTC CACTAGAATT TTAAGTTTCA TGAAACCAAAG GAGCCTGAGG
CAP-B -GCACTTAT GAGTCTCTAG AATGTTCAAT TGTCAAGCCTC CACTAGAATT TTAAGTTTCA TGAAACCAAAG GAGCCTGAGG

A-NW GTCCAAAGGAC AGTGAAGTGA GCGCACATAG TAGGTGGTCC ATAAAAACT TGTGAAAGA AACCGATGAA CAAATGAGTG
CAP-A GTCCAAAGGAC AGTGGTTGAG CCG ACATGG AAGGTGGTAA ATAAAAACT TGTGAA GA AACCGATGAA CAAATGATTG
B-NW -TCCAAAGGAC AGTGAAGTGA GCGCACATAG TAGGTGGTCC ATAAAAACT TGTGAAAGA AACCGATGAA CAAATGAAATG
CAP-B -TCCAAAGGAC AGTGAAGTGA GCGCACATAG TAGGTGGTCC ATAAAAACT TGTGAAAGA AACCGATGAA CAAATGAAATG

A-NW AGCTGCTCAG ATGCCCTTGG AGAAAAGAAC TGTTTTATGC ACATCACTGC CAGGGCTGCA GAGGGGAAA GAAGGCAGAG
CAP-A AGCTGCTCAA ATGCCCTTGG AGAAAAGAAC TGTTTTATGC ACATCACTGC CAGGGCTGCA GGTGGGAAA GAAGGCAGAG
B-NW AGCTGCTCAG ATGCCCTTGG AGAAAAGAAC TGTTTTATGC ACATCACTGC CAGGGCTGCA GGTGGGAAA GAAGGCAGAG
CAP-B AGCTGCTCAG ATGCCCTTGG AGAAAAGAAC TGTTTTATGC ACATCACTGC CAGGGCTGCA GGTGGGAAA GAAGGCAGAG

A-NW TGGGCGGGGA GGAAGAAAGGA ATGGTGGGAG AGGGTGGGGG AGCTCCTCTA ATTCTGTGCG GTGGGGGTCA CATAGGGTGG
CAP-A TGGGCGGGGA GGAAGAAAGGA ATGGTGGGAG AGGGTGGGGG AGCTCCTCTA ATTCTGTGCG GTGGGGGTCA CATAGGGTGG
B-NW TGGGCGGGGA GGAAGAAAGGA ATGGTGGGAG AGGGTGGGGG AGCTCCTCTA ATTCTGTGCG GTGGGGGTCA CATAGGGTGG
CAP-B TGGGCGGGGA GGAAGAAAGGA ATGGTGGGAG AGGGTGGGGG AGCTCCTCTA ATTCTGTGCG GTGGGGGTCA CATAGGGTGG

A-NW AGTCAATGGGG TGCTACCCCTC AGCCTGGGCT TCCCTCCAGG CCTGGCGCAC CTCGATCTTC ACCTTGAAAAG GTTTGATCTG
CAP-A AGTCAATGGGG TGCTACCCCTC AGCCTGGGCT TCCCTCCAGG CCTGGCGCAC CTCGATCTTC ACCTTGAAAAG GTTTGATCTG
B-NW AGTCAATGGGG TGCTACCCCTC AGCCTGGGCT TCCCTCCAGG CCTGGCGCAC CTCGATCTTC ACCTTGAAAAG GTTTGATCTG
CAP-B AGTCAATGGGG TGCTACCCCTC AGCCTGGGCT TCCCTCCAGG CCTGGCGCAC CTCGATCTTC ACCTTGAAAAG GTTTGATCTG

A-NW CAAAGCAGAGA CTGGCATGGC CCTCCAGAGA GGGTAGCTTC CCATCATCCG GGATCTCCAG GTTGAACCTC TCGACGAGCC
CAP-A CAAAGCAGAGA CTGGCATGGC CCTCCAGAGA GGGTAGCTTC CCATCATCCG GGATCTCCAG GTTGAACCTC TCGACGAGCC
B-NW CAAAGCAGAGA CTGGCATGGC CCTCCAGAGA GGGTAGCTTC CCATCATCCG GGATCTCCAG GTTGAACCTC TCGACGAGCC
CAP-B CAAAGCAGAGA CTGGCATGGC CCTCCAGAGA GGGTAGCTTC CCATCATCCG GGATCTCCAG GTTGAACCTC TCGACGAGCC

A-NW GGGACATGAA GAGGAAAGGC TCCTGGCGGG CTAGCATCTC ACCTACGCAAG GAGCGGGGTC CTGCTCCAAA GGGCAAGTAG
CAP-A GGGACATGAA GAGGAAAGGC TCCTGGCGGG CTAGCATCTC ACCTACGCAAG GAGCGGGGTC CTGCTCCAAA GGGCAAGTAG
B-NW GGGACATGAA GAGGAAAGGC TCCTGGCGGG CTAGCATCTC ACCTACGCAAG GAGCGGGGTC CTGCTCCAAA GGGCAAGTAG
CAP-B GGGACATGAA GAGGAAAGGC TCCTGGCGGG CTAGCATCTC ACCTACGCAAG GAGCGGGGTC CTGCTCCAAA GGGCAAGTAG

A-NW CTTAATGATG GCGAGATGAG TTGCGTCCCT GTGGGGTCCA AGAAGCGCTC TCGGACACA AGGAGCCATC AGCGAGGGGC
CAP-A CTTAATGATG GCGAGATGAG TTGCGTCCCT GTGGGGTCCA AGAAGCGCTC TCGGACACA AGGAGCCATC AGCGAGGGGC
B-NW CTTAATGATG GCGAGATGAG TTGCGTCCCT GTGGGGTCCA AGAAGCGCTC TCGGACACA AGGAGCCATC AGCGAGGGGC
CAP-B CTTAATGATG GCGAGATGAG TTGCGTCCCT GTGGGGTCCA AGAAGCGCTC TCGGACACA AGGAGCCATC AGCGAGGGGC

A-NW CAGGGTCACT GTCAAGTAGAG GAGTAGGGGG AGGACAAAGC CCTGCCTGGG GAGCCCCAGT CTGAGAAAGT GGACAGCCTT
CAP-A CAGGGTCACT GTCAAGTAGAG GAGTAGGGGG AGGACAAAGC CCTGCCTGGG GAGCCCCAGT CTGAGAAAGT GGACAGCCTT
B-NW CAGGGTCACT GTCAAGTAGAG GAGTAGGGGG AGGACAAAGC CCTGCCTGGG GAGCCCCAGT CTGAGAAAGT GGACAGCCTT
CAP-B CAGGGTCACT GTCAAGTAGAG GAGTAGGGGG AGGACAAAGC CCTGCCTGGG GAGCCCCAGT CTGAGAAAGT GGACAGCCTT

A-NW ATTTGGGTGG AGGGGGGGCG GCAATCTTAA GAGAGA --- T AAATGCCCTG CTCTCAGGTA GCCCTTAAAT ACACAGAGAA
CAP-A ATTTGGGTGG AGGGGGGGCG GCAATCTTAA GAGAGA --- T AAATGCCCTG CTCTCAGGTA GCCCTTAAAT ACACAGAGAA
B-NW ATTTGGGTGG AGGGGGGGCG GCGGCAATCT TAGGAGAGAT AAATGCCCTG CTCTCAGGTA GCCCTTAAAT ACACAGAGAA
CAP-B ATTTGGGTGG AGGGGGGGCG GCGGCAATCT TAGGAGAGAT AAATGCCCTG CTCTCAGGTA GCCCTTAAAT ACACAGAGAA

A-NW GATGGAAATAT TCAAGAAAGGA TGAAGAAAGT TATTAACAT GAAAAATGGCA CCATCCACA GAGAGAGGGG CTGACTTCAC
CAP-A GATGGAAATAT TCAAGAAAGGA TGAAGAAAGT TATTAACAT GAAAAATGGCA CCATCCACA GAGAGAGGGG CTGACTTCAC
B-NW GATGGAAATAT TCAAGAAAGGA TGAAGAAAGT TATTAACAT GAAAAATGGCA CCATCCACA GAGAGAGGGG CTGACTTCAC
CAP-B GATGGAAATAT TCAAGAAAGGA TGAAGAAAGT TATTAACAT GAAAAATGGCA CCATCCACA GAGAGAGGGG CTGACTTCAC

A-NW AGGAGACATA TGCCCTATGGG CCTTCTCAGA AGAGGAGGCT GCAGTTGGCG GAAAGGAGAT GGGGAAACCA GCTGTGAAGA
CAP-A AGGAGACATA TGCCCTATGGG CCTTCTCAGA AGAGGAGGCT GCAGTTGGCG GAAAGGAGAT GGGGAAACCA GCTGTGAAGA
B-NW AGGAGACATA TGCCCTATGGG CCTTCTCAGA AGAGGAGGCT GCAGTTGGCG GAAAGGAGAT GGGGAAACCA GCTGTGAAGA
CAP-B AGGAGACATA TGCCCTATGGG CCTTCTCAGA AGAGGAGGCT GCAGTTGGCG GAAAGGAGAT GGGGAAACCA GCTGTGAAGA

A-NW ATCTGGGTAA GTGTAGGGTA GGTAGGGGT GTCAACAGGT CCATGTAAGT GAGTCAAGGA GTCTTCTAGC AGGAAAGCTTG
CAP-A ATCTGGGTAA GTGTAGGGTA GGTAGGGGT GTCAACAGGT CCATGTAAGT GAGTCAAGGA GTCTTCTAGC AGGAAAGCTTG
B-NW ATCTGGGTAA GTGTAGGGTA GGTAGGGGT GTCAACAGGT CCATGTAAGT GAGTCAAGGA GTCTTCTAGC AGGAAAGCTTG
CAP-B ATCTGGGTAA GTGTAGGGTA GGTAGGGGT GTCAACAGGT CCATGTAAGT GAGTCAAGGA GTCTTCTAGC AGGAAAGCTTG

A-NW CCAAGAAAGGA AGCAATGCTG GAGAAGTGG CACAGAGGGG GCTGTGTGGC TGGGAGGGCA GACAGAAAC GGGACTCAC
CAP-A CCAAGAAAGGA AGCAATGCTG GAGAAGTGG CACAGAGGGG GCTGTGTGGC TGGGAGGGCA GACAGAAAC GGGACTCAC
B-NW CCAAGAAAGGA AGCAATGCTG GAGAAGTGG CACAGAGGGG GCTGTGTGGC TGGGAGGGCA GACAGAAAC GGGACTCAC
CAP-B CCAAGAAAGGA AGCAATGCTG GAGAAGTGG CACAGAGGGG GCTGTGTGGC TGGGAGGGCA GACAGAAAC GGGACTCAC

A-NW GGGCATGAAC AGGTGGGGAT GCTGCCACTC CTTCTCACTG TGATGCAAGT CCCACAGGTT GACCAACAAC TCTGTGCCTT
CAP-A GGGCATGAAC AGGTGGGGAT GCTGCCACTC CTTCTCACTG TGATGCAAGT CCCACAGGTT GACCAACAAC TCTGTGCCTT
B-NW GGGCATGAAC AGGTGGGGAT GCTGCCACTC CTTCTCACTG TGATGCAAGT CCCACAGGTT GACCAACAAC TCTGTGCCTT
CAP-B GGGCATGAAC AGGTGGGGAT GCTGCCACTC CTTCTCACTG TGATGCAAGT CCCACAGGTT GACCAACAAC TCTGTGCCTT

A-NW TGTCAATGGT AAGGTCGCCA ATGCTGCTCC AGAGAGGGAG AGGAAAAAGG GGTCTGGAAT ATCGTCTTC CCCACCTTCT
CAP-A TGTCAATGGT AAGGTCGCCA ATGCTGCTCC AGAGAGGGAG AGGAAAAAGG GGTCTGGAAT ATCGTCTTC CCCACCTTCT
B-NW TGTCAATGGT AAGGTCGCCA ATGCTGCTCC AGAGAGGGAG AGGAAAAAGG GGTCTGGAAT ATCGTCTTC CCCACCTTCT
CAP-B TGTCAATGGT AAGGTCGCCA ATGCTGCTCC A

A-NW CTGCCCCAAG TCTCCGTGGT CCTACCCCTG GGGAAAAAGG TCCTTTGCTC CCTCATGCC GCTCCCTCAT GCCCCCTCCC
CAP-A CTGCCCCAAG TCTCCGTGGT CATTACCCCTG GGGAAAAAGG TCCTTTGCTC CCTCATGCC GCTCCCTCAT GCCCCCTCCC
B-NW CTGCCCCAAG TCTCCGTGGT CATTACCCCTG GGGAAAAAGG TCCTTTGCTC CCTCATGCC GCTCCCTCAT GCCCCCTCCC
CAP-B CTGCCCCAAG TCTCCGTGGT CATTACCCCTG GGGAAAAAGG TCCTTTGCTC CCTCATGCC GCTCCCTCAT GCCCCCTCCC

A-NW TCATGCCAC TCCTGCATCT TCTCCTCAGC TTCTGCCAGA TTGAATATGA ATCCAGCCCT CTCT
CAP-A TCATGCCAC TCCTGCATCT TCTCCTCAGC TTCTGCCAGA TTGAATATGA ATCCAGCC
B-NW TCATGCCAC TCCTGCATCT TCTCCTCAGC TTCTGCCAGA TTGAATATGA ATCCAGCC
CAP-B TCATGCCAC TCCTGCATCT TCTCCTCAGC TTCTGCCAGA TTGAATATGA ATCCAGCC

```

Figura 26: Alineamiento fragmentos amplificados A, B (con primers 3, 4 y 5) y las dos regiones del contig NW_001503883.

Estos resultados nos permiten explicar la desviación altamente significativa del equilibrio de Hardy-Weinberg provocada por el exceso de heterocigotos en la población. Por esto, este polimorfismo ha sido eliminado del panel de SNPs para estudios de trazabilidad a pesar de estar recomendado por otros trabajos. Además de eliminar este SNP se recomienda no utilizar ningún otro marcador ubicado en esta región con el fin de no introducir errores en la identificación de individuos.

Por otra parte dada la elevada homología que existe entre el gen bovino y su ortólogo humano, se recomienda tener especial cuidado con este gen en estudios en humanos. Este gen al tener una importante función en el metabolismo tiene implicaciones en diversos procesos en humanos, de hecho se han llevado a cabo estudios de asociación de este gen CYP17A1 con cáncer de próstata o mamario (Sarma *et al.* 2008; Severi *et al.* 2008) en los que no hay resultados definitivos de dicha asociación.

Es importante tener en cuenta que al revisar las bases de datos de secuencias se hacen llamadas de atención de este gen acerca de su alineamiento múltiple en diferentes contig, pudiendo conducir a errores en la interpretación de los resultados debido a la posibilidad de duplicación de este gen.

5.2.2.1.2. Análisis de variabilidad en SNP del cromosoma 1

En la selección de los polimorfismos que serían analizados como parte de la búsqueda de un panel para ser empleado en identificación y trazabilidad fueron elegidos cuatro SNPs pertenecientes a regiones exónicas dentro de los que estaba el polimorfismo en el codón 12 de la transferrina bovina en el cromosoma 1. Los tres polimorfismos restantes ubicados en regiones exónicas habían sido estudiados anteriormente por su asociación con caracteres cuantitativos.

Debido a la función fisiológica de la transferrina así como por estar ubicada en una región involucrada en varios caracteres cuantitativos, se ha

estudiado la variabilidad de este polimorfismo en una muestra de vacas con valores estimados de producción extremos en cantidad de grasa en la leche.

El genotipado del SNP c.1455A>G (U02564) del exón 12 del gen de la transferrina bovina para verificar si existía asociación con la cantidad de grasa en la leche se realizó mediante PCR alelo específica, utilizando PCR estándar o a tiempo real.

En la Figura 27 se puede apreciar una imagen de PCR a tiempo real con los posibles resultados del genotipado, donde una diferencia de 6 ciclos de amplificación de un alelo a otro era suficiente para el correcto genotipado de las muestras (Germer *et al.* 2000); para la validación de los genotipos se utilizaron controles obtenidos por secuenciación.

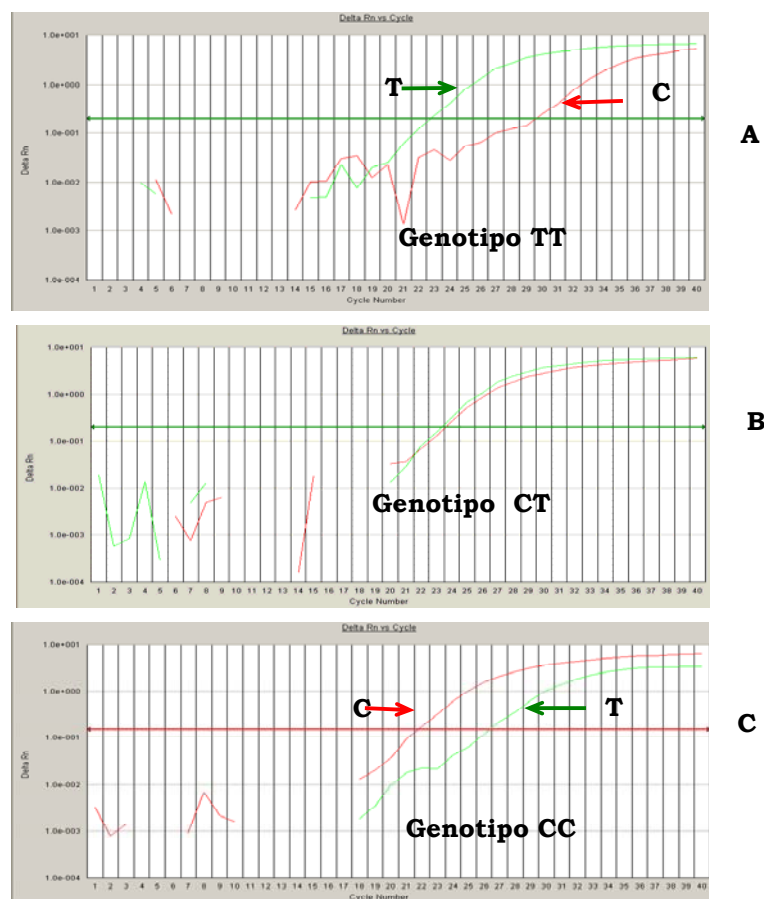


Figura 27: Imágenes de las reacciones de PCR a tiempo real para el genotipado. A: muestra TT. B: muestra CT. C: muestra CC.

A partir del resultado de los genotipos se calcularon las frecuencias alélicas en ambos grupos de muestras, los resultados se muestran en la Tabla 35.

Tabla 35: Frecuencias alélicas del SNP por grupos de muestras.

Alelos		
	C	T
Cola Alta (n=117)	0,517	0,483
Cola baja (n= 94)	0,351	0,649
Total (n=211)	0,44	0,56

El análisis de asociación del SNP con el contenido de grasa en la leche en los dos grupos de animales de raza Holstein-Friesian se realizó usando un contraste binomial de proporciones mediante una prueba chi cuadrado. Como resultado obtuvimos un valor de X^2 de 11,65 ($p < 0,0006$) evidenciándose una clara asociación entre este SNP y la cantidad de grasa en la leche, estando el alelo T asociado a baja cantidad y el C a una alta cantidad (Sanz *et al.* 2010).

Este polimorfismo produce un cambio aminoacídico Asp→Gly en la posición 469 de la transferina bovina, sin embargo este aminoácido no parece estar involucrado en la función de unión al hierro de esta proteína o cualquier otra función que no sea estructural. Para intentar explicar la clara asociación que se produce nos planteamos la búsqueda de algún otro polimorfismo situado en esta región y verificar si se producía también alguna asociación.

Dada la presencia del microsatélite BM1824 empleado en el estudio de trazabilidad y que está mapeado en el cromosoma 1 en 122.391 cM a solo 3.514 cM del gen de la transferrina se realizó el genotipado de este marcador mediante PCR fluorescente con posterior visualización mediante electroforesis capilar en un secuenciador ABI Prism 3130 (*Applied Biosystems*), de esta manera verificaríamos la asociación con la cantidad de

grasa en la leche en los dos grupos de muestras y comprobaríamos el desequilibrio de ligamiento con el SNP.

Los resultados de las frecuencias alélicas y los resultados de la asociación entre el microsatélite y la cantidad de grasa se muestran en la Tabla 36, evidenciándose asociación de los alelos 180 y 188 con la cantidad de grasa en la leche ($p < 0.05$).

Tabla 36: Asociación entre Microsatélite y Cola de grasa.

Frecuencia génica				
Alelo	Valores P (error tipo I)	Cola1(Alta)	Cola2(Baja)	Dif.
A(178)	0.4319	0.1304	0.1576	-0.0272
B(180)	0.0210	0.3130	0.2120	0.1010
C(182)	0.3537	0.2043	0.1685	0.0358
D(188)	0.0180	0.3217	0.4348	-0.1131
E(190)	0.8442	0.0304	0.0272	0.0032

Partiendo del hecho de la existencia de la asociación se llevó a cabo el análisis del desequilibrio de ligamiento entre el SNP de la transferrina y el microsatélite BM1824 encontrándose un desequilibrio altamente significativo ($p < 0,001$) de los alelos 180 y 190 del microsatélite con el SNP en el grupo seleccionado para la cola alta de cantidad de grasa en la leche (Tabla 37).

Tabla 37: Desequilibrio de ligamiento entre SNP de transferrina y microsatélite.

Valores P (error tipo I)		
Alelo	Cola1(Alta)	Cola2(Baja)
A(178)	0.1676	0.0724
B(180)	0.0008	0.4016
C(182)	0.4482	0.3985
D(188)	0.1353	0.5884
E(190)	0.0054	0.4676

Si bien existe asociación entre el microsatélite con la cantidad de grasa esta asociación debe estar provocada por el elevado desequilibrio de ligamiento que existe entre estos loci.

La influencia de la transferrina en la cantidad de grasa en la leche podría ser debido a la actividad lipolítica de esta que ha sido demostrada en numerosos trabajos (Rumberger *et al.* 2004; Green *et al.* 2006). Estudios recientes han demostrado el efecto de la transferrina y del hierro en la actividad lipolítica del suero, posiblemente a través de la reacción de Fenton catalizada por el hierro mediante la formación de radicales libres, como por ejemplo los radicales hidroxilos (Rumberger *et al.* 2004). Estos radicales pueden inducir la peroxidación lipídica que según se ha demostrado incrementa la lipólisis (Rejholcova *et al.* 1988). Por lo que este incremento en la tasa de lipólisis inducida por la transferrina unida al hierro podría resultar en el incremento de los ácidos grasos libres circulando en sangre (Rumberger *et al.* 2004).

En el proceso metabólico de secreción de leche en la ubre se necesitan una gran cantidad de nutrientes traídos a la ubre por la sangre. Aproximadamente el 17-45% de la grasa en la leche se forma a partir del acetato y el 8-25% a partir del butirato circulantes en sangre, hecho que podría explicar al menos parcialmente la asociación entre el SNP de la transferrina con la cantidad de grasa, debido a que la cantidad de ácidos grasos libres circulando en la sangre es superior a la basal debido al incremento de la tasa de lipólisis.

Mediante el estudio de la variabilidad de este polimorfismo en la muestra seleccionada se puede llegar a conclusiones preliminares acerca de la posible implicación de este polimorfismo en la cantidad de grasa en la leche que presenten los animales que presenten el genotipo favorable, siendo una aplicación adicional de los SNPs a considerar en análisis de identificación, verificación de trazabilidad o en los programas de selección.

5.2.2.2. Probabilidades de exclusión

Los valores de PE por locus y combinados para las poblaciones estudiadas se muestran en la Tabla 38. Se puede observar que el panel de 19 SNPs tiene un bajo poder de exclusión, con un valor de 0,9696 en la población total. Comparando nuestros resultados con los obtenidos por otros autores analizando paneles con un mayor número de marcadores, podemos decir que nuestros resultados son satisfactorios. Van Eenennaam y colaboradores (2007) obtuvieron una PEcomb de 0,956 utilizando 28 SNPs o Herraiz y colaboradores (2005) analizando 43 SNPs obtuvieron una PEcomb de 0,9772. Estos resultados podrían ser debidos a la elevada heterocigosidad que muestran en general estos polimorfismos en las poblaciones estudiadas.

Tabla 38: Probabilidades de exclusión por locus y combinado.

Locus	Parda de Montaña	Pirenaica I	Pirenaica II	Pirenaica I+II	Total
SNP_C1	0,18457	0,18688	0,16789	0,17766	0,18110
SNP_C4	0,18672	0,16869	0,18545	0,18230	0,18463
SNP_C5	0,12088	0,03762	0,01959	0,02577	0,07673
SNP_C6	0,17695	0,16407	0,16014	0,16149	0,16922
SNP_C7	0,18748	0,18748	0,10815	0,16243	0,18039
SNP_C8	0,17768	0,18740	0,18717	0,18743	0,18542
SNP_C21	0,16268	0,18268	0,17580	0,17854	0,17319
SNP_C23	0,18614	0,17807	0,18388	0,18240	0,18431
SNP_C24	0,18711	0,17922	0,18749	0,18656	0,18682
SNP_C2	0,17776	0,16298	0,14024	0,14934	0,16455
SNP_C10	0,15403	0,10010	0,05983	0,07501	0,11987
SNP_C11	0,06240	0,18652	0,12628	0,16164	0,13577
SNP_C13	0,18573	0,17190	0,18635	0,18367	0,18716
SNP_C17	0,18608	0,17766	0,18485	0,18298	0,18445
SNP_C18	0,18732	0,18576	0,16483	0,17491	0,18296
SNP_C19	0,11637	0,18417	0,15040	0,16807	0,18429
SNP_C20	0,17602	0,15481	0,05933	0,10557	0,14870
SNP_C12	0,17243	0,17896	0,14491	0,16090	0,16644
SNP_C29	0,17100	0,18586	0,18734	0,18702	0,18617
PEcomb	0,9688	0,9690	0,95196	0,96237	0,9696

Se han ordenado los valores de probabilidad de exclusión por locus considerando la población total (Tabla 39). Se ha comprobado, como era de esperar, que los que muestran los mayores poderes de exclusión se corresponden con los locus más polimórficos.

Tabla 39: Valores teóricos de probabilidades de exclusión de los diferentes SNPs calculados en la población total ordenados por su grado de polimorfismo.

Locus	PE
SNP_C13	0,18716
SNP_C24	0,18682
SNP_C29	0,18617
SNP_C8	0,18542
SNP_C4	0,18463
SNP_C17	0,18445
SNP_C23	0,18431
SNP_C19	0,18429
SNP_C18	0,18296
SNP_C1	0,18110
SNP_C7	0,18039
SNP_C21	0,17319
SNP_C6	0,16922
SNP_C12	0,16644
SNP_C2	0,16455
SNP_C20	0,14870
SNP_C11	0,13577
SNP_C10	0,11987
SNP_C5	0,07673

Las probabilidades de exclusión parentales (PEP1, PEP2 y PEP3) fueron calculadas asumiendo las tres situaciones diferentes descritas en el capítulo de Material y Métodos, utilizando la información de genealogías disponible. Los resultados se muestran en la Tabla 40.

Tabla 40: Probabilidades de exclusión parentales acumulativas calculadas para diferentes tríos de familias de las razas Parda de Montaña y Pirenaica.

	Parda de Montaña	Pirenaica
PEP.1	0,962246553	0,9674028
PEP.2	0,838522169	0,8410522
PEP.3	0,995325001	0,9961394

El poder de exclusión parental acumulativo (PEP) es una medida de la probabilidad de la asignación parental correcta. Los resultados obtenidos con los 19 SNPs muestran el bajo poder de este panel. Se puede apreciar que en ambas poblaciones el menor valor se obtiene para el caso PEP.2 donde a partir de los genotipos de un progenitor y un descendiente se excluye la relación entre ellos. El mayor valor se obtiene para el caso PEP.3 que excluye ambos progenitores a partir de los genotipos de ambos progenitores y el descendiente. En un estudio en la raza Galloway, en el que han sido analizados 43 SNPs se han encontrado valores de 0,9929 para PEP1, 0,9497 para PEP2 y 0,9999 para PEP3 (Herraez *et al.* 2005), de manera que nuestros resultados se pueden considerar satisfactorios ya que hemos analizado aproximadamente la mitad SNPs.

5.2.2.3. Probabilidades de identidad teóricas y empíricas

Desde el punto de vista de trazabilidad el cálculo del valor de la probabilidad de identidad es un valor muy útil. Se han calculado estos valores por locus y los resultados se muestran en la Tabla 41 ordenados en orden ascendente de probabilidad. Se observa que las probabilidades de identidad para los diferentes loci difieren entre poblaciones, corroborándose el hecho de que los SNPs manifiestan comportamientos muy diferentes entre poblaciones.

Tabla 41: Probabilidades de identidad teóricas por locus.

Parda de Montaña		Pirenaica I		Pirenaica II		Pirenaica I+II		Total	
SNP_C7	0,375	SNP_C7	0,375	SNP_C24	0,375	SNP_C8	0,3751	SNP_C13	0,3757
SNP_C18	0,3754	SNP_C8	0,3752	SNP_C29	0,3753	SNP_C29	0,376	SNP_C24	0,3764
SNP_C24	0,3758	SNP_C1	0,3762	SNP_C8	0,3757	SNP_C24	0,3769	SNP_C29	0,3777
SNP_C4	0,3766	SNP_C11	0,377	SNP_C13	0,3774	SNP_C13	0,3829	SNP_C8	0,3792
SNP_C23	0,3777	SNP_C29	0,3783	SNP_C4	0,3788	SNP_C17	0,3843	SNP_C4	0,3809
SNP_C17	0,3779	SNP_C18	0,3785	SNP_C17	0,3799	SNP_C23	0,3856	SNP_C17	0,3812
SNP_C13	0,3786	SNP_C19	0,3818	SNP_C23	0,3819	SNP_C4	0,3858	SNP_C23	0,3815
SNP_C1	0,381	SNP_C21	0,385	SNP_C21	0,4012	SNP_C21	0,3941	SNP_C18	0,3844
SNP_C2	0,3959	SNP_C24	0,3926	SNP_C1	0,4212	SNP_C1	0,3961	SNP_C1	0,3884
SNP_C8	0,3961	SNP_C12	0,3932	SNP_C18	0,426	SNP_C18	0,4025	SNP_C7	0,39
SNP_C6	0,3977	SNP_C23	0,3952	SNP_C6	0,4421	SNP_C19	0,4191	SNP_C21	0,4066
SNP_C20	0,3999	SNP_C17	0,3961	SNP_C19	0,4703	SNP_C7	0,4336	SNP_C6	0,4162
SNP_C12	0,4084	SNP_C13	0,4097	SNP_C12	0,4803	SNP_C11	0,4357	SNP_C12	0,4232
SNP_C29	0,4119	SNP_C4	0,4176	SNP_C2	0,4939	SNP_C6	0,436	SNP_C2	0,4281
SNP_C21	0,4329	SNP_C6	0,4293	SNP_C11	0,5366	SNP_C12	0,4376	SNP_C19	0,4666
SNP_C10	0,4562	SNP_C2	0,4321	SNP_C7	0,5955	SNP_C2	0,4694	SNP_C20	0,4712
SNP_C5	0,556	SNP_C20	0,4541	SNP_C10	0,767	SNP_C20	0,6065	SNP_C11	0,5094
SNP_C19	0,6158	SNP_C10	0,6251	SNP_C20	0,7688	SNP_C10	0,7133	SNP_C10	0,5593
SNP_C11	0,7593	SNP_C5	0,8526	SNP_C5	0,9217	SNP_C5	0,8983	SNP_C5	0,7071

Se ha calculado cómo varía la probabilidad de identidad a medida que se incrementa el número de loci genotipados (Tabla 42). Se puede ver que no en todas las poblaciones la informatividad de los marcadores es la misma, lo cual está relacionado con la estructura genética de cada población tal y como se ha discutido para el caso de los marcadores microsatélites.

Considerando el panel completo de SNPs se observa que la probabilidad de encontrar dos individuos en la población que tengan los mismos genotipos es del orden de 10^{-8} , siendo de $8,83E-08$ para la población de Parda de Montaña y de $2,01E-07$ para la población Pirenaica I+II, para la totalidad de las muestras la probabilidad de identidad fue de $8,87E-08$ (Tabla 43), es decir la probabilidad de encontrar individuos que compartan los mismos perfiles genéticos considerando la población total es de 1:11,27 millones.

Tabla 42: Variación de la probabilidad de identidad teórica con el incremento de locus genotipados.

Nº. SNPs	Parda de Montaña		Pirenaica I		Pirenaica II		Pirenaica I+II		Total	
	Locus	Incremento PI	Locus	Incremento PI	Locus	Incremento PI	Locus	Incremento PI	Locus	Incremento PI
1	SNP_C7	0,375	SNP_C7	0,375	SNP_C24	0,375	SNP_C8	0,3751	SNP_C13	0,3757
2	SNP_C18	0,1407721	SNP_C8	0,140716	SNP_C29	0,14075	SNP_C29	0,1410367	SNP_C24	0,1413967
3	SNP_C24	5,29E-02	SNP_C1	5,29E-02	SNP_C8	5,29E-02	SNP_C24	5,32E-02	SNP_C29	5,34E-02
4	SNP_C4	0,0199202	SNP_C11	2,00E-02	SNP_C13	2,00E-02	SNP_C13	2,04E-02	SNP_C8	2,03E-02
5	SNP_C23	7,52E-03	SNP_C29	7,55E-03	SNP_C4	7,56E-03	SNP_C17	7,82E-03	SNP_C4	7,71E-03
6	SNP_C17	2,84E-03	SNP_C18	2,86E-03	SNP_C17	2,87E-03	SNP_C23	3,02E-03	SNP_C17	2,94E-03
7	SNP_C13	1,08E-03	SNP_C19	1,09E-03	SNP_C23	1,10E-03	SNP_C4	1,16E-03	SNP_C23	1,12E-03
8	SNP_C1	4,10E-04	SNP_C21	4,20E-04	SNP_C21	4,40E-04	SNP_C21	4,59E-04	SNP_C18	4,31E-04
9	SNP_C2	1,62E-04	SNP_C24	1,65E-04	SNP_C1	1,85E-04	SNP_C1	1,82E-04	SNP_C1	1,68E-04
10	SNP_C8	6,43E-05	SNP_C12	6,49E-05	SNP_C18	7,90E-05	SNP_C18	7,31E-05	SNP_C7	6,53E-05
11	SNP_C6	2,56E-05	SNP_C23	2,56E-05	SNP_C6	3,49E-05	SNP_C19	3,06E-05	SNP_C21	2,66E-05
12	SNP_C20	1,02E-05	SNP_C17	1,02E-05	SNP_C19	1,64E-05	SNP_C7	1,33E-05	SNP_C6	1,11E-05
13	SNP_C12	4,18E-06	SNP_C13	4,16E-06	SNP_C12	7,89E-06	SNP_C11	5,79E-06	SNP_C12	4,68E-06
14	SNP_C29	1,72E-06	SNP_C4	1,74E-06	SNP_C2	3,90E-06	SNP_C6	2,52E-06	SNP_C2	2,00E-06
15	SNP_C21	7,45E-07	SNP_C6	7,46E-07	SNP_C11	2,09E-06	SNP_C12	1,10E-06	SNP_C19	9,35E-07
16	SNP_C10	3,40E-07	SNP_C2	3,22E-07	SNP_C7	1,24E-06	SNP_C2	5,18E-07	SNP_C20	4,40E-07
17	SNP_C5	1,89E-07	SNP_C20	1,46E-07	SNP_C10	9,55E-07	SNP_C20	3,14E-07	SNP_C11	2,24E-07
18	SNP_C19	1,16E-07	SNP_C10	9,15E-08	SNP_C20	7,34E-07	SNP_C10	2,24E-07	SNP_C10	1,25E-07
19	SNP_C11	8,83E-08	SNP_C5	7,80E-08	SNP_C5	6,76E-07	SNP_C5	2,01E-07	SNP_C5	8,87E-08

Tabla 43: Valores teóricos de probabilidades de identidad del panel de 19 SNPs.

Población	PIt
Parda de Montaña	$8,83 \times 10^{-8}$
Pirenaica I	$7,80 \times 10^{-8}$
Pirenaica II	$6,76 \times 10^{-7}$
Pirenaica I+II	$2,01 \times 10^{-7}$
Total	$8,87 \times 10^{-8}$

Las probabilidades de identidad teóricas obtenidas con nuestro panel de 19 SNPs han sido menores a las publicadas para otras poblaciones europeas (Mountain Brown, Piedmontese, Italian Holstein, Belgian Blue, Marchigiana, Charolais, Italian Red Pied, Aosta Red Pied, Rendena, Chianina y Reggiana) donde con un panel de 16 SNPs se obtuvieron valores del orden de 10^{-4} cuando se consideraron grupos de individuos de una misma raza o del orden de 10^{-5} para grupos mixtos (Capoferri *et al.* 2006). En estudios en las razas German Holstein, German Fleckvieh, German Braunvieh, Kerry y Angus la PI para 37 SNPs ha sido del orden de 10^{-13} (Werner *et al.* 2004). Así mismo en otros trabajos se han obtenido valores de 2.0×10^{-13} al testar 32 SNPs (Heaton *et al.* 2002) en muestras pertenecientes al MARC Beef Cattle Diversity Panel versión 2.1 (Heaton *et al.* 2001a), o de 1.9×10^{-10} en muestras pertenecientes al MARC Purebred Angus Panel versión 1.0 (Heaton *et al.* 2001a). En la raza Galloway un estudio con 43 SNPs dio un valor de PI de 5.3×10^{-11} .

Teniendo en cuenta el número de marcadores que comprende nuestro panel, podemos afirmar que nuestros resultados son satisfactorios y discriminantes ya que comparado con estudios realizados previamente siempre obtenemos probabilidades menores.

Cuando se realiza el cálculo empírico podemos apreciar que los resultados son más discriminantes ya que se trabaja directamente con los genotipos y no se hace una estimación teórica a partir de las frecuencias alélicas de la población. En la Tabla 44 se muestran las probabilidades de

identidad empíricas para cada SNP ordenadas según el grado de informatividad del marcador. Se puede observar que el comportamiento de los SNPs varía entre poblaciones y que existen diferencias entre estos valores empíricos y los valores teóricos (Tabla 41).

Tabla 44: Probabilidades de identidad empíricas por locus.

Parda de Montaña		Pirenaica I		Pirenaica II		Pirenaica I+II		Total	
SNP_C24	0,3362	SNP_C11	0,346	SNP_C8	0,3441	SNP_C8	0,3571	SNP_C29	0,3563
SNP_C17	0,3394	SNP_C18	0,366	SNP_C24	0,3576	SNP_C29	0,3713	SNP_C24	0,3641
SNP_C4	0,3644	SNP_C7	0,367	SNP_C29	0,3647	SNP_C23	0,3824	SNP_C17	0,3689
SNP_C1	0,3786	SNP_C29	0,371	SNP_C17	0,3785	SNP_C24	0,39	SNP_C8	0,3719
SNP_C23	0,3786	SNP_C1	0,372	SNP_C23	0,3849	SNP_C17	0,3956	SNP_C13	0,3751
SNP_C13	0,3797	SNP_C8	0,38	SNP_C13	0,4034	SNP_C4	0,4079	SNP_C23	0,3797
SNP_C7	0,388	SNP_C23	0,385	SNP_C4	0,4175	SNP_C13	0,4098	SNP_C4	0,3886
SNP_C18	0,4019	SNP_C21	0,424	SNP_C21	0,4384	SNP_C1	0,4292	SNP_C7	0,3887
SNP_C2	0,4137	SNP_C17	0,43	SNP_C1	0,4867	SNP_C21	0,4342	SNP_C18	0,4
SNP_C20	0,4232	SNP_C12	0,431	SNP_C18	0,4963	SNP_C18	0,439	SNP_C1	0,4067
SNP_C8	0,4262	SNP_C19	0,441	SNP_C6	0,5137	SNP_C19	0,4762	SNP_C21	0,4589
SNP_C29	0,4415	SNP_C4	0,453	SNP_C19	0,5621	SNP_C7	0,4882	SNP_C6	0,4747
SNP_C6	0,4415	SNP_C13	0,454	SNP_C12	0,5704	SNP_C11	0,4946	SNP_C12	0,4904
SNP_C12	0,4773	SNP_C6	0,491	SNP_C2	0,6009	SNP_C12	0,5063	SNP_C2	0,4904
SNP_C21	0,5085	SNP_C2	0,507	SNP_C11	0,6546	SNP_C6	0,5074	SNP_C19	0,5551
SNP_C10	0,5405	SNP_C24	0,507	SNP_C7	0,7184	SNP_C2	0,5673	SNP_C20	0,5587
SNP_C5	0,6722	SNP_C20	0,527	SNP_C10	0,8634	SNP_C20	0,7291	SNP_C11	0,6142
SNP_C19	0,735	SNP_C10	0,741	SNP_C20	0,8647	SNP_C10	0,8232	SNP_C10	0,6791
SNP_C11	0,8553	SNP_C5	0,916	SNP_C5	0,9592	SNP_C5	0,946	SNP_C5	0,818

Se han calculado cómo varían las probabilidades de identidad empíricas incrementando el número de loci genotipados y se puede observar que para el caso de la población de Parda de Montaña con un set de 10 marcadores SNPs es suficiente para alcanzar una probabilidad 0, esto no ocurre para las poblaciones de Pirenaica (Tabla 45). En la subpoblación I de Pirenaica a partir de 9 SNPs ya no disminuye más la PI; en la Pirenaica II a partir de 11 y en la Pirenaica total con 14 SNPs. Considerando la población total con un panel de 12 marcadores es suficiente para alcanzar la probabilidad mínima. Evidentemente el número de marcadores requeridos para alcanzar la probabilidad de identidad mínima es un aspecto que varía entre poblaciones y que depende de la estructura genética de la población en cuestión.

Tabla 45: Probabilidades de identidad empíricas incrementando el número de SNPs.

N° SNPs	Parda de Montaña		Pirenaica I		Pirenaica II		Pirenaica I+II		Total	
	Locus	Incremento PI	Locus	Incremento PI	Locus	Incremento PI	Locus	Incremento PI	Locus	Incremento PI
1	SNP_C24	0,3362	SNP_C11	0,3469	SNP_C8	0,3441	SNP_C8	0,3571	SNP_C29	0,3563
2	SNP_C17	0,1104131	SNP_C18	0,1333	SNP_C24	0,12209738	SNP_C29	0,13607	SNP_C24	0,133821
3	SNP_C4	4,00E-02	SNP_C7	4,64E-02	SNP_C29	5,11E-02	SNP_C23	4,88E-02	SNP_C17	5,06E-02
4	SNP_C1	1,48E-02	SNP_C29	2,03E-02	SNP_C17	0,02429467	SNP_C24	2,09E-02	SNP_C8	1,88E-02
5	SNP_C23	4,76E-03	SNP_C1	1,55E-02	SNP_C23	7,31E-03	SNP_C17	9,04E-03	SNP_C13	7,51E-03
6	SNP_C13	0,00194	SNP_C8	7,73E-03	SNP_C13	3,21E-03	SNP_C4	3,81E-03	SNP_C23	2,61E-03
7	SNP_C7	1,16E-03	SNP_C23	2,90E-03	SNP_C4	1,60E-03	SNP_C13	1,55E-03	SNP_C4	8,58E-04
8	SNP_C18	5,82E-04	SNP_C21	1,93E-03	SNP_C21	1,07E-03	SNP_C1	8,84E-04	SNP_C7	4,65E-04
9	SNP_C2	3,88E-04	SNP_C17	9,66E-04	SNP_C1	8,02E-04	SNP_C21	4,42E-04	SNP_C18	3,25E-04
10	SNP_C20	0	SNP_C12	9,66E-04	SNP_C18	8,02E-04	SNP_C18	4,49E-04	SNP_C1	2,16E-04
11	SNP_C8		SNP_C19	9,66E-04	SNP_C6	5,35E-04	SNP_C19	4,49E-04	SNP_C21	1,80E-04
12	SNP_C29		SNP_C4	9,66E-04	SNP_C19	5,35E-04	SNP_C7	4,49E-04	SNP_C6	1,08E-04
13	SNP_C6		SNP_C13	9,66E-04	SNP_C12	5,35E-04	SNP_C11	4,49E-04	SNP_C12	1,08E-04
14	SNP_C12		SNP_C6	9,66E-04	SNP_C2	5,35E-04	SNP_C12	3,37E-04	SNP_C2	1,08E-04
15	SNP_C21		SNP_C2	9,66E-04	SNP_C11	5,35E-04	SNP_C6	3,37E-04	SNP_C19	1,08E-04
16	SNP_C10		SNP_C24	9,66E-04	SNP_C7	5,35E-04	SNP_C2	3,37E-04	SNP_C20	1,08E-04
17	SNP_C5		SNP_C20	9,66E-04	SNP_C10	5,35E-04	SNP_C20	3,37E-04	SNP_C11	1,08E-04
18	SNP_C19		SNP_C10	9,66E-04	SNP_C20	5,35E-04	SNP_C10	3,37E-04	SNP_C10	1,08E-04
19	SNP_C11		SNP_C5	9,66E-04	SNP_C5	5,35E-04	SNP_C5	3,37E-04	SNP_C5	1,08E-04

Debido a las diferencias que se muestran entre los valores de las probabilidades de identidad teóricas y empíricas se ha calculado la relación entre dichos valores (Tabla 46). Se ha visto, en general, que el valor teórico es menor que el empírico, lo que muestra que el valor teórico sobrevalora la probabilidad de identidad del marcador. En los casos donde el valor teórico es mayor que el empírico la relación es >1 . Esta situación aparece fundamentalmente en los marcadores más polimórficos que son los más adecuados para análisis de identificación y trazabilidad, estimando mejor la probabilidad de identidad.

5.2.2.4. Estimación de parámetros genéticos en la descendencia a partir de parámetros genéticos de los progenitores utilizando SNPs

De la misma forma que se ha hecho para los marcadores microsatélites se ha realizado la estimación de parámetros genéticos de los descendientes en función de parámetros de los progenitores utilizando el panel de SNPs.

Se llevó a cabo el análisis de correlación de cada uno de los parámetros calculados en los descendientes con los calculados en los progenitores. En la Tabla 47 se pueden observar los coeficientes de correlación de Pearson obtenidos entre las diferentes variables. Se puede apreciar que existe correlación estadísticamente significativa entre las variables de los progenitores y los descendientes. Los coeficientes muestran una correlación positiva entre las probabilidades de identidad de progenitores y descendientes, y negativa entre las probabilidades de identidad de los descendientes con respecto a las heterocigosidades de sus progenitores. Lo mismo ocurre con las heterocigosidades, que muestran correlaciones positivas entre progenitores y descendientes, y negativas entre las heterocigosidades de los descendientes respecto a las probabilidades de identidad de los progenitores (Figura 28).

Tabla 46: Relación entre los valores de probabilidades de identidad teóricos y empíricos.

Parda de Montaña				Pirenaica I				Pirenaica II				Pirenaica I+II				Gobal			
Loci	Pit	Ple	Pit/Ple	Loci	Pit	Ple	Pit/Ple	Loci	Pit	Ple	Pit/Ple	Loci	Pit	Ple	Pit/Ple	Loci	Pit	Ple	Pit/Ple
SNP_C7	0,375	0,388	0,9665	SNP_C7	0,375	0,367	1,0218	SNP_C24	0,375	0,3576	1,0487	SNP_C8	0,3751	0,3571	1,0504	SNP_C13	0,3757	0,3751	1,0016
SNP_C18	0,3754	0,4019	0,9341	SNP_C8	0,3752	0,38	0,9874	SNP_C29	0,3753	0,3647	1,0291	SNP_C29	0,376	0,3713	1,0127	SNP_C24	0,3764	0,3641	1,0338
SNP_C24	0,3758	0,3362	1,1178	SNP_C1	0,3762	0,372	1,0113	SNP_C8	0,3757	0,3441	1,0918	SNP_C24	0,3769	0,39	0,9664	SNP_C29	0,3777	0,3563	1,0601
SNP_C4	0,3766	0,3644	1,0335	SNP_C11	0,377	0,346	1,0896	SNP_C13	0,3774	0,4034	0,9355	SNP_C13	0,3829	0,4098	0,9344	SNP_C8	0,3792	0,3719	1,0196
SNP_C23	0,3777	0,3786	0,9976	SNP_C29	0,3783	0,371	1,0197	SNP_C4	0,3788	0,4175	0,9073	SNP_C17	0,3843	0,3956	0,9714	SNP_C4	0,3809	0,3886	0,9802
SNP_C17	0,3779	0,3394	1,1134	SNP_C18	0,3785	0,366	1,0342	SNP_C17	0,3799	0,3785	1,0037	SNP_C23	0,3856	0,3824	1,0084	SNP_C17	0,3812	0,3689	1,0333
SNP_C13	0,3786	0,3797	0,9971	SNP_C19	0,3818	0,441	0,8658	SNP_C23	0,3819	0,3849	0,9922	SNP_C4	0,3858	0,4079	0,9458	SNP_C23	0,3815	0,3797	1,0047
SNP_C1	0,381	0,3786	1,0063	SNP_C21	0,385	0,424	0,9080	SNP_C21	0,4012	0,4384	0,9151	SNP_C21	0,3941	0,4342	0,9076	SNP_C18	0,3844	0,4	0,9610
SNP_C2	0,3959	0,4137	0,9570	SNP_C24	0,3926	0,507	0,7744	SNP_C1	0,4212	0,4867	0,8654	SNP_C1	0,3961	0,4292	0,9229	SNP_C1	0,3884	0,4067	0,9550
SNP_C8	0,3961	0,4262	0,9294	SNP_C12	0,3932	0,431	0,9123	SNP_C18	0,426	0,4963	0,8584	SNP_C18	0,4025	0,439	0,9169	SNP_C7	0,39	0,3887	1,0033
SNP_C6	0,3977	0,4415	0,9008	SNP_C23	0,3952	0,385	1,0265	SNP_C6	0,4421	0,5137	0,8606	SNP_C19	0,4191	0,4762	0,8801	SNP_C21	0,4066	0,4589	0,8860
SNP_C20	0,3999	0,4232	0,9449	SNP_C17	0,3961	0,43	0,9212	SNP_C19	0,4703	0,5621	0,8367	SNP_C7	0,4336	0,4882	0,8882	SNP_C6	0,4162	0,4747	0,8768
SNP_C12	0,4084	0,4773	0,8556	SNP_C13	0,4097	0,454	0,9024	SNP_C12	0,4803	0,5704	0,8420	SNP_C11	0,4357	0,4946	0,8809	SNP_C12	0,4232	0,4904	0,8630
SNP_C29	0,4119	0,4415	0,9330	SNP_C4	0,4176	0,453	0,9219	SNP_C2	0,4939	0,6009	0,8219	SNP_C6	0,436	0,5074	0,8593	SNP_C2	0,4281	0,4904	0,8730
SNP_C21	0,4329	0,5085	0,8513	SNP_C6	0,4293	0,491	0,8743	SNP_C11	0,5366	0,6546	0,8197	SNP_C12	0,4376	0,5063	0,8643	SNP_C19	0,4666	0,5551	0,8406
SNP_C10	0,4562	0,5405	0,8440	SNP_C2	0,4321	0,507	0,8523	SNP_C7	0,5955	0,7184	0,8289	SNP_C2	0,4694	0,5673	0,8274	SNP_C20	0,4712	0,5587	0,8434
SNP_C5	0,556	0,6722	0,8271	SNP_C20	0,4541	0,527	0,8617	SNP_C10	0,767	0,8634	0,8883	SNP_C20	0,6065	0,7291	0,8318	SNP_C11	0,5094	0,6142	0,8294
SNP_C19	0,6158	0,735	0,8378	SNP_C10	0,6251	0,741	0,8436	SNP_C20	0,7688	0,8647	0,8891	SNP_C10	0,7133	0,8232	0,8665	SNP_C10	0,5593	0,6791	0,8236
SNP_C11	0,7593	0,8553	0,8878	SNP_C5	0,8526	0,916	0,9308	SNP_C5	0,9217	0,9592	0,9609	SNP_C5	0,8983	0,946	0,9496	SNP_C5	0,7071	0,818	0,8644

Pit: Probabilidad de identidad teórica

Ple: Probabilidad de identidad empírica

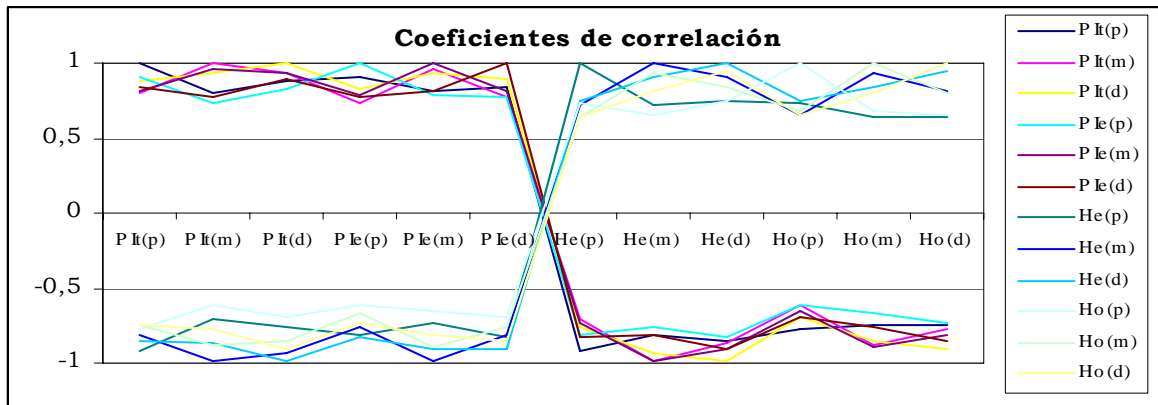


Figura 28: Representación de los coeficientes de correlación de Pearson entre las diferentes variables.

Debido a la existencia de correlación tan elevada de los diferentes parámetros entre progenitores y descendientes, se llevaron a cabo los análisis de regresión para modelizar la relación existente entre los diferentes parámetros.

5.2.2.4.1. Análisis de regresión para las probabilidades de identidad de los descendientes

Debido a la clara existencia de correlación entre los parámetros de los progenitores y descendientes se llevó a cabo el análisis de regresión para intentar explicar la variación de unos en función de otros. Utilizando el método Stepwise se determinaron cuales eran los parámetros con mejor capacidad de predicción por grupos poblacionales (Tabla 48). Se pudo observar que la mayoría de los modelos obtenidos involucran al menos dos variables, excepto en el grupo total, cuando se consideran todos los individuos como una misma población.

Tabla 47: Coeficientes de correlación de Pearson entre parámetros de progenitores y descendientes para SNPs (** $p < 0,01$, * $p < 0,05$).

		PIt(p)	PIt(m)	PIt(d)	PIe(p)	PIe(m)	PIe(d)	He(p)	He(m)	He(d)	Ho(p)	Ho(m)	Ho(d)
SNPs	PIt(p)	1	0,797(**)	0,878(**)	0,913(**)	0,813(**)	0,835(**)	-0,918(**)	-0,812(**)	-0,860(**)	-0,776(**)	-0,745(**)	-0,748(**)
	PIt(m)	0,797(**)	1	0,927(**)	0,740(**)	0,964(**)	0,774(**)	-0,702(**)	-0,981(**)	-0,871(**)	-0,615(**)	-0,880(**)	-0,768(**)
	PIt(d)	0,878(**)	0,927(**)	1	0,830(**)	0,931(**)	0,897(**)	-0,763(**)	-0,937(**)	-0,982(**)	-0,697(**)	-0,858(**)	-0,909(**)
	PIe(p)	0,913(**)	0,740(**)	0,830(**)	1	0,783(**)	0,773(**)	-0,818(**)	-0,764(**)	-0,827(**)	-0,608(**)	-0,672(**)	-0,731(**)
	PIe(m)	0,813(**)	0,964(**)	0,931(**)	0,783(**)	1	0,817(**)	-0,727(**)	-0,991(**)	-0,905(**)	-0,652(**)	-0,887(**)	-0,813(**)
	PIe(d)	0,835(**)	0,774(**)	0,897(**)	0,773(**)	0,817(**)	1	-0,831(**)	-0,812(**)	-0,912(**)	-0,695(**)	-0,762(**)	-0,849(**)
	He(p)	-0,918(**)	-0,702(**)	-0,763(**)	-0,818(**)	-0,727(**)	-0,831(**)	1	0,718(**)	0,744(**)	0,736(**)	0,644(**)	0,636(**)
	He(m)	-0,812(**)	-0,981(**)	-0,937(**)	-0,764(**)	-0,991(**)	-0,812(**)	0,718(**)	1	0,904(**)	0,659(**)	0,930(**)	0,816(**)
	He(d)	-0,860(**)	-0,871(**)	-0,982(**)	-0,827(**)	-0,905(**)	-0,912(**)	0,744(**)	0,904(**)	1	0,741(**)	0,846(**)	0,949(**)
	Ho(p)	-0,776(**)	-0,615(**)	-0,697(**)	-0,608(**)	-0,652(**)	-0,695(**)	0,736(**)	0,659(**)	0,741(**)	1	0,675(**)	0,655(**)
	Ho(m)	-0,745(**)	-0,880(**)	-0,858(**)	-0,672(**)	-0,887(**)	-0,762(**)	0,644(**)	0,930(**)	0,846(**)	0,675(**)	1	0,796(**)
	Ho(d)	-0,748(**)	-0,768(**)	-0,909(**)	-0,731(**)	-0,813(**)	-0,849(**)	0,636(**)	0,816(**)	0,949(**)	0,655(**)	0,796(**)	1

PIt: Probabilidades de identidad teóricas

PIe: Probabilidades de identidad empíricas

He: Heterocigosidades esperadas

Ho: Heterocigosidades observadas

(p): padres

(m): madres

(d): descendientes

Tabla 48: Variables incluidas en los modelos de regresión usando el método Stepwise.

Marcador	Grupo	Modelo	Variables incluidas
SNPs	Total	1	PIt(p)
	Parda de Montaña	1	PIt(p)
		2	He(m)
	Pirenaica I	1	PIt(p)
		2	PIt(m)
	Pirenaica I+II	1	PIt(m)
		2	PIt(p)

El ajuste de estos modelos a los datos lo podemos verificar mediante el cálculo de los coeficientes de determinación (Tabla 49). Todos los modelos muestran un ajuste elevado; considerando el mejor modelo por población observamos en todos los casos valores superiores al 97%. Nótese que en los grupos donde se obtiene más de un modelo, el ajuste no aumenta considerablemente, cuando se ven involucradas un mayor número de variables predictoras. En nuestro caso siempre trabajamos con los modelos de mayor R^2 , que más se ajusten sin tener en cuenta el número de variables que involucre.

Tabla 49: Coeficientes de determinación para cada uno de los modelos de regresión.

Marcador	Grupo	Modelo	R² ajustado
SNPs	Total	1	0,978621496
	Parda de Montaña	1	0,966150663
		2	0,97868736
	Pirenaica I	1	0,981041399
		2	0,994325923
	Pirenaica I+II	1	0,980476864
		2	0,985516879

En la Tabla 50 podemos encontrar los coeficientes de las rectas de regresión pasando por el origen de coordenadas, que permiten estimar la variación de la probabilidad de identidad de los descendientes a partir de las diferentes variables con mejor capacidad de predicción obtenidas en cada modelo.

Tabla 50: Coeficientes de las rectas de regresión pasando por el origen de coordenadas.

Marcador	Grupo	Modelo	Variables	Coeficientes estandarizados	
				B	Std. Error
SNPs	Total	1	PIt(p)	1,104547472	0,037431682
	Parda de Montaña	1	PIt(p)	0,907141556	0,038918032
		2	PIt(p)	0,647621492	0,082253874
			He(m)	0,277711996	0,081580879
	Pirenaica I	1	PIt(p)	0,930344286	0,029655525
		2	PIt(m)	0,564639193	0,085964058
			PIt(p)	0,472720826	0,071535337
	Pirenaica I+II	1	PIt(m)	1,159272335	0,03750916
		2	PIt(m)	0,741989769	0,158161617
			PIt(p)	0,368400735	0,136689977

De la tabla anterior podemos definir los modelos en notación matemática para los diferentes grupos que permiten predecir el comportamiento de las probabilidades de identidad en los descendientes (PIest(d)).

Población total (Parda de Montaña y Pirenaica I+II):

$$PIest(d) = 1,105 * PI t(p)$$

Pardas de Montaña:

$$1) PIest(d) = 0,907 * PI t(p)$$

$$2) PIest(d) = 0,648 * PI t(p) + 0,278 * He(m)$$

Pirenaica I:

$$1) PIest(d) = 0,93 * PI t(p)$$

$$2) PIest(d) = 0,565 * PI t(m) + 0,473 * PI t(p)$$

Pirenaica I + II:

$$1) PIest(d) = 1,159 * PI t(m)$$

$$2) PIest(d) = 0,742 * PI t(m) + 0,368 * PI t(p)$$

Con estas fórmulas matemáticas sería posible realizar la predicción del comportamiento de las Probabilidades de identidad en los descendientes a

partir de datos genotípicos de los progenitores con una elevada fiabilidad, lo cual está respaldado por los altos R^2 .

5.2.2.4.1.1. Comparación entre las PI teóricas, empíricas y estimadas en los descendientes

Los valores estimados de las PI de los descendientes utilizando los modelos matemáticos anteriores se han comparado con los valores teóricos y empíricos obtenidos. La comparación se ha llevado a cabo mediante la prueba t de Student para muestras pareadas para ver si existen diferencias entre ambos grupos de valores. Los resultados se pueden observar en la Tabla 51 y gráficamente en la Figura 29.

En general se puede apreciar que en la mayoría de los grupos existen diferencias estadísticamente significativas entre los valores de PI estimadas por regresión y los valores teóricos. Únicamente no existen diferencias significativas en la población de Parda de Montaña que es el grupo más grande y que incluye un mayor número de progenitores machos.

En general el valor teórico es menor que el empírico y el estimado por regresión, evidenciándose una vez más que el valor teórico sobrevalora la capacidad del marcador, ya que toma principalmente en consideración la frecuencia alélica más que la distribución de los alelos en la población. Al ser los SNPs marcadores bialélicos con un bajo contenido de información polimórfica (PIC) es muy importante considerar la distribución de los alelos en la población en la estimación de cualquier parámetro.

Tabla 51: Comparación de las Probabilidades de identidad empírica [PIe(d)], teórica [PIt(d)] y estimada por regresión [PIest(d)] de los descendientes por subpoblaciones.

Parda de Montaña				Pirenaica I				Pirenaica I+II				Poblaciones combinadas			
Loci	PIe(d)	PIt(d)	PIest(d)	Loci	PIe(d)	PIt(d)	PIest(d)	Loci	PIe(d)	PIt(d)	PIest(d)	Loci	PIe(d)	PIt(d)	PIest(d)
SNP_C7	0,388	0,375	0,3817	SNP_C7	0,367	0,375	0,3983	SNP_C8	0,3571	0,3751	0,4179	SNP_C13	0,3751	0,3757	0,4158
SNP_C18	0,4019	0,3754	0,3847	SNP_C8	0,38	0,3752	0,3931	SNP_C29	0,3713	0,376	0,4258	SNP_C24	0,3641	0,3764	0,4502
SNP_C24	0,3362	0,3758	0,3888	SNP_C1	0,372	0,3762	0,4035	SNP_C24	0,39	0,3769	0,4994	SNP_C29	0,3563	0,3777	0,4280
SNP_C4	0,3644	0,3766	0,3816	SNP_C11	0,346	0,377	0,4127	SNP_C13	0,4098	0,3829	0,4394	SNP_C8	0,3719	0,3792	0,4280
SNP_C23	0,3786	0,3777	0,3805	SNP_C29	0,371	0,3783	0,4069	SNP_C17	0,3956	0,3843	0,4497	SNP_C4	0,3886	0,3809	0,4174
SNP_C17	0,3394	0,3779	0,3824	SNP_C18	0,366	0,3785	0,3965	SNP_C23	0,3824	0,3856	0,4390	SNP_C17	0,3689	0,3812	0,4202
SNP_C13	0,3797	0,3786	0,3815	SNP_C19	0,441	0,3818	0,4518	SNP_C4	0,4079	0,3858	0,4585	SNP_C23	0,3797	0,3815	0,4147
SNP_C1	0,3786	0,381	0,3973	SNP_C21	0,424	0,385	0,3893	SNP_C21	0,4342	0,3941	0,4168	SNP_C18	0,4	0,3844	0,4188
SNP_C2	0,4137	0,3959	0,3647	SNP_C24	0,507	0,3926	0,4965	SNP_C1	0,4292	0,3961	0,4293	SNP_C1	0,4067	0,3884	0,4261
SNP_C8	0,4262	0,3961	0,3891	SNP_C12	0,431	0,3932	0,3943	SNP_C18	0,439	0,4025	0,4236	SNP_C7	0,3887	0,39	0,4147
SNP_C6	0,4415	0,3977	0,4053	SNP_C23	0,385	0,3952	0,4071	SNP_C19	0,4762	0,4191	0,4846	SNP_C21	0,4589	0,4066	0,4502
SNP_C20	0,4232	0,3999	0,3635	SNP_C17	0,43	0,3961	0,4246	SNP_C7	0,4882	0,4336	0,4499	SNP_C6	0,4747	0,4162	0,4921
SNP_C12	0,4773	0,4084	0,3784	SNP_C13	0,454	0,4097	0,4161	SNP_C11	0,4946	0,4357	0,4507	SNP_C12	0,4904	0,4232	0,4261
SNP_C29	0,4415	0,4119	0,4205	SNP_C4	0,453	0,4176	0,4092	SNP_C6	0,5074	0,436	0,4560	SNP_C2	0,4904	0,4281	0,4243
SNP_C21	0,5085	0,4329	0,4320	SNP_C6	0,491	0,4293	0,4423	SNP_C12	0,5063	0,4376	0,4234	SNP_C19	0,5551	0,4666	0,6632
SNP_C10	0,5405	0,4562	0,3975	SNP_C2	0,507	0,4321	0,4451	SNP_C2	0,5673	0,4694	0,4685	SNP_C20	0,5587	0,4712	0,4318
SNP_C5	0,6722	0,556	0,5144	SNP_C20	0,527	0,4541	0,5596	SNP_C20	0,7291	0,6065	0,6214	SNP_C11	0,6142	0,5094	0,5564
SNP_C19	0,735	0,6158	0,5871	SNP_C10	0,741	0,6251	0,7584	SNP_C10	0,8232	0,7133	0,7815	SNP_C10	0,6791	0,5593	0,5194
SNP_C11	0,8553	0,7593	0,4411	SNP_C5	0,916	0,8526	0,9105	SNP_C5	0,946	0,8983	1,0000	SNP_C5	0,818	0,7071	0,7734

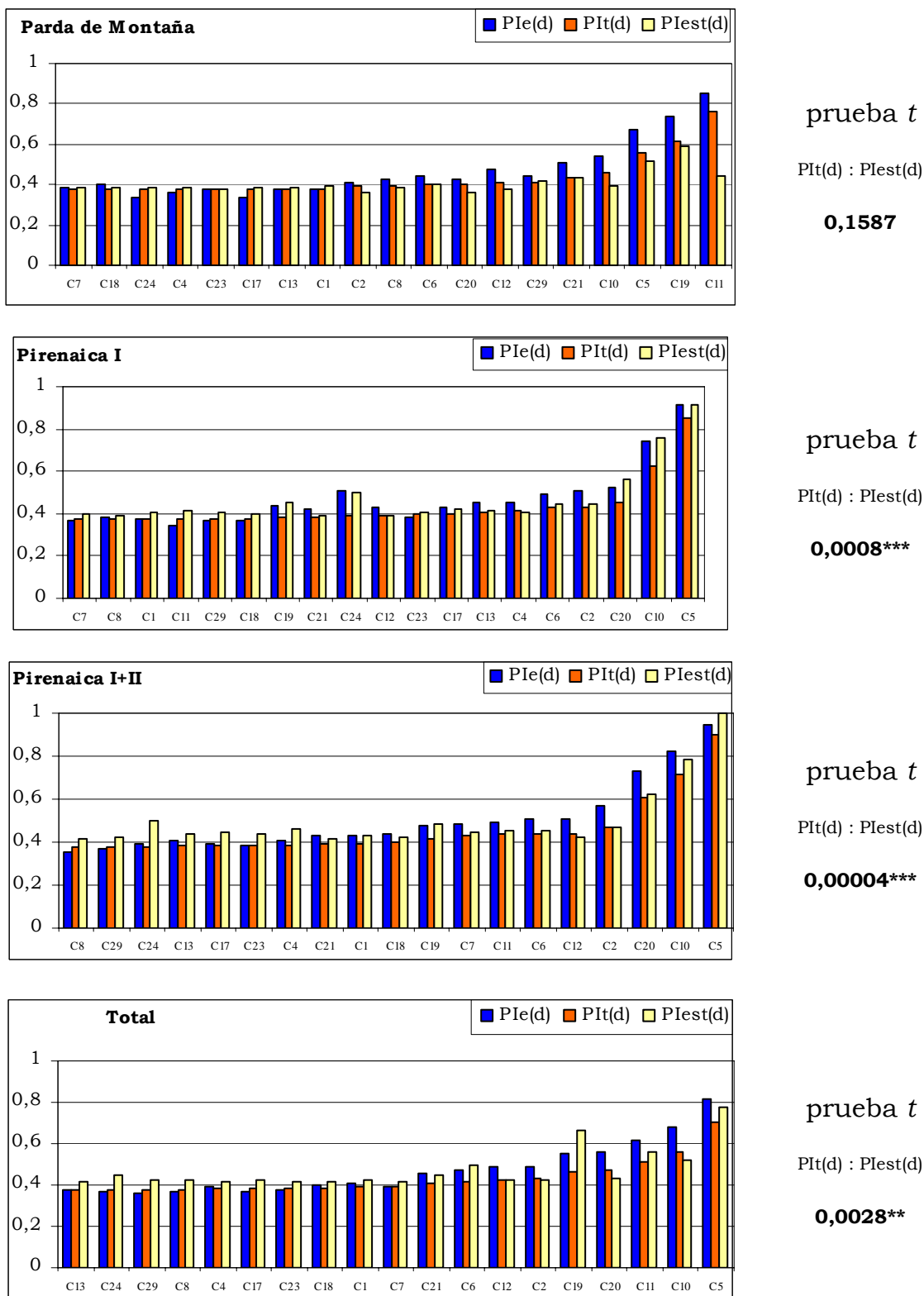


Figura 29: Representación gráfica de las PI empíricas, teóricas y estimadas de los descendientes con SNPs. Valores de las pruebas “*t*” de student para muestras pareadas para las PIt(d) y PIest(d). Valores de las pruebas *t* por muestras pareadas para PI teóricos y estimados.

5.2.2.4.2. Análisis de regresión para las heterocigosidades de los descendientes

En la Tabla 47 se puede observar que existe correlación entre las heterocigosidades de los descendientes y los diferentes parámetros de los progenitores. Para estimar la variación de las heterocigosidades observadas en los descendientes en función de dichos parámetros, se llevó a cabo el análisis de regresión. En la Tabla 52 se encuentran las variables incluidas en los modelos de regresión obtenidos.

Tabla 52: Variables incluidas en los modelos de regresión usando el método Stepwise.

Marcador	Grupo	Modelo	Variables incluidas
SNPs	Total	1	He(m)
		2	PIt(m)
	Parda de Montaña	1	He(m)
		2	PIt(p)
		3	He(p)
	Pirenaica I	1	He(m)
	Pirenaica I+ II	1	He(m)

En todas las poblaciones se puede observar que a partir de las heterocigosidades esperadas de las madres se puede predecir el comportamiento de las heterocigosidades de los descendientes, aunque existen diferentes modelos que involucran otras variables.

En la Tabla 53 se muestran los valores de R^2 y puede observarse que los ajustes son superiores al 95%, siendo el modelo 2 de la población total el que muestra un mejor ajuste superior al 99%.

Tabla 53: Coeficientes de determinación para cada modelo.

Marcador	Grupo	Modelo	R^2 ajustado
SNPs	Total	1	0,97096658
		2	0,99155804
	Parda de Montaña	1	0,96282305
		2	0,97443846
		3	0,98021598
	Pirenaica I	1	0,96354384
	Pirenaica I+II	1	0,95764744

Para crear un modelo matemático que nos permita hacer la predicción de las heterocigosidades de los descendientes a partir de los parámetros de los progenitores se han calculado los coeficientes de las rectas de regresión pasando por el origen de coordenadas (Tabla 54). En general se observa que todos los modelos predictivos dependientes de las heterocigosidades esperadas de las madres se pueden representar como rectas pasando por el origen de coordenadas.

Tabla 54: Coeficientes de las rectas de regresión pasando por el origen de coordenadas.

Marcador	Grupo	Modelo	Variables	Coeficientes estandarizados	
				B	Std. Error
SNPs	Total	1	He(m)	0,85750905	0,033991268
		2	He(m)	1,353954884	0,076317525
			Plt(m)	-0,562265005	0,083905897
	Parda de Montaña	1	He(m)	0,891503206	0,040148459
		2	He(m)	1,140504843	0,088672202
			Plt(p)	-0,2708706	0,089403697
		3	He(m)	0,599111665	0,235005004
			He(p)	0,929906991	0,380760699
			Plt(p)	-0,676634261	0,183821578
	Pirenaica I	1	He(m)	0,9821951	0,043786343
	Pirenaica I+II	1	He(m)	0,877036193	0,042264221

Los modelos en notación matemática se muestran a continuación. Utilizando estas fórmulas podremos a partir de diferentes variables calculadas con datos de los progenitores predecir el valor de la heterocigosidad que mostrarían los descendientes de los diferentes grupos.

Población total (Parda de Montaña y Pirenaica I+II):

$$1) Hest(d) = 0,858 * He(m)$$

$$2) Hest(d) = 1,354 * He(m) - 0,562 * Plt(m)$$

Parda de Montaña:

$$1) Hest(d) = 0,89 * He(m)$$

$$2) Hest(d) = 1,14 * He(m) - 0,271 * Plt(m)$$

$$3) Hest(d) = 0,599 * He(m) + 0,93 * He(p) - 0,676634261 * Plt(p)$$

Pirenaica I:

$$Hest(d) = 0,982 * He(m)$$

Pirenaica I + II:

$$Hest(d) = 0,877 * He(m)$$

5.2.2.4.2.1. Comparación entre las heterocigosidades esperadas, observadas y estimadas en los descendientes

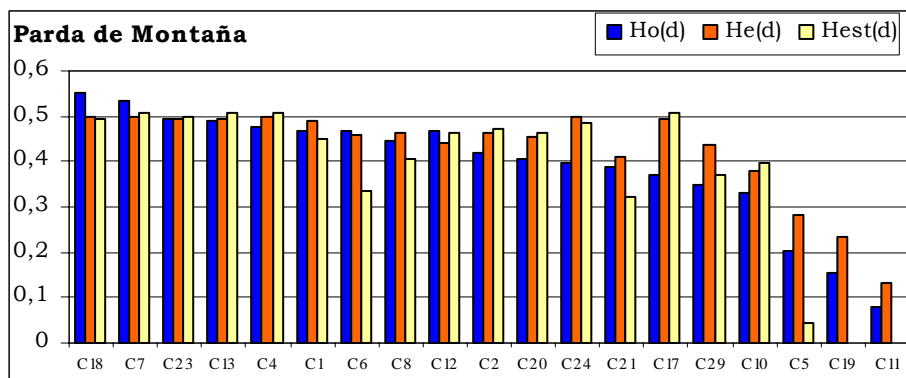
Utilizando las fórmulas matemáticas anteriores se han estimado las heterocigosidades en los descendientes y se han comparado con las heterocigosidades esperadas y observadas en las diferentes poblaciones. Se han realizado las pruebas *t* de Student para muestras pareadas tratando de verificar si existen diferencias entre los valores esperados y los estimados mediante regresión (Tabla 55 y Figura 30).

Para los SNPs la estimación por regresión de las heterocigosidades en los descendientes a partir de los progenitores es un método fiable lo cual está avalado por los elevados coeficientes de determinación que hemos obtenido. Existen diferencias significativas en la mayoría de los grupos entre los valores estimados por regresión y los valores teóricos, las mayores diferencias se encuentran en la población total debido a la mezcla de individuos de diferentes razas. Además estos resultados indican que en la teoría se sobrevalora la capacidad del marcador ya que en general los valores esperados son mayores con respecto a los observados o estimados por regresión.

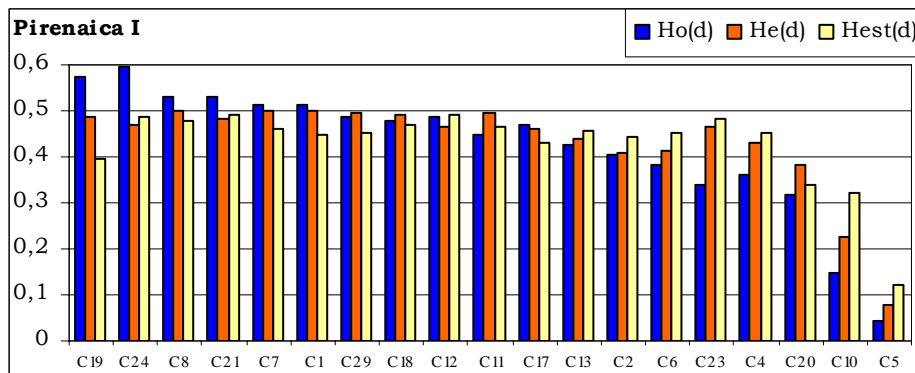
En análisis de trazabilidad en las poblaciones incluidas en el presente estudio, en caso de tener genotipados solamente los progenitores podríamos haciendo uso de estos modelos matemáticos dar un valor aproximado del comportamiento de la heterocigosidad observada en los descendientes.

Tabla 55: Comparación de las heterocigosidades observadas [Ho(d)], esperadas [He(d)] y estimadas por regresión [Hest(d)] en los descendientes.

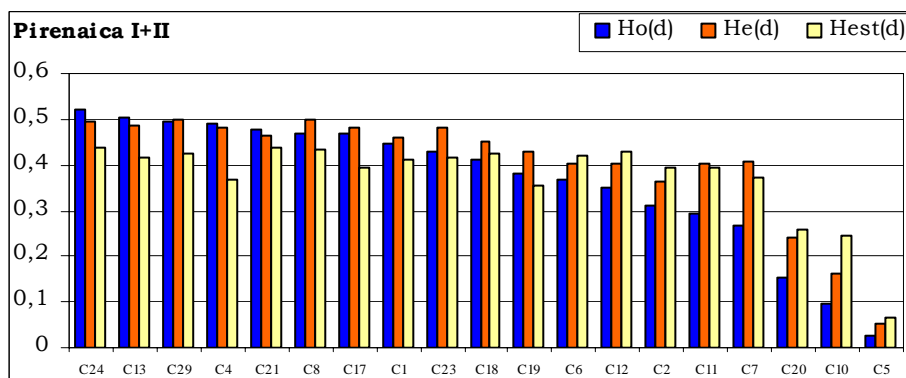
Parda de Montaña				Pirenaica I				Pirenaica I+II				Total			
Locus	Ho(d)	He(d)	Hest(d)	Locus	Ho(d)	He(d)	Hest(d)	Locus	Ho(d)	He(d)	Hest(d)	Locus	Ho(d)	He(d)	Hest(d)
SNP_C18	0,553	0,499	0,4959336	SNP_C19	0,575	0,487	0,3970226	SNP_C24	0,522	0,496	0,4378861	SNP_C13	0,498	0,499	0,460364
SNP_C7	0,534	0,5	0,5076983	SNP_C24	0,596	0,468	0,4884468	SNP_C13	0,504	0,485	0,4141194	SNP_C4	0,484	0,489	0,4216784
SNP_C23	0,495	0,495	0,4974539	SNP_C8	0,532	0,5	0,4781358	SNP_C29	0,493	0,498	0,4256958	SNP_C18	0,472	0,482	0,4519412
SNP_C13	0,49	0,493	0,5066927	SNP_C21	0,532	0,481	0,491	SNP_C4	0,49	0,48	0,3697432	SNP_C24	0,469	0,497	0,46625
SNP_C4	0,476	0,497	0,5075186	SNP_C7	0,511	0,5	0,462031	SNP_C21	0,479	0,465	0,4373599	SNP_C8	0,46	0,492	0,4338668
SNP_C1	0,466	0,488	0,448816	SNP_C1	0,511	0,498	0,4471046	SNP_C8	0,469	0,5	0,4346412	SNP_C23	0,456	0,487	0,4360204
SNP_C6	0,466	0,46	0,3362983	SNP_C29	0,489	0,494	0,45172	SNP_C17	0,469	0,482	0,3942992	SNP_C1	0,456	0,475	0,4412294
SNP_C8	0,447	0,462	0,4070322	SNP_C18	0,478	0,493	0,4705744	SNP_C1	0,448	0,462	0,4102606	SNP_C21	0,441	0,446	0,4537678
SNP_C12	0,466	0,443	0,4622665	SNP_C12	0,489	0,467	0,491	SNP_C23	0,428	0,48	0,414821	SNP_C29	0,433	0,495	0,465923
SNP_C2	0,418	0,463	0,470573	SNP_C11	0,447	0,496	0,4648788	SNP_C18	0,413	0,452	0,423591	SNP_C17	0,427	0,488	0,445999
SNP_C20	0,408	0,456	0,4625618	SNP_C17	0,468	0,462	0,4318836	SNP_C19	0,382	0,428	0,3529925	SNP_C6	0,407	0,432	0,4028836
SNP_C24	0,398	0,499	0,484729	SNP_C13	0,426	0,441	0,4549606	SNP_C6	0,366	0,405	0,4203461	SNP_C12	0,399	0,422	0,4332128
SNP_C21	0,388	0,409	0,3240011	SNP_C2	0,404	0,41	0,4447478	SNP_C12	0,352	0,403	0,4309578	SNP_C7	0,379	0,472	0,4243506
SNP_C17	0,369	0,494	0,5087892	SNP_C6	0,383	0,414	0,4525056	SNP_C2	0,313	0,365	0,3959655	SNP_C2	0,356	0,415	0,3727588
SNP_C29	0,35	0,438	0,3689438	SNP_C23	0,34	0,464	0,4825548	SNP_C11	0,292	0,405	0,3924575	SNP_C19	0,287	0,369	0,2474392
SNP_C10	0,33	0,38	0,3962558	SNP_C4	0,362	0,43	0,4539786	SNP_C7	0,269	0,408	0,3733389	SNP_C20	0,258	0,364	0,2437196
SNP_C5	0,204	0,281	0,0422546	SNP_C20	0,319	0,383	0,3394774	SNP_C20	0,152	0,24	0,2592412	SNP_C11	0,202	0,324	0,211472
SNP_C19	0,155	0,233	-0,1895	SNP_C10	0,149	0,226	0,3222924	SNP_C10	0,097	0,163	0,2458231	SNP_C10	0,194	0,279	0,2844002
SNP_C11	0,078	0,134	-0,1157	SNP_C5	0,043	0,078	0,1221608	SNP_C5	0,028	0,053	0,064898	SNP_C5	0,101	0,168	0,0853124



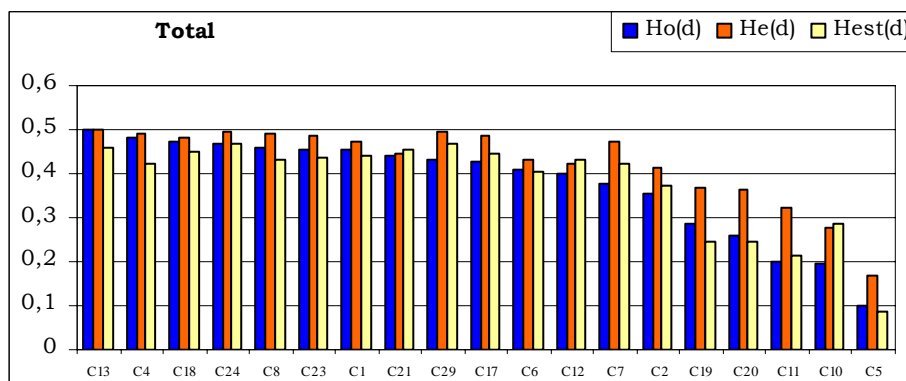
prueba *t*
 He(d): Ho_est(d)
0,0326*



prueba *t*
 He(d): Ho_est(d)
0,8146



prueba *t*
 He(d): Ho_est(d)
0,0177*



prueba *t*
 He(d): Ho_est(d)
0,0000***

Figura 30: Representación gráfica de las Heterocigosidades observadas, esperadas y estimadas de los descendientes. Valores de las pruebas “*t*” de Student para muestras pareadas para las He(d) y Hest(d).

5.3. Asignación de individuos a población como método complementario a trazabilidad: Microsatélites y SNPs

Utilizando los resultados obtenidos para los 9 microsatélites analizados en este trabajo fue posible realizar la asignación de individuos a raza mediante dos metodologías. Primeramente mediante el programa WHICHRUN v4.1 (Banks & Eichert 2000) que permite la asignación de individuos a la población más probable a partir de datos genotípicos multilocus; y mediante la construcción de un árbol filogenético a partir de las distancias de los alelos compartidos considerando cada animal de forma individual usando el programa MICROSAT (Minch 1995)

Usando el programa informático WHICHRUN v4.1 (Banks & Eichert 2000) se llevó a cabo la asignación de las 111 muestras de carne genotipadas a la población de pertenencia más probable, pudiéndose asignar correctamente 105 muestras a la raza correspondiente, lo cual supone un 94,58% (Figura 31).

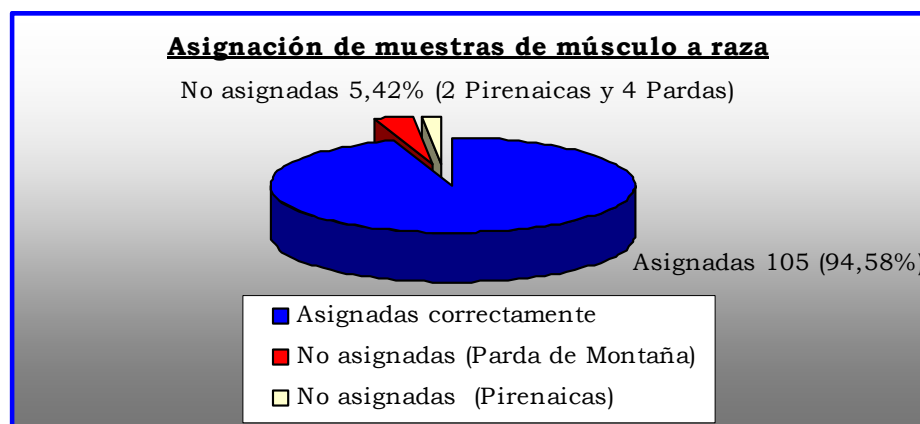


Figura 31: Gráfico que representa el porcentaje de asignación de muestras a raza.

Del 5.42% que no fue posible asignar correctamente 4 eran de raza Parda de Montaña y 2 de Pirenaica. Además se pudo determinar que tres (2,71%) de las muestras correspondían con casos en los que no se pudo establecer el vínculo correcto entre carne-animal, por lo que utilizando

métodos genéticos no solo podemos detectar fraudes cuando verificamos trazabilidad, sino también realizar la asignación de individuos a raza.

Empleando el mismo programa se realizó la asignación de muestras de carne provenientes de individuos de raza Pirenaica a su correspondiente subpoblación. Se obtuvo que de las 66 muestras no se asignaron correctamente a su respectiva subpoblación 8 de ellas (cuatro de cada subpoblación). Estos resultados pueden considerarse lógicos debido a que al estar trabajando con dos subpoblaciones de la misma raza las frecuencias alélicas son muy similares no existiendo diferencias estadísticamente significativas entre ellas, por lo que el error en la asignación puede ser mayor.

Utilizando resultados previamente obtenidos en nuestro laboratorio para las razas Mallorquina, Casta Navarra, Betizu, Toro de Lidia, Menorquina, Pirenaica, Monchina y Serrana de Teruel, se testó el poder de asignación de las 66 muestras de carne de Pirenaica a su raza y se obtuvo que 60 de ellas que representaban el 90,91% fueron asignadas correctamente; las restantes 6 muestras, es decir el 9,09% se asignó a la raza Serrana de Teruel la cual poseía la menor distancia D_A (Nei *et al.* 1983) al realizar la matriz de distancias (Tabla 56) entre las 8 razas analizadas (Zaragoza 2007).

Tabla 56 Matriz de distancias D_A Nei's (1983).

	Mallorquina	Casta Navarra	Betizu	Pirenaica	Toro de Lidia	Menorquina	Monchina
Casta Navarra	0,2484						
Betizu	0,3217	0,1785					
Pirenaica	0,2407	0,1286	0,1805				
Toro de Lidia	0,2787	0,1072	0,2248	0,1563			
Menorquina	0,2528	0,1804	0,2263	0,1601	0,2039		
Monchina	0,2709	0,1180	0,1686	0,1111	0,1545	0,1614	
Serrana de Teruel	0,2696	0,1180	0,1870	0,0956	0,1712	0,1550	0,0698

Otro método de estudio de la variabilidad genética empleado para realizar la asignación de individuos a raza utilizando los marcadores microsatélites ha sido la construcción del árbol filogenético con NJ a partir de las distancias de los alelos compartidos de todos los progenitores de ambas razas y las muestras de carne de sus descendientes (Figura 32). Del total de muestras analizadas para la asignación se puede apreciar que aunque existe armonía en la distribución de los individuos según su raza, es claro que hay un porcentaje no muy elevado donde la asignación no es correcta que comprende además de las muestras en las que no se pudo verificar la trazabilidad otras en las que la trazabilidad genética estaba garantizada. Esto puede ser debido a las semejanzas genéticas que pueden poseer estas muestras ya que las razas Parda de Montaña y Pirenaica están muy cercanas genéticamente, apareciendo en el mismo cluster en un dendograma construido a partir de la distancia D_A y la metodología NJ (Zaragoza 2007).

Estos resultados nos permiten concluir que los marcadores microstélites además de ser una herramienta muy útil en la identificación y verificación de la trazabilidad nos permite realizar la asignación de individuos a raza con una probabilidad dada. Por tanto en análisis de trazabilidad es una aplicación muy útil, ya que teniendo una muestra de carne problema si contamos con datos de frecuencias de las posibles poblaciones de donde pueda provenir dicha muestra problema se puede realizar la asignación de la muestra de carne a la posible población de origen.

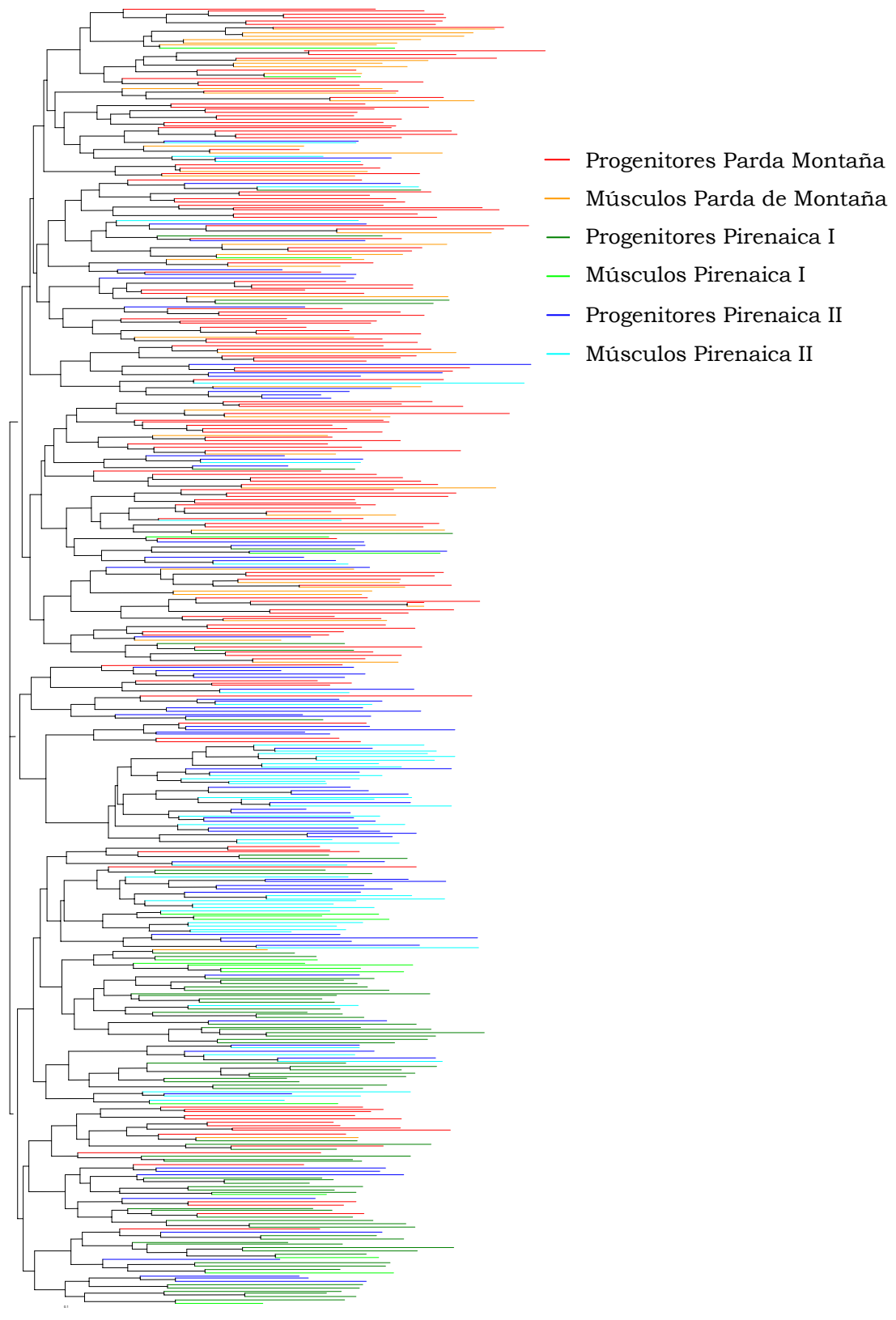


Figura 32: Dendrograma NJ construido a partir de la distancia de los alelos compartidos entre los progenitores y muestras de músculo de las explotaciones analizadas.

Los resultados genotípicos de los 19 SNPs analizados permitieron llevar a cabo, mediante la metodología de la aproximación probable, la asignación de individuos a raza. Se utilizó el programa WHICHRUN v4.1 (Banks & Eichert 2000) y se encontró que del total de 111 muestras de carne fue posible asignar correctamente a su raza de origen 89 muestras (80,2%). De las 22 que no fueron asignadas correctamente 10 pertenecían a la raza Parda de Montaña y 12 a Pirenaica (Figura 33).

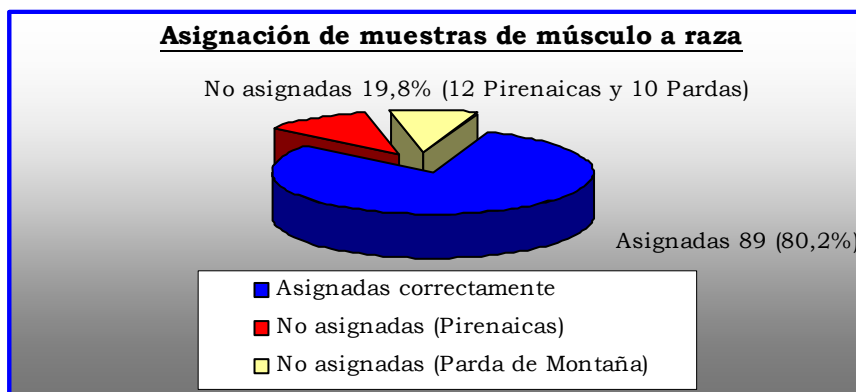


Figura 33: Gráfico que representa el porcentaje de asignación de muestras a raza.

Teniendo en cuenta que contábamos con muestras de carne de dos subpoblaciones de raza Pirenaica se testó mediante la misma metodología la asignación de las 66 muestras de carne a su subpoblación. En este caso se asignaron correctamente 52 muestras (78,8%), del resto que no fue posible asignar 7 pertenecían a la subpoblación I y 7 a la subpoblación II (15,6%).

En general se aprecia que la capacidad de asignación con este tipo de marcadores es baja lo que puede estar dado por su bajo polimorfismo. Ya que ha sido reportado previamente que el éxito de la asignación puede ser incrementado si se aumenta el número de marcadores genotipados esta puede ser una estrategia a tener en cuenta si se quiere tener un mayor porcentaje de asignación de individuos a raza.

5.4. Comparación de la aplicabilidad a los sistemas de trazabilidad de ambos tipos de marcadores moleculares: Microsatélites y SNPs

Para tratar de cuantificar cual de los dos métodos ofrece mejores resultados para ser utilizado como método de identificación y verificación de sistemas de trazabilidad, en este apartado realizaremos a modo de resumen una comparación de la efectividad de ambos tipos de marcadores.

Como se ha visto a lo largo del desarrollo del Proyecto, se han analizado dos tipos de marcadores moleculares: microsatélites y SNPs en dos razas bovinas del norte de Aragón; una población de Parda de Montaña y dos subpoblaciones de Pirenaica. Estas razas son de especial interés en la Comunidad de Aragón ya que en un futuro no muy lejano carnes frescas de estas razas serán amparadas por lo que va a denominarse Indicación Geográfica Protegida “Ternera Montañas de Aragón”. En este sentido los análisis de trazabilidad pueden ser extremadamente importantes.

5.4.1. Comparación de parámetros: PI y PE

Dos parámetros importantes a tener en cuenta a la hora de comparar ambos paneles de marcadores, son las probabilidades de identidad (PI) y las probabilidades de exclusión (PE) (Tabla 57).

Tabla 57: Comparación de las Probabilidades de identidad (PI) y las Probabilidades de exclusión (PE) a partir de microsatélites y SNPs.

		Microsatélites	SNPs
Parda de Montaña	PI	1,49E-09	8,83E-08
	PE	0,9987	0,9688
Pirenaica I	PI	1,34E-08	7,80E-08
	PE	0,9968	0,9690
Pirenaica II	PI	7,04E-08	6,76E-07
	PE	0,9941	0,95196
Pirenaica I+II	PI	9,29E-09	2,01E-07
	PE	0,9974	0,96237
Total	PI	7,23E-10	8,87E-08
	PE	0,9991	0,9696

La probabilidad de identidad obtenida para el panel de 9 microsatélites ha oscilado entre 10^{-8} y 10^{-10} dependiendo de la población analizada. Para la población de Parda de Montaña la probabilidad de encontrar dos individuos que compartan los mismos genotipos ha sido de 1:671 millones; para Pirenaica I de 1:74,6 millones; para Pirenaica II, 1:14,2 millones y para la Pirenaica I+II de 1:108 millones. En la población total hemos observado que la probabilidad de encontrar dos individuos con iguales perfiles genéticos es de 1:1380 millones, por lo que trabajando con poblaciones de individuos no emparentados o de diferentes razas la probabilidad de encontrar individuos iguales es extremadamente baja.

Las diferencias observadas entre todas las poblaciones vienen dadas fundamentalmente por el grado de polimorfismo existente en cada una. Las pequeñas diferencias que se muestran entre las dos subpoblaciones de Pirenaica pueden estar determinadas por el número de reproductores existentes en cada una, siendo menor en el caso de la Pirenaica II donde se pueden apreciar valores ligeramente menores que los obtenidos para la Pirenaica I. De cualquier forma, se evidencia que los valores de probabilidades de identidad son muy bajos en todos los casos.

Para el panel de 19 SNPs los valores de PI han oscilado entre 10^{-7} y 10^{-8} . Para la población de Parda de Montaña la probabilidad ha sido 1:11,3 millones; para Pirenaica I, 1:12,8 millones, para Pirenaica II, 1:1,48 millones y para Pirenaica I+II de 1:5 millones. En la población total se ha visto que con este panel la probabilidad de encontrar individuos iguales es de 1:11,3 millones. Las diferencias entre las poblaciones analizadas depende del grado de polimorfismo de estas y del número de reproductores utilizados en cada una de las poblaciones.

El valor de PI obtenido para los SNPs es comparable a la probabilidad de $4,3 \times 10^{-8}$ (1:23 millones) obtenida al genotipar 20 loci en razas de carne americanas (Heaton *et al.* 2005). En estudios realizados en razas europeas utilizando un panel de 43 SNPs se obtuvieron valores de PI de $5,3 \times 10^{-11}$ (1:18867 millones) (Herraez *et al.* 2005).

Podemos concluir que para el panel de 9 microsatélites se obtienen mejores resultados en cuanto a la probabilidad de identidad con respecto a los 19 SNPs (Figura 34). Por lo que para obtener resultados semejantes en cuanto a probabilidades de identidad se precisaría un mayor número de SNPs en comparación con la cantidad de microsatélites necesarios.

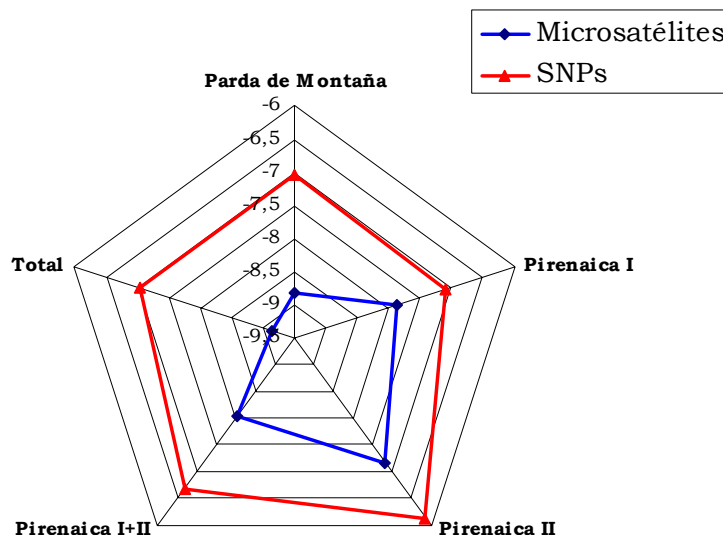


Figura 34: Logaritmos de las probabilidades de identidad con ambos paneles de marcadores moleculares por poblaciones.

A continuación vamos a comparar los valores de Probabilidad de Exclusión (PE), medida de la habilidad de un marcador de identificar la paternidad genética excluyendo todos los restantes candidatos, usando los paneles de 9 microsatélites y 19 SNPs. Según puede observarse en la Tabla 57, los microsatélites muestran mejores poderes de exclusión en todas las poblaciones, debido a que al ser marcadores multialélicos presentan mayor Contenido de información polimórfica que los marcadores SNPs que son marcadores bialélicos (Herraez *et al.* 2005). Una representación gráfica de dichos valores se muestra en la Figura 35.

Con el panel de 9 microsatélites se obtienen para todas las poblaciones analizadas probabilidades de exclusión superiores al 99%, siendo para la población total de 99,91%. Para el panel de 19 SNPs los valores obtenidos no

son tan elevados, oscilando en las diferentes poblaciones alrededor del 96%, el mayor valor se obtuvo al considerar la población total (96,96%). Los resultados obtenidos para los SNPs son más bajos que los obtenidos para los microsatélites y son comparables a los obtenidos en la raza Galloway donde con un panel de 43 SNPs se obtuvo una probabilidad de exclusión de 97,72% (Herraez *et al.* 2005).

Las diferencias en los valores de PE entre los dos paneles están influenciadas evidentemente por el número de alelos por locus de cada tipo de marcador; en este sentido los microsatélites al ser multialélicos con mayor PIC serán preferibles a los SNPs que son bialélicos (Herraez *et al.* 2005).

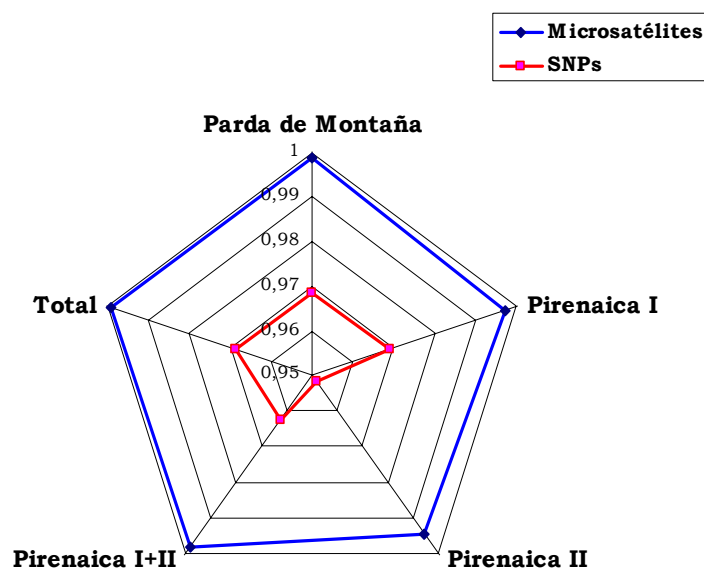


Figura 35: Probabilidades de exclusión obtenidas para las diferentes poblaciones y ambos paneles de marcadores.

5.4.2. Costes económicos

Otro aspecto a tener en cuenta para determinar la aplicabilidad de cada uno de los métodos para verificar la trazabilidad son los costes económicos que acarrear la identificación genética. Existe una gran

diversidad de metodologías disponibles para el genotipado de SNPs, se estima que la producción de un genotipo está en el orden de 0,01 \$ (Heaton *et al.* 2005), sin embargo dependiendo de la metodología utilizada para el genotipado los costes pueden variar.

En nuestro trabajo se realizó una reacción múltiplex PCR fluorescente con posterior electroforesis capilar para el genotipado de los 9 microsatélites. Para el genotipado de los 19 SNPs fue necesario la realización de dos reacciones de PCR estándares múltiplex y las posteriores reacciones de minisequenciación empleando el kit SNaPshot de Applied Biosystems y finalmente las correspondientes electroforesis capilares. Por lo que en las condiciones de nuestro trabajo el coste económico y en tiempo que conlleva el genotipado de los 19 SNPs es superior al coste que genera el genotipado de los 9 microsatélites.

La utilización de otro tipo de metodología y plataformas permite la automatización en el genotipado de SNPs, lo que sugiere un ahorro económico y de tiempo considerable, pero también acarrea un desembolso inicial importante en la adquisición de las plataformas.

Por otra parte la utilización de SNPs en trazabilidad permite establecer una firma digital de ADN, lo que sería una alternativa muy fiable en el etiquetado de los productos. Los SNPs al ser marcadores bialélicos, es posible utilizar un código binario de 0 y 1 y establecer una firma digital de ADN (Fries & Durstewitz 2001; Werner *et al.* 2002) que podría ser traducido a un código de barras y aparecer en el etiquetado final del producto.

En este caso las cuatro posibilidades serían codificadas de forma digital, de manera que para un locus en un individuo existen cuatro posibilidades diferentes:

heterocigótico → 11

homocigótico para la base 1 → 10

homocigótico para la base 2 → 01

falta de genotipo → 00

Por ejemplo un animal genotipado para el panel de 19 SNPs en un orden definido, tendrá asociada una firma digital como se muestra a continuación en la Figura 36.

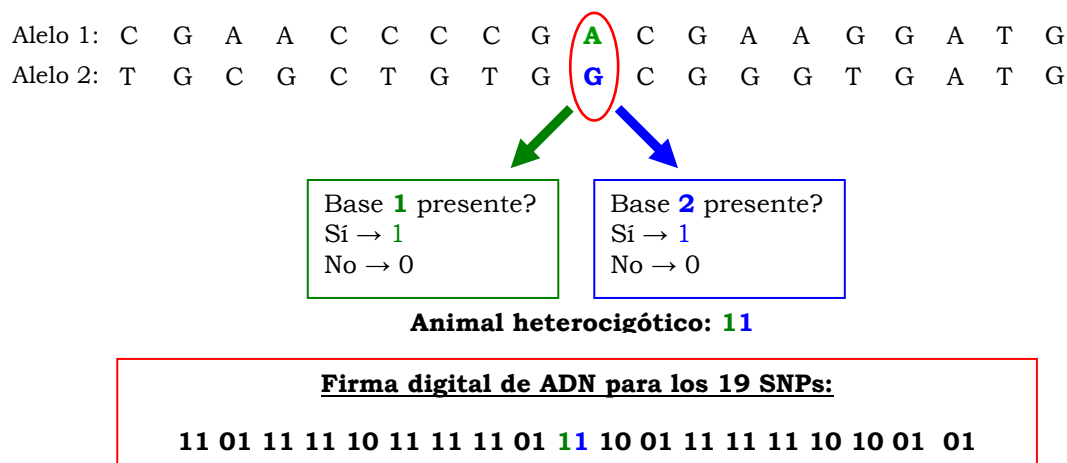


Figura 36: Firma digital para los 19 SNPs.

De esta manera todos los individuos pueden ser identificados mediante una firma digital de ADN en el caso que se utilice el panel de SNPs, lo que supondría una ventaja en la automatización del proceso, no siendo posible para el caso de microsatélites.

Con respecto al panel de SNPs, una ventaja que se puede considerar es la posibilidad de genotipar polimorfismos involucrados con caracteres de interés cuantitativos al mismo tiempo que se lleve a cabo la verificación de la trazabilidad.

5.4.3. Estimación de parámetros de los descendientes a partir de los progenitores

La capacidad de estimación de las probabilidades de identidad y las heterociguidades en los descendientes a partir de parámetros genéticos de los progenitores ha sido demostrada con ambos paneles de marcadores moleculares, avalado por los elevados coeficientes de determinación obtenidos en todos los casos. Al comparar los modelos matemáticos obtenidos por regresión para ambos tipos de marcadores podemos ver que

en general para los microsatélites es posible predecir el comportamiento de las variables de los descendientes en función de una variable de los progenitores. Para los SNPs la estimación generalmente está en función de más de una variable, siendo modelos más complejos. Estas diferencias pueden estar dadas por la naturaleza de ambos tipos de marcadores y por tanto de los diferentes PIC de cada uno.

Por otra parte la comparación de los valores teóricos/esperados y estimados por regresión para las dos variables de los descendientes a partir de los progenitores se llevó a cabo mediante la prueba paramétrica *t* de Student para muestras pareadas. El resumen de las significaciones obtenidas se muestra en la Tabla 58.

Tabla 58: Significación obtenida al comparar valores teóricos / esperados y estimados de las probabilidades de identidad y heterocigosidades de los descendientes con ambos paneles de marcadores en las diferentes poblaciones.

	Microsatélites				SNPs			
	Parda de Montaña	Pirenaica I	Pirenaica I+II	Total	Parda de Montaña	Pirenaica I	Pirenaica I+II	Total
PI			**	**		***	***	**
H				***	*		*	***

Se ha observado que para los microsatélites no existen diferencias estadísticamente significativas entre los valores teóricos y estimados por regresión en ambas variables para las poblaciones de Parda de Montaña y Pirenaica I. Sin embargo cuando se realizó la prueba para la población Pirenaica I+II y la población total, si se observan diferencias entre las dos variables.

Para los SNPs se han observado en la mayoría de los casos diferencias estadísticamente significativas entre los valores teóricos y los estimados por regresión, dadas por la naturaleza bialélica de este tipo de marcadores y el

bajo PIC que muestran, por lo que la capacidad de predicción es menor si se compara con los microsatélites.

5.4.4. Comparación de la asignación

Como método complementario para la trazabilidad y seguimiento de los animales y productos, se ha propuesto la posibilidad de asignar los individuos a una raza o población. La comparación del poder de asignación de individuos a raza con ambos paneles de marcadores se ha llevado a cabo utilizando la metodología de aproximación probable mediante el programa WHICHRUN v4.1 (Banks & Eichert 2000). En la Figura 37 se puede observar la capacidad de asignación usando los marcadores microsatélites y SNPs.

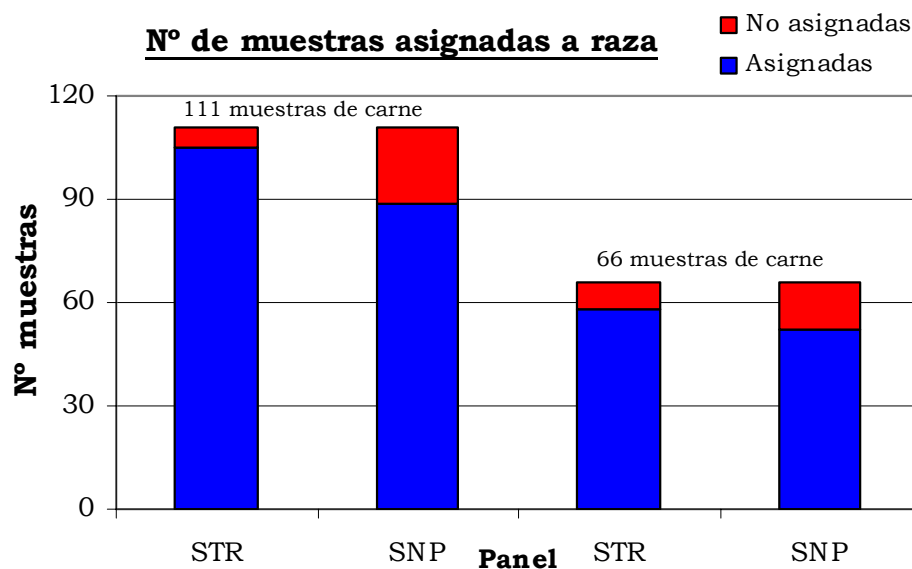


Figura 37: Asignación de muestras a raza con ambos paneles de marcadores moleculares.

La asignación de las 111 muestras de carne a su respectiva raza muestra mejores resultados para el panel de microsatélites, con el que se ha asignado correctamente el 94,58% de las muestras a su respectiva raza; con el panel de SNPs ha sido posible asignarse el 87,88%. Las 66 muestras de carne de individuos de raza Pirenaica han sido asignadas a su respectiva subpoblación con el panel de microsatélites en el 80,2% de los casos, mientras que para los SNPs la asignación ha sido del 78,8%.

En general se puede observar que el panel de microsatélites muestra mejor capacidad de asignación, lo cual puede estar dado por el grado de polimorfismo de este tipo de marcadores con respecto a los marcadores bialélicos SNPs. Ya que el número de loci analizados puede influir en el éxito de la asignación racial, es probable que el número de asignaciones se incremente si aumenta el número de loci analizados.

Por todo lo expuesto anteriormente ambos paneles son elegibles y eficaces para ser utilizados en estudios de identificación y trazabilidad.

5.4.5. Verificación de la correspondencia carne-sangre de animal de procedencia

Para verificar el sistema de trazabilidad convencional se compararon los perfiles genéticos de las 111 muestras de carne de matadero, con los perfiles obtenidos a partir de las muestras de sangre identificadas con el mismo número según el sistema convencional de trazabilidad. Ambos procedimientos de genotipado permitieron reconstituir los pares “pieza de carne” – “animal de procedencia” en 102 casos (91,89%). Las diferentes causas por las que no fue posible reconstituir los pares en las 9 muestras restantes (8,11%) se muestran a continuación (Tabla 59).

Según se aprecia en la Tabla 59 además fue detectado un error en la toma de muestras gracias al uso de este sistema genético. Este es el caso de la muestra 5775 que a pesar de ser verificable mediante el sistema de trazabilidad, se comprobó que existía una segunda muestra que compartía sus mismos perfiles genéticos para ambos paneles de marcadores. En este caso se pudo determinar que era igual a la muestra identificada como su hermana gemela, pero como es sabido dos individuos solamente comparten los mismos perfiles genéticos si son gemelos monocigóticos. Esto nos hizo dudar pues eran muestras de diferente sexo, de manera que ante la sospecha de tener duplicada la muestra de sangre se realizó la PCR-RFLP para determinar el sexo de la muestra siguiendo la metodología descrita por Osta (1994). Se encontró que las dos muestras de sangre recibidas como

gemelos de diferente sexo eran ambas procedentes de un individuo macho. Por esto al comparar el perfil de la muestra de carne con la base de datos dio como resultado dos individuos con iguales perfiles genéticos. Por un lado se podía verificar el sistema de trazabilidad, pero por otro se detectaba un error en la toma de la muestra.

Tabla 59: Casos donde no fue posible verificar el sistema de trazabilidad convencional.

ID carne	ID sangre	Verificación sistema trazabilidad convencional mediante marcadores moleculares
5252 (I)	4847 (I)	No correspondencia. Muestra carne 5252 igual a muestra de sangre 4872 para ambos paneles de marcadores (muestras cruzadas).
5254 (II)	4872 (II)	No correspondencia. Muestra carne 5254 igual a muestra de sangre 4847 para ambos paneles de marcadores (muestras cruzadas).
5256 (III)	4857 (III)	No correspondencia. Perfiles discordantes para los 9 microsatélites y para 9 SNPs. Origen desconocido.
5507 (IV)	4932 (IV)	No correspondencia. La muestra de carne es de origen desconocido, discordancias para 6 microsatélites y 9 SNPs. No coincide el perfil genético con ninguna de las muestras genotipadas y es incompatible con padre y madre. Sin embargo, la supuesta muestra correspondiente de sangre si es compatible con los progenitores asignados en la base de datos.
5632 (V)	5605 (V)	No correspondencia. Discordancias en 7 microsatélites y 10 SNPs, pero perfil genético coincide con la 5574 para ambos paneles de marcadores.
5767 (VI)	5195 (VI)	No correspondencia. Discordancias en los 9 microsatélites. Origen desconocido. Genotipado parcial.
5774 (VII)	5214 (VII)	No correspondencia. Muestra de carne 5774 igual a muestra de sangre 5227 para ambos paneles de marcadores (muestras cruzadas).
5775 (VIII)	5215 (VIII)	Existe correspondencia total entre la muestra de carne 5775 y su correspondiente animal de origen (muestra de sangre 5215), sin embargo sale además total correspondencia con la 5216 que es gemela con la 5215.
5781 (IX)	5227 (IX)	No correspondencia. Muestra de carne 5781 igual a muestra de sangre 5214 para ambos paneles de marcadores (muestras cruzadas).

En las condiciones de nuestro trabajo no se detectan diferencias entre la informatividad de los dos tipos de marcadores empleados; con ambos

tipos se resuelven los mismos casos y se puede llegar a las mismas conclusiones.

5.5 Estudio piloto de trazabilidad

La metodología genética descrita en esta memoria ha sido utilizada como método de verificación del sistema de trazabilidad convencional establecido en el Reglamento 1760/2000 de la Comunidad Europea y de aplicación rutinaria en la industria cárnica.

Para el establecimiento de este sistema genético se ha realizado desde el inicio un control exhaustivo de los progenitores y descendientes en dos explotaciones de razas Parda de Montaña y Pirenaica, en el período comprendido entre 2004 y 2005. Para esto se tomaron muestras de sangre de todos los progenitores y descendientes. Posteriormente se tomaron muestras de carne de los descendientes que fueron sacrificados en matadero. La totalidad de las muestras recibidas quedaron registradas conjuntamente con toda la información disponible en la base de datos diseñada en el Laboratorio de Genética (LAGENBIO). Esta base de datos ha permitido gestionar todos los datos de las muestras, incluir sus perfiles genéticos una vez analizadas con los dos paneles de marcadores moleculares microsatélites y SNPs. Mediante la aplicación web diseñada se han podido realizar distintas búsquedas por grupos, realizar análisis de filiación y establecer el vínculo entre la muestra de carne, la muestra de sangre del animal de procedencia y por tanto la explotación, mediante comparación de perfiles genéticos utilizando los dos procedimientos de genotipado. De esta forma se ha autenticando la trazabilidad en el 91,89% de los casos. El 8% restante en el que no fue posible autenticar el sistema de trazabilidad se atribuyó en su mayoría a errores en el etiquetado, aunque también se detectó la duplicación de una muestra.

Teniendo en cuenta que el presente trabajo se ha realizado en condiciones reales, se podría estimar que el porcentaje de error no detectado mediante el sistema convencional de trazabilidad en este caso específico ha

estado en el orden del 8%, por lo que con este sistema genético, los errores podrían minimizarse.

Por otra parte, con este sistema y mediante la comparación de perfiles genéticos entre progenitores y descendientes, uno de los objetivos del presente trabajo era establecer un sistema de trazabilidad basado en el genotipado de progenitores. Se ha intentado asignar las muestras de carne a alguno de los progenitores genotipados llevando a cabo un análisis de filiación, tratando de esta forma de hacer un seguimiento sin necesidad de genotipar los descendientes en vida. Fue posible asignar la muestra de carne a sus progenitores en el 8% de los casos. En el 20% de los casos se asignó el padre y dos o tres posibles madres, todas de la misma raza y explotación, debido fundamentalmente al elevado grado de relación entre los individuos, así como el bajo número de reproductores que conformaban las explotaciones incluidas en nuestro estudio. Mediante aproximaciones probables se han realizado las asignaciones de las muestras de carne a las explotaciones de origen basándonos en los perfiles genéticos de los progenitores, en un porcentaje elevado de las muestras de carne. Esta aplicación supone una importante herramienta en el ámbito comercial ya que puede permitir verificar el origen de la carne principalmente cuando se analizan carnes pertenecientes a marcas comerciales o con denominaciones de origen a nivel de país (Negrini *et al.* 2008). Los productos que contienen en su etiquetado una indicación geográfica dan una imagen positiva del producto (Kapferer 1992), lo que incrementa la transparencia y la confianza de los consumidores.

Desde un punto de vista práctico los dos paneles de marcadores moleculares utilizados representan una herramienta muy útil en el control de libros genealógicos que puede ser explotada por las diferentes asociaciones de ganaderos, así como en la certificación de la calidad de marcas o denominaciones de origen. Además, el sistema permite la defensa en casos de detectar conductas sospechosas tales como el entierro ilegal de animales que han muerto accidentalmente o la detección de drogas ilegales en la carne (Vazquez *et al.* 2004).

Este sistema usado conjuntamente con los procedimientos administrativos, constituiría un sistema más eficaz en el sentido de evitar fraudes o para mejorar la trazabilidad en ganado bovino. En este contexto la constitución de un banco de datos incluyendo el genotipo de cada individuo podría ser muy útil. Además el sistema puede ser apropiadamente adaptado a otras especies en las cuales el valor de los animales individuales sea alto, como cerdos, ovejas o caballos.

6. CONCLUSIONES

6. CONCLUSIONES

De acuerdo a los resultados obtenidos las principales conclusiones del presente trabajo son las siguientes:

1. Los dos procedimientos múltiplex de genotipado diseñados para ambos paneles de marcadores (9 microsatélites y 19 SNPs), permitieron determinar satisfactoriamente los perfiles genéticos del total de muestras utilizadas en el estudio de trazabilidad.
2. Los dos paneles de marcadores utilizados han permitido identificar el animal de origen de la muestra de carne en el 92% de los casos, detectándose errores en el sistema de trazabilidad convencional gracias al empleo de esta metodología.
3. La base de datos diseñada y el algoritmo asociado que permite realizar la gestión del sistema de trazabilidad podría ser una herramienta muy útil para los ganaderos, además de su fácil manejo y accesibilidad.
4. Las probabilidades de exclusión (PE) obtenidas con ambos paneles de microsatélites y SNPs (99,91 y 96,83% respectivamente) demuestran que ambos conjuntos de marcadores pueden ser utilizados en estudios de identificación y trazabilidad.
5. Las probabilidades de identidades (PI) calculadas para ambos paneles ($7,16E-10$ para microsatélites y $8,92E-08$ para SNPs) demuestran que ambos son aplicables a estudios de trazabilidad.
6. Se han obtenido modelos matemáticos que permiten estimar los diferentes parámetros genéticos de los descendientes en función de los parámetros genéticos de los progenitores
7. La asignación de muestras de carne a su población de origen basada en la información de los progenitores ha sido posible mediante métodos probabilísticos, no mediante la comparación directa de genotipos.
8. Se ha descartado la utilización de un SNP recomendado previamente para estudios de trazabilidad e identificación que mostraba

frecuencias alélicas próximas a 0,5 por su ubicación en una región repetida del genoma bovino.

9. Uno de los SNPs incluidos en el panel establecido en la presente memoria (gen Transferrina), ha mostrado una asociación significativa con la cantidad de grasa en la leche en ganado vacuno.

7. BIBLIOGRAFÍA

7. BIBLIOGRAFÍA

- Ajmone P., Milanese E. & Negrini R. (2004) "Trazabilidad de la raza con métodos moleculares". *Bruna* 2004.
- Ammendrup S. & Fussel A.E. (2001) Legislative requirements for the identification and traceability of farm animals within the European Union. *Rev Sci Tech* **20**, 437-44.
- Andersen J.S. & Mann M. (2000) Functional genomics by mass spectrometry. *FEBS Lett* **480**, 25-31.
- Arana A., Soret B., Lasa I. & Alfonso L. (2002) Meat traceability using DNA markers: application to the beef industry. *Meat Science* **61**, 367-73.
- Arranz J.J., Coppieters W., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mezer C., Riquet J., Simon P., Vanmanshoven P., Wagenaar D. & Georges M. (1998) A QTL affecting milk yield and composition maps to bovine chromosome 20: a confirmation. *Anim Genet* **29**, 107-15.
- Banks M.A. & Eichert W. (2000) WHICHRUN (version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. *J Hered* **91**, 87-9.
- Baumung R., Simianer H. & Hoffmann I. (2004) Genetic diversity studies in farm animals - a survey. *J. Anim. Breed. Genet.* **121**, 361-73.
- Ben-Ari G., Zenvirth D., Sherman A., Simchen G., Lavi U. & Hillel J. (2005) Application of SNPs for assessing biodiversity and phylogeny among yeast strains. *Heredity* **95**, 493-501.
- Bishop M.D., Kappes S.M., Keele J.W., Stone R.T., Sunden S.L.F., Hawkins G.A., Toldo S.S., Fries R., Grosz M.D., Yoo J.Y. & Beattie C.W. (1994) A Genetic-Linkage Map for Cattle. *Genetics* **136**, 619-39.
- Blott S.C., Williams J.L. & Haley C.S. (1999) Discriminating among cattle breeds using genetic markers. *Heredity* **82**, 613-9.
- Boenke A. (2002) Food safety in Europe (FOSIE): Risk assessment of chemicals in food and diet - Preface. *Food and Chemical Toxicology* **40**, 137-.

- Botstein D., White R.L., Skolnick M. & Davis R.W. (1980) Construction of a Genetic-Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *American Journal of Human Genetics* **32**, 314-31.
- Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R. & Cavalli-Sforza L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455-7.
- Brambilla G. & De Filippis S. (2005) Trends in animal feed composition and the possible consequences on residue tests. *Analytica Chimica Acta* **529**, 7-13.
- Brinkmann B., Klintschar M., Neuhuber F., Huhne J. & Rolf B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* **62**, 1408-15.
- Brookes A.J. (1999) The essence of SNPs. *Gene* **234**, 177-86.
- Buchanan F. & Thue T.D. (1998) Intrabreed polymorphic information content of microsatellites in cattle and sheep. *Can. J. Anim. Sci.* **78**: 425-428.
- Canon J., Tupac-Yupanqui I., Garcia-Atance M.A., Cortes O., Garcia D., Fernandez J. & Dunner S. (2008) Genetic variation within the Lidia bovine breed. *Anim Genet* **39**, 439-45.
- Capoferri R., Bongioni G., Galli A. & Aleandri R. (2006) Genetic control of conventional labeling through the bovine meat production chain by single nucleotide polymorphisms using real-time PCR. *Journal of Food Protection* **69**, 1971-7.
- Capoferri R., Galli A. & Bonguioni G. (2005) Molecular traceability in meat producing animals by SNPs. *4° World Italian Beef Cattle Congress, Italy, April-May 2005*.
- Caporale V., Giovannini A., Di Francesco C. & Calistri P. (2001) Importance of the traceability of animals and animal products in epidemiology. *Revue Scientifique Et Technique De L Office International Des Epizooties* **20**, 372-8.
- Carrión D., Day, A., Evans G., Mitsuhashi T., Archibald A., Haley C., Andersson L. & Plastow G. (2003) The use of MC1R and KIT genotypes for breeds characterization. *Arch. Zootec* **52**, 237-44.

- Casas E., Shackleford S.D., Keele J.W., Stone R.T., Kappes S.M. & Koohmaraie M. (2000) Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *Journal of Animal Science* **78**, 560-9.
- Casas E., White S.N., Riley D.G., Smith T.P.L., Brenneman R.A., Olson T.A., Johnson D.D., Coleman S.W., Bennett G.L. & Chase C.C. (2005) Assessment of single nucleotide polymorphisms in genes residing on chromosomes 14 and 29 for association with carcass composition traits in *Bos indicus* cattle. *Journal of Animal Science* **83**, 13-9.
- Ciampolini R., Moazami-Goudarzi K., Vaiman D., Dillmann C., Mazzanti E., Foulley J.L., Leveziel H. & Cianci D. (1995) Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. *J Anim Sci* **73**, 3259-68.
- Cornide M.T. (2002) Marcadores moleculares: nuevos horizontes en la genética y la selección de las plantas. Ed. Félix Varela, La Habana, Cuba.
- Cornuet J.M., Piry S., Luikart G., Estoup A. & Solignac M. (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.
- Cunningham E.P. (2000) Traceability from farm to consumption. 46^o Congreso Internacional de Ciencia y Tecnología de Carnes.
- Cunningham E.P. & Meghen C.M. (2001) Biological identification systems: genetic markers. *Rev Sci Tech* **20**, 491-9.
- Cydne L.H., Clinton S., Jeanette M.W., Katherine D.L., Theresa N., Bruce B. & Sean W. (2000) Practical applications of genotypic surveys for forensic STR testing. *Forensic Science International* **112**, 91-109.
- Chakraborty R. & Jin L. (1993) A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. *EXS* **67**, 153-75.
- Chambers G.K. & MacAvoy E.S. (2000) Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **126**, 455-76.

- Dalvit C., De Marchi M., Dal Zotto R., Zanetti E., Meuwissen T. & Cassandro M. (2008) Genetic characterization of the Burlina cattle breed using microsatellites markers. *J Anim Breed Genet* **125**, 137-44.
- Dietz A.B., Womack J.E., Swarbrick P.A. & Crawford A.M. (1993) Assignment of five polymorphic ovine microsatellites to bovine syntenic groups. *Anim Genet* **24**, 433-6.
- Felmer R., Chavez R., Catrileo A. & Rojas C. (2006) Tecnologías actuales y emergentes para la identificación animal y su aplicación en la trazabilidad animal. *Arch. med. vet.*, **38**, 197-206.
- Felsestein J. (1995) PHYLIP (Phylogeny Inference Package) Version 3.5c. University of Washington.
- Ferreira M.A. & Grattapaglia D. (1995) Introdução ao uso de marcadores moleculares em análise genética. (2nd ed.) *Embrapa*.
- Forbes S.H., Hogg J.T., Buchanan F.C., Crawford A.M. & Allendorf F.W. (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. *Mol Biol Evol* **12**, 1106-13.
- Fries R. & Durstewitz G. (2001) Digital DNA signatures for animal tagging. *Nat Biotechnol* **19**, 508.
- Fries R., Eggen A. & Stranzinger G. (1990) The bovine genome contains polymorphic microsatellites. *Genomics* **8**, 403-6.
- Garber R.A. & Morris J.W. (1983) General equations for the average power of exclusion for genetic systems of n co-dominant alleles in one-parent cases of disputed parentage. In: *Inclusion Probabilities in Parentage Testing*. R. H. Walker. ed. American Association of Blood Banks, Arlington, Virginia, 277-80.
- Georges M. & Massey J.M. (1992) Polymorphic DNA markers in Bovidae. *Patent WO 92/13102 (1992)*.
- Germer S., Holland M.J. & Higuchi R. (2000) High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome Res* **10**, 258-66.
- Giovambattista G., Ripoli M.V., De Luca J.C., Mirol P.M., Liron J.P. & Dulout F.N. (2000) Male-mediated introgression of *Bos indicus* genes into Argentine and Bolivian Creole cattle breeds. *Anim Genet* **31**, 302-5.

- Giuffra E., Tornsten A., Marklund S., Bongcam-Rudloff E., Chardon P., Kijas J.M., Anderson S.I., Archibald A.L. & Andersson L. (2002) A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking KIT. *Mamm Genome* **13**, 569-77.
- Glowatzki-Mullis M.L., Gaillard C., Wigger G. & Fries R. (1995) Microsatellite-based parentage control in cattle. *Anim Genet* **26**, 7-12.
- Goffaux F., China B., Dams L., Clinquart A. & Daube G. (2005) Development of a genetic traceability test in pig based on single nucleotide polymorphism detection. *Forensic Sci Int* **151**, 239-47.
- Gomez-Raya L., Priest K., Rauw W.M., Okomo-Adhiambo M., Thain D., Bruce B., Rink A., Torell R., Grellman L., Narayanan R. & Beattie C.W. (2008) The value of DNA paternity identification in beef cattle: examples from Nevada's free-range ranches. *J Anim Sci* **86**, 17-24.
- Green A., Basile R. & Rumberger J.M. (2006) Transferrin and iron induce insulin resistance of glucose transport in adipocytes. *Metabolism* **55**, 1042-5.
- Griffiths A.J.F., Miller J.H., Suzuki D.T., Lewontin R.C. & Gelbart W.M. (2000) *Genética*. 7^a ed. McGraw Hill- Interamericana.
- Grundel H. & Reetz I. (1981) Exclusion probabilities obtained by biochemical polymorphisms in dogs. *Animal Blood Groups and Biochemical Genetics* **12**, 123-32.
- Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**, 95-8.
- Hamada H., Petrino M.G. & Kakunaga T. (1982) A Novel Repeated Element with Z-DNA-Forming Potential Is Widely Found in Evolutionarily Diverse Eukaryotic Genomes. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79**, 6465-9.
- Heaton M.P., Chitko-McKnown C.G., Grosse W.M., Keele J.W., Keen J.E. & Laegreid W.W. (2001a) Interleukin-8 haplotype structure from nucleotide sequence variation in commercial populations of U.S. beef cattle. *Mamm Genome* **12**, 219-26.

- Heaton M.P., Grosse W.M., Kappes S.M., Keele J.W., Chitko-McKown C.G., Cundiff L.V., Braun A., Little D.P. & Laegreid W.W. (2001b) Estimation of DNA sequence diversity in bovine cytokine genes. *Mamm Genome* **12**, 32-7.
- Heaton M.P., Harhay G.P., Bennett G.L., Stone R.T., Grosse W.M., Casas E., Keele J.W., Smith T.P.L., Chitko-McKown C.G. & Laegreid W.W. (2002) Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. *Mammalian Genome* **13**, 272-81.
- Heaton M.P., Keen J.E., Clawson M.L., Harhay G.P., Bauer N., Shultz C., Green B.T., Durso L., Chitko-McKown C.G. & Laegreid W.W. (2005) Use of bovine single nucleotide polymorphism markers to verify sample tracking in beef processing. *J Am Vet Med Assoc* **226**, 1311-4.
- Herraez D.L., Schafer H., Mosner J., Fries H.R. & Wink M. (2005) Comparison of microsatellite and single nucleotide polymorphism markers for the genetic analysis of a Galloway cattle population. *Zeitschrift Fur Naturforschung C-a Journal of Biosciences* **60**, 637-43.
- Heyen D.W., Beaver J.E., Da Y., Evert R.E., Green C., Bates S.R., Ziegler J.S. & Lewin H.A. (1997) Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semiautomated parentage testing. *Anim Genet* **28**, 21-7.
- Ilbery B. & Kneafsey M. (2000) Producer constructions of quality in regional speciality food production: a case study from south west England. *Journal of Rural Studies* **16**, 210-30.
- Imai K., Matsughige T., Watanabe T., Sugimoto Y. & Ihara N. (2007) Mapping of a quantitative trait locus for beef marbling on bovine chromosome 9 in purebred Japanese black cattle. *Anim Biotechnol* **18**, 75-80.
- Inagaki S., Yamamoto Y., Doi Y., Takata T., Ishikawa T., Yoshitome K., Miyaishi S. & Ishizu H. (2002) Typing of Y chromosome single nucleotide polymorphisms in a Japanese population by a multiplexed single nucleotide primer extension reaction. *Leg Med (Tokyo)* **4**, 202-6.
- Jamieson A. (1965) The genetics of transferrins in cattle. *Heredity* **20**, 419-41.

- Jamieson A. (1994) The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. *Anim Genet* **25 Suppl 1**, 37-44.
- Jamieson A. & Taylor S.C. (1997) Comparisons of three probability formulae for parentage exclusion. *Anim Genet* **28**, 397-400.
- Kapferer J.N. (1992) Strategic brand management: New approaches to creating and evaluating brand equity. *London: Koran Page*.
- Kemp S.J., Hishida O., Wambugu J., Rink A., Longeri M.L., Ma R.Z., Da Y., Lewin H.A., Barendse W. & Teale A.J. (1995) A panel of polymorphic bovine, ovine and caprine microsatellite markers. *Anim Genet* **26**, 299-306.
- Kennedy G.C., Matsuzaki H., Dong S., Liu W.M., Huang J., Liu G., Su X., Cao M., Chen W., Zhang J., Liu W., Yang G., Di X., Ryder T., He Z., Surti U., Phillips M.S., Boyce-Jacino M.T., Fodor S.P. & Jones K.W. (2003) Large-scale genotyping of complex DNA. *Nat Biotechnol* **21**, 1233-7.
- Kijas J.M., Wales R., Tornsten A., Chardon P., Moller M. & Andersson L. (1998) Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* **150**, 1177-85.
- Kondrashov A.S. (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* **21**, 12-27.
- Koreth J., OLeary J.J. & McGee J.O. (1996) Microsatellites and PCR genomic analysis. *Journal of Pathology* **178**, 239-48.
- Krawczak M. (1999) Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* **20**, 1676-81.
- Kruger K., Gaillard C., Stranzinger G. & Rieder S. (2005) Phylogenetic analysis and species allocation of individual equids using microsatellite data. *J Anim Breed Genet* **122 Suppl 1**, 78-86.
- Kruglyak L. (1997) The use of a genetic map of biallelic markers in linkage studies. *Nat Genet* **17**, 21-4.
- Lahiri D.K., Bye S., Nurnberger J.I., Jr., Hodes M.E. & Crisp M. (1992) A non-organic and non-enzymatic extraction method gives higher yields

- of genomic DNA from whole-blood samples than do nine other methods tested. *J Biochem Biophys Methods* **25**, 193-205.
- Laurent P. & Rodellar C. (2001) Characterization of a single nucleotide polymorphism in the coding sequence of the bovine transferrin gene. *Mutat Res* **458**, 1-5.
- Li Y., Wenzel F., Holzgreve W. & Hahn S. (2006) Genotyping fetal paternally inherited SNPs by MALDI-TOF MS using cell-free fetal DNA in maternal plasma: influence of size fractionation. *Electrophoresis* **27**, 3889-96.
- Lindblad-Toh K., Winchester E., Daly M.J., Wang D.G., Hirschhorn J.N., Laviolette J.P., Ardlie K., Reich D.E., Robinson E., Sklar P., Shah N., Thomas D., Fan J.B., Gingeras T., Warrington J., Patil N., Hudson T.J. & Lander E.S. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* **24**, 381-6.
- Liron J.P., Peral-Garcia P. & Giovambattista G. (2006) Genetic characterization of Argentine and Bolivian Creole cattle breeds assessed through microsatellites. *J Hered* **97**, 331-9.
- Litt M. & Luty J.A. (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* **44**, 397-401.
- Liu Q., Thorland E.C., Heit J.A. & Sommer S.S. (1997) Overlapping PCR for bidirectional PCR amplification of specific alleles: a rapid one-tube method for simultaneously differentiating homozygotes and heterozygotes. *Genome Res* **7**, 389-98.
- Livak K.J., Marmaro J. & Todd J.A. (1995) Towards fully automated genome-wide polymorphism screening. *Nat Genet* **9**, 341-2.
- Lundqvist E., Johansson I. & Ingelman-Sundberg M. (1999) Genetic mechanisms for duplication and multiduplication of the human CYP2D6 gene and methods for detection of duplicated CYP2D6 genes. *Gene* **226**, 327-38.
- Markovtsova L., Marjoram P. & Tavaré S. (2000) The age of a unique event polymorphism. *Genetics* **156**, 401-9.

- Martin-Burriel I., Garcia-Muro E. & Zaragoza P. (1999) Genetic diversity analysis of six Spanish native cattle breeds using microsatellites. *Anim Genet* **30**, 177-82.
- Martin-Burriel I., Rodellar C., Lenstra J.A., Sanz A., Cons C., Osta R., Reta M., De Arguello S. & Zaragoza P. (2007) Genetic Diversity and Relationships of Endangered Spanish Cattle Breeds. *J Hered.*
- Maudet C., Luikart G. & Taberlet P. (2002) Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. *J Anim Sci* **80**, 942-50.
- McKean J.D. (2001) The importance of traceability for public health and consumer protection. *Revue Scientifique Et Technique De L Office International Des Epizooties* **20**, 363-71.
- Meghen C.N., Scott C.S., Bradley D.G., MacHugh D.E., Loftus R.T. & Cunningham E.P. (1998) DNA based traceability techniques for the meat industry. *Anim. Genet.*, **29**, 48-9.
- Mertens P.P.C., Maan S., Samuel A. & Attoui H. (2005) Orbivirus, Reoviridae. Virus Taxonomy, VIIIth Report of the ICTV, Elsevier/Academic Press, London, pp. 466-483. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.).
- Minch E. (1995) MICROSAT 1.4d, a program for calculating distances from microsatellite data. Available from <http://hpgl.stanford.edu/projects/microsat/>.
- Moazami-Goudarzi K., Laloe D., Furet J.P. & Grosclaude F. (1997) Analysis of genetic relationships between 10 cattle breeds with 17 microsatellites. *Anim Genet* **28**, 338-45.
- Mommens G., Van Zeveren A. & Peelman L.J. (1998) Effectiveness of bovine microsatellites in resolving paternity cases in American bison, Bison bison L. *Animal Genetics* **29**, 12-8.
- Moore S.S. & Byrne K. (1993) Dinucleotide polymorphism at the bovine calmodulin independent adenylylase locus. *Anim Genet* **24**, 150.
- Moran C. (1993) Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes. *J Hered* **84**, 274-80.

- Mullis K.B. & Faloona F.A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155**, 335-50.
- Nachman M.W. & Crowell S.L. (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297-304.
- Negrini R., Nicoloso L., Crepaldi P., Milanese E., Colli L., Chegiani F., Pariset L., Dunner S., Leveziel H., Williams J.L. & Ajmone Marsan P. (2009) Assessing SNP markers for assigning individuals to cattle populations. *Anim Genet* **40**, 18-26.
- Negrini R., Nicoloso L., Crepaldi P., Milanese E., Marino R., Perini D., Pariset L., Dunner S., Leveziel H., Williams J.L. & Ajmone Marsan P. (2008) Traceability of four European Protected Geographical Indication (PGI) beef products using Single Nucleotide Polymorphisms (SNP) and Bayesian statistics. *Meat Science* **80**, 1212-7
- Nei M., Tajima F. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J Mol Evol* **19**, 153-70.
- Nielsen R. (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* **154**, 931-42.
- Ordovas L. (2006) Tesis doctoral: "Análisis de genes candidatos a QTLs relacionados con el metabolismo lipídico en la especie bovina".
- Ordovas L., Roy R., Zaragoza P., Hayes H., Eggen A. & Rodellar C. (2005) Assignment of the solute carrier family 27 member 1 (SLC27A1) gene to bovine chromosome 7. *Anim Genet* **36**, 352-3.
- Ordovas L., Roy R., Zaragoza P. & Rodellar C. (2006) Structural and functional characterization of the bovine solute carrier family 27 member 1 (SLC27A1) gene. *Cytogenet Genome Res* **115**, 115-22.
- Ordovas L., Zaragoza P., Altarriba J. & Rodellar C. (2008) Identification of 14 new Single Nucleotide Polymorphism in the bovine *SLC27A1* gene and evaluation of their association with milk fat content. *Journal of Dairy Research* **75**, 129-34.
- Orrù L., Catillo G., Napolitano F., De Matteis G., Scatà M.C., F. S. & Moiolib B. (2009) Characterization of a SNPs panel for meat traceability in six cattle breeds *Food Control* **20**, 856-60.

- Orru L., Napolitano F., Catillo G. & Moioli B. (2006) Meat molecular traceability: how to choose the best set of microsatellites? . *Meat Science*, 312–7.
- Osta R. (1994) Caracterización genética de proteínas lácteas y sexaje de embriones en ganado vacuno mediante la aplicación de la biotecnología al análisis del ADN. *Tesis doctoral. Facultad de Veterinaria, Universidad de Zaragoza.*
- Paetkau D., Calvert W., Stirling I. & Strobeck C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* **4**, 347-54.
- Page B.T., Casas E., Heaton M.P., Cullen N.G., Hyndman D.L., Morris C.A., Crawford A.M., Wheeler T.L., Koohmaraie M., Keele J.W. & Smith T.P. (2002) Evaluation of single-nucleotide polymorphisms in CAPN1 for association with meat tenderness in cattle. *J Anim Sci* **80**, 3077-85.
- Page R.D.M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357-8.
- Pakstis A.J., Speed W.C., Kidd J.R. & Kidd K.K. (2007) Candidate SNPs for a universal individual identification panel. *Hum Genet* **121**, 305-17.
- Pastinen T., Partanen J. & Syvanen A.C. (1996) Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. *Clin Chem* **42**, 1391-7.
- Peelman L.J., Mortiaux F., Van Zeveren A., Dansercoer A., Mommens G., Coopman F., Bouquet Y., Burny A., Renaville R. & Portetelle D. (1998) Evaluation of the genetic variability of 23 bovine microsatellite markers in four Belgian cattle breeds. *Animal Genetics* **29**, 161-7.
- Peiris J.S., de Jong M.D. & Guan Y. (2007) Avian influenza virus (H5N1): a threat to human health. *Clin Microbiol Rev* **20**, 243-67.
- Polineni P. (2004) Developing a web accessible integrated database and visualization tool for bovine quantitative trait loci. . *A Thesis for the degree of Master of Science. Sri Venkateswara University.*
- Portelle V., Haezebroeck F. & Renaville. M.a.R. (2000) Tracabilité dans la filière animale. *Biotechnol. Agron. Soc. Environ* **4**, 233–40.

- Pritchard J.K., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59.
- Puntaric D., Smit Z., Bosnir J. & Topolovec J. (2000) Small countries and the dioxin scandal: how to control imported food? *Croat Med J* **41**, 150-3.
- Rannala B. & Mountain J.L. (1997) Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci U S A* **94**, 9197-201.
- Raymond M. & Rousset F. (1995) Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**, 248-9.
- Rejholcova M., Wilhelm J. & Svoboda P. (1988) Lipid peroxidation inhibits norepinephrine-stimulated lipolysis in rat adipocytes. Reduction of beta-adreno-ceptor number. *Biochem Biophys Res Commun* **150**, 802-10.
- Rengmark, A. H., Slettan A., Skaala Ø., ; L.Ø. & Lingaas F. (2006) Genetic variability in wild and farmed Atlantic salmon (*Salmo salar*) strains estimated by SNP and microsatellites. *Aquaculture* **253**, 229-37.
- Ritz L.R., Glowatzki-Mullis M.-L., MacHugh D.E. & Gaillard C. (2000) Phylogenetic analysis of the tribe Bovini using microsatellites. *Animal Genetics* **31**, 178-85.
- Rodellar C., Sanz A., Osta R., Martín-Burriel I., Zarazaga I. & Zaragoza P. (2006) "Trazabilidad genética de los productos cárnicos en rumiantes".
- Ronaghi M., Uhlen M. & Nyren P. (1998) A sequencing method based on real-time pyrophosphate. *Science* **281**, 363-+.
- Roy R., Gautier M., Hayes H., Laurent P., Osta R., Zaragoza P., Eggen A. & Rodellar C. (2001) Assignment of the fatty acid synthase (FASN) gene to bovine chromosome 19 (19q22) by in situ hybridization and confirmation by somatic cell hybrid mapping. *Cytogenet Cell Genet* **93**, 141-2.
- Roy R., Ordovas L., Zaragoza P., Romero A., Moreno C., Altarriba J. & Rodellar C. (2006) Association of polymorphisms in the bovine FASN gene with milk-fat content. *Anim Genet* **37**, 215-8.

- Rumberger J.M., Peters T., Jr., Burrington C. & Green A. (2004) Transferrin and iron contribute to the lipolytic effect of serum in isolated adipocytes. *Diabetes* **53**, 2535-41.
- Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mullikin J.C., Mortimore B.J., Willey D.L., Hunt S.E., Cole C.G., Coggill P.C., Rice C.M., Ning Z., Rogers J., Bentley D.R., Kwok P.Y., Mardis E.R., Yeh R.T., Schultz B., Cook L., Davenport R., Dante M., Fulton L., Hillier L., Waterston R.H., McPherson J.D., Gilman B., Schaffner S., Van Etten W.J., Reich D., Higgins J., Daly M.J., Blumenstiel B., Baldwin J., Stange-Thomann N., Zody M.C., Linton L., Lander E.S. & Altshuler D. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-33.
- Saitou N. & Nei M. (1987) The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution* **4**, 406-25.
- Sancristobal-Gaudy M., Renand G., Amigues Y., Boscher M.Y., Leveziel H. & Bibe B. (2000) Tracabilité individuelle des viandes bovines à l'aide de marqueurs génétiques. *INRA Prod. Anim. Genet.*, **13**, 269-76.
- Sanger F., Nicklen S. & Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7.
- Sanz A., Martín-Burriel I., Rodellar C., Osta R., Sanz A., Abril F. & Zaragoza P. (2007) Caracterización genética de la población bovina Serrana Negra de Teruel. *Archivos de Zootecnia* **56**, 461-5.
- Sanz A., Ordovás L., Serrano C., Zaragoza P., Altarriba J. & Rodellar C. (2010) SNP in the coding region of bovine transferrin is associated with milk fat yield. *Genetics and molecular Research. Aceptado para publicar*.
- Sanz A., Uffo O., Miranda I. & Martínez S. (2002) Empleo de los microsatélites para determinar paternidad en bovinos cubanos. . *Rev. Salud Anim.* **24**, 166-9.
- Sarma A.V., Dunn R.L., Lange L.A., Ray A., Wang Y., Lange E.M. & Cooney K.A. (2008) Genetic polymorphisms in CYP17, CYP3A4, CYP19A1,

- SRD5A2, IGF-1, and IGFBP-3 and prostate cancer risk in African-American men: the Flint Men's Health Study. *Prostate* **68**, 296-305.
- Sauer S., Gelfand D.H., Boussicault F., Bauer K., Reichert F. & Gut I.G. (2002) Facile method for automated genotyping of single nucleotide polymorphisms by mass spectrometry. *Nucleic Acids Res* **30**, e22.
- Schena M., Heller R.A., Theriault T.P., Konrad K., Lachenmeier E. & Davis R.W. (1998) Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol* **16**, 301-6.
- Schnabel R.D., Ward T.J. & Derr J.N. (2000) Validation of 15 microsatellites for parentage testing in North American bison, *Bison bison* and domestic cattle. *Anim Genet* **31**, 360-6.
- Schopen G.C., Bovenhuis H., Visker M.H. & van Arendonk J.A. (2008) Comparison of information content for microsatellites and SNPs in poultry and cattle. *Anim Genet* **39**, 451-3.
- Severi G., Hayes V.M., Tesoriero A.A., Southey M.C., Hoang H.N., Padilla E.J., Morris H.A., English D.R., Sutherland R.L., Boyle P., Hopper J.L. & Giles G.G. (2008) The rs743572 common variant in the promoter of CYP17A1 is not associated with prostate cancer risk or circulating hormonal levels. *BJU Int* **101**, 492-6.
- Shackell G.H., Tate M.L. & Anderson R.M. (2001) Installing a DNA-based traceability system in the meat industry. *Proc. Assoc. Advmt. Anim. Breed. Genet*, 533-6.
- Sherman G.B., Kachman S.D., Hungerford L.L., Rupp G.P., Fox C.P., Brown M.D., Feuz B.M. & Holm T.R. (2004) Impact of candidate sire number and sire relatedness on DNA polymorphism-based measures of exclusion probability and probability of unambiguous parentage. *Anim Genet* **35**, 220-6.
- Smith P.G., Cousens S.N., d' Huillard Aignaux J.N., Ward H.J. & Will R.G. (2004) The epidemiology of variant Creutzfeldt-Jakob disease. *Curr Top Microbiol Immunol* **284**, 161-91.
- Smith T.P., Casas E., Rexroad C.E., 3rd, Kappes S.M. & Keele J.W. (2000) Bovine CAPN1 maps to a region of BTA29 containing a quantitative trait locus for meat tenderness. *J Anim Sci* **78**, 2589-94.

- Snelling W.M., Casas E., Stone R.T., Keele J.W., Harhay G.P., Bennett G.L. & Smith T.P. (2005) Linkage mapping bovine EST-based SNP. *BMC Genomics* **6**, 74.
- Stanford K., Stitt J., Kellar A.J. & McAllister T.A. (2001) Traceability in cattle and small ruminants in Canada. *Revue Scientifique et Technique de l'Office International des Epizooties* **20**, 510–22.
- Steffen P., Eggen A., Dietz A.B., Womack J.E., Stranzinger G. & Fries R. (1993) Isolation and mapping of polymorphic microsatellites in cattle. *Anim Genet* **24**, 121-4.
- Stone R.T., Casas E., Smith T.P., Keele J.W., Harhay G., Bennett G.L., Koochmaraie M., Wheeler T.L., Shackelford S.D. & Snelling W.M. (2005) Identification of genetic markers for fat deposition and meat tenderness on bovine chromosome 5: development of a low-density single nucleotide polymorphism map. *J Anim Sci* **83**, 2280-8.
- Stone R.T., Grosse W.M., Casas E., Smith T.P.L., Keele J.W. & Bennett G.L. (2002) Use of bovine EST data and human genomic sequences to map 100 gene-specific bovine markers. *Mammalian Genome* **13**, 211-5.
- Taboada C.D. (1995) Aislamiento de ADN total a partir de biopsias. . *Protocolos generales*.
- Tapio I., Varv S., Bennewitz J., Maleviciute J., Fimland E., Grislis Z., Meuwissen T.H., Miceikiene I., Olsaker I., Viinalass H., Vilkki J. & Kantanen J. (2006) Prioritization for conservation of northern European cattle breeds based on analysis of microsatellite data. *Conserv Biol* **20**, 1768-79.
- Tautz D. (1989) Hypervariability of Simple Sequences as a General Source for Polymorphic DNA Markers. *Nucleic Acids Research* **17**, 6463-71.
- Thirstrup J.P., Pertoldi C. & Loeschcke V. (2008) Genetic analysis, breed assignment and conservation priorities of three native Danish horse breeds. *Anim Genet* **39**, 496-505.
- Thomas J.H. (2007) Rapid birth-death evolution specific to xenobiotic cytochrome P450 genes in vertebrates. *PLoS Genet* **3**, e67.

- Thomson R., Pritchard J.K., Shen P., Oefner P.J. & Feldman M.W. (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A* **97**, 7360-5.
- Toldo S.S., Fries R., Steffen P., Neiberghs H.L., Barendse W., Womack J.E., Hetzel D.J.S. & Stranzinger G. (1993) Physically Mapped, Cosmid-Derived Microsatellite Markers as Anchor Loci on Bovine Chromosomes. *Mammalian Genome* **4**, 720-7.
- Twito T., Weigend S., Blum S., Granevitze Z., Feldman M.W., Perl-Treves R., Lavi U. & Hillel J. (2007) Biodiversity of 20 chicken breeds assessed by SNPs located in gene regions. *Cytogenet Genome Res* **117**, 319-26.
- Uffo O. (2003) Aplicación de los marcadores al estudio de la biodiversidad del ganado bovino cubano. *Tesis doctoral*.
- Usha A.P., Simpson S.P. & Williams J.L. (1995) Probability of random sire exclusion using microsatellite markers for parentage verification. *Anim Genet* **26**, 155-61.
- Vaiman D., Mercier D., Moazami-Goudarzi K., Eggen A., Ciampolini R., Lepingle A., Velmala R., Kaukinen J., Varvio S.L., Martin P. & et al. (1994) A set of 99 cattle microsatellites: characterization, synteny mapping, and polymorphism. *Mamm Genome* **5**, 288-97.
- Vallat B. (2001) Preface - Traceability of animals and animal products. *Revue Scientifique Et Technique De L Office International Des Epizooties* **20**, 359-.
- Van Eenennaam A.L., Weaber R.L., Drake D.J., Penedo M.C., Quaas R.L., Garrick D.J. & Pollak E.J. (2007) DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting. *J Anim Sci* **85**, 3159-69.
- Van Laere A.S., Nguyen M., Braunschweig M., Nezer C., Collette C., Moreau L., Archibald A.L., Haley C.S., Buys N., Tally M., Andersson G., Georges M. & Andersson L. (2003) A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832-6.

- Vazquez J.F., Perez T., Urena F., Gudín E., Albornoz J. & Dominguez A. (2004) Practical application of DNA fingerprinting to trace beef. *Journal of Food Protection* **67**, 972-9.
- Verbeke W., Frewer L.J., Scholderer J. & De Brabander H.F. (2007) Why consumers behave as they do with respect to food safety and risk information. *Analytica Chimica Acta* **586**, 2-7.
- Viana J.L., Bouzada J.A., Prado C., Areán H., Muíño R., López M. & Fernández A. (2005) Trazabilidad de la carne de vacuno mediante el análisis de 17 Marcadores microsatélite de ADN. *CONGRESOS/AIDA2005*.
- Vignal A., Milan D., SanCristobal M. & Eggen A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* **34**, 275-305.
- Wang D.G., Fan J.B., Siao C.J., Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., Spencer J., Kruglyak L., Stein L., Hsie L., Topaloglou T., Hubbell E., Robinson E., Mittmann M., Morris M.S., Shen N., Kilburn D., Rioux J., Nusbaum C., Rozen S., Hudson T.J., Lipshutz R., Chee M. & Lander E.S. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077-82.
- Weber J.L. & May P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* **44**, 388-96.
- Weir B.S. (1996) Genetic data analysis II. Methods for discrete population genetic data. *Sunderland, Massachusetts: Sinauer Associates, Inc. Publishers*.
- Weller J.I., Seroussi E. & Ron M. (2006) Estimation of the number of genetic markers required for individual animal identification accounting for genotyping errors. *Anim Genet* **37**, 387-9.
- Werner F.A., Durstewitz G., Habermann F.A., Thaller G., Kramer W., Kollers S., Buitkamp J., Georges M., Brem G., Mosner J. & Fries R. (2004) Detection and characterization of SNPs useful for identity control and

- parentage testing in major European dairy breeds. *Anim Genet* **35**, 44-9.
- Werner F.A.O., Durstewitz G., Thaller G., Barkemeyer H., Irps U., Monsner J. & Fries R. (2002) Standardized presentation of SNP genotypes for paternity testing, individual identification and genetic distance analysis. . *I.S.A.G.*
- Winter A., Kramer W., Werner F.A., Kollers S., Kata S., Durstewitz G., Buitkamp J., Womack J.E., Thaller G. & Fries R. (2002) Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proc Natl Acad Sci U S A* **99**, 9300-5.
- Zaragoza P. (2007) Informe Final del Contrato de Investigación sobre “Caracterización genética de la raza Parda de Montaña (Aragón) para su fomento y recuperación. Comparación con otras razas bovinas autóctonas españolas”.
- Zaragoza P., Rodellar C., Sanz A., Ordovás L. & Zarazaga I. (2003) Estrategias genéticas para la optimización de la Calidad y Seguridad Alimentaria. *Jornadas de Investigación del Sector de Salud III. Zaragoza, 12 de diciembre de 2003.*

