UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS BIOLÓGICAS

Departamento de Genética

TESIS DOCTORAL

Estrategias de manejo poblacional para revertir introgresión de genes exógenos

Management strategies to remove exogenous introgression

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Carmen Amador Arranz

Director

Jesús Fernández Martín

Madrid, 2013

Estrategias de manejo poblacional para revertir introgresión de genes exógenos

Carmen Amador Arranz

TESIS DOCTORAL
MADRID 2012

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE GENÉTICA

ESTRATEGIAS DE MANEJO POBLACIONAL PARA REVERTIR INTROGRESIÓN DE GENES EXÓGENOS

MANAGEMENT STRATEGIES TO REMOVE EXOGENOUS INTROGRESSION

Carmen Amador Arranz

TESIS DOCTORAL

MADRID 2012

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE GENÉTICA

ESTRATEGIAS DE MANEJO POBLACIONAL PARA REVERTIR INTROGRESIÓN DE GENES EXÓGENOS

MEMORIA PARA OPTAR AL GRADO DE DOCTOR EN BIOLOGÍA

MENCIÓN EUROPEA

Carmen Amador Arranz

Vº Bº del Director

Dr. Jesús Fernández Martín

La Doctoranda

Carmen Amador Arranz

MADRID 2012

Esta Tesis Doctoral se realizó en el Instituto de Investigación y Tecnología Agraria y Alimentaria (INIA) de Madrid, y ha sido financiada con una beca predoctoral FPI-INIA y con el proyecto CGL2009-13278-C02-02 (Plan Nacional de Ciencia y Tecnología).

Los trabajos expuestos en los Capítulos 3 y 4 se realizaron parcialmente durante estancias predoctorales en el Department of Animal and Aquacultural Sciences de la UMB en Ås (Noruega) y en el Department of Primary Industries, en Bundoora (Australia) respectivamente.

Caminante, son tus huellas
el camino, y nada más

Antonio Machado, Proverbios y Cantares

DON'T PANIC

Douglas Adams, The Hitchhiker's guide to the Galaxy

Sin duda, a mi madre

AGRADECIMIENTOS

Mucha gente no entiende qué es una Tesis. La línea que separa si lo que hacemos es trabajar o estudiar es difícil de trazar en algunas ocasiones y combinamos las desventajas de lo laboral y lo académico. Afortunadamente esta carrera es, a veces, muy agradecida, y además termina con un libro en el que se puede dedicar un par de páginas a la gente que tenemos alrededor. Porque sin ellos, sin vosotros, es imposible terminar un doctorado. Por eso agradezco a las siguientes personas su apoyo y contribución a este trabajo:

A mi director, Jesús Fernández, por la confianza desde el primer día y por ser tan buen jefe. Por tantas oportunidades para ir a congresos y hacer estancias y enseñarme lo mejor de la buena ciencia.

A Miguel Ángel Toro y Beatriz Villanueva por dar otros puntos de vista. Por sus buenas ideas, su buen criterio y su apoyo. Beatriz, no volveré a poner un punto final en una presentación de *Power Point*.

A todos mis compañeros del INIA por tan buenos ratos científicos y lúdicos. A María Muñoz, que con la cosa de ser la predecesora me ha ido solucionando todas las dudas (en persona y online). Porque es una de las mejores personas que conozco y una gran amiga. A Dafne y María Saura, compañeras de despacho en el peor de los momentos, cuyo buen gusto ha marcado muchas decisiones importantes de esta Tesis, estéticas y profesionales, y a Judit, porque a veces las dudas y los dramas pasan al despacho de enfrente. Además nos lo hemos pasado estupendamente. A Ana porque siempre está pendiente de todo, dispuesta a echar una mano y organizar una cena. A Ángeles por todas las conversaciones de camino a casa y su ayuda constante. A Cristina Meneses, mi compañera de cursos y viajes, que aunque últimamente nos vemos sólo en el bus, sigue bien atenta a mis novedades. A Estefânia, que de vez en cuando trae un bizcocho sin gluten. A Óscar y a Fernando, que están siempre en minoría, pero organizan porras y ven series. A todo el grupo pasado y presente de cafés, comidas y congresos: Amanda, Chari, Evangelina, Gloria, Fabián, Silvia, Enrique…, por las risas y los buenos momentos. A toda la gente que cuando me ve por el pasillo me pregunta cómo va la cosa: Carmen Rodríguez, Luis Alberto, Almudena, Luis, María Gil… Gracias.

I would like to thank Prof. Henner Simianer, Theo Meuwissen, Ben Hayes and Hans Daetwyler for their warm welcome and help in scientific matters during my

research stays. Many people helped me during that time, especially in social outings which made the stays really nice experiences. Thanks to Filippo, Christian, Malena, Florence and Eduardo from Göttingen; Sergio, Lorena, Thorben and Damien from Ås and Hans and Jennie from Melbourne.

A la gente del Departamento de Genética de la UCM dónde empecé este camino. A Carlos López-Fanjul y Aurora García-Dorado sin cuya ayuda no estaría aquí. A todos los amigos que aún tengo allí, especialmente a Diego, Patri y Natalia. Y a Bea mucho más, que me cambió el título de la Tesis sin necesidad de mandarle demasiados emails.

A Maialen y a Bea porque aunque pasan millones de años entre vez y vez que nos vemos, parece que nada ha cambiado. A Rafa, Pablo, Asun y Clara por su amistad desde primero (o segundo) de carrera hasta niveles postdoctorales. A Gise y a Araceli porque no hay nada mejor que tomarse un vino español como unas señoras. A Juan, porque viene a verme al INIA aunque no toque. A Carlos, Isa y Jesús, por todas las noches en el Johnny llenas de expectativas y de sueños. A Adri, mi eterna compañera de cuarto, porque no hay nadie como ella y, aunque hablemos poco, la echo de menos todos los días.

A mi familia. A mi madre, a la que le debo todo. Porque me ha enseñado a valerme por mí misma, pero no puedo pasar un día sin pedirle consejo. A mi abuelo, el mejor ejemplo de persona que conozco. A mi abuela, que se perdió esta etapa de mi vida, pero todo lo que me enseñó influye sobre mis decisiones cada día. Lo que he aprendido de ellos es esencial para mi vida y la persona que soy, pero también, sin saberlo, me transmitieron la curiosidad, el rigor científico y la búsqueda de porqués. A Ramón, Alicia, Andrea y Paola, cuyo apoyo es indispensable. A Amparo y Carmen, porque la ciencia es muchas veces como hacer manualidades, y de eso me enseñaron mucho.

Y, dejando lo mejor para el final, a Gonzalo. Porque esta Tesis empezó con una mudanza, y desde entonces, las cosas son más fáciles. Porque estamos juntos en esto, y nos queda mucho camino por recorrer. Estoy deseando ver a dónde nos lleva.

Allons-y!

Los capítulos de esta Tesis se corresponden con las siguientes publicaciones científicas:

1. Amador, C., Toro, M.A. & Fernández, J. 2011 *Removing exogenous information using pedigree data*. Conservation Genetics. 12:1565–1573.
2. Amador, C., Toro, M.A. & Fernández, J. *Molecular markers allow to remove introgressed genetic background: a simulation study*. PLoS One (submitted).
3. Amador, C., Fernández, J. & Meuwissen, T.H.E. *Advantages of using molecular coancestry in the removal of introgressed genetic information*. Genetics Selection Evolution (submitted).
4. Amador, C., Hayes, B.J. & Daetwyler, H.D. *Estimation of genomic breed proportions to remove introgressed genetic material: an example in sheep*. Under preparation.

ÍNDICE

RESUMEN

El intercambio genético entre diferentes poblaciones es un fenómeno habitual en la naturaleza con importantes implicaciones adaptativas y evolutivas. También ha sido una práctica habitual en la gestión de especies domésticas. Sin embargo, en algunos casos un aporte genético proveniente de otra población puede ser indeseado, y con consecuencias negativas sobre la conservación de especies y razas, tanto salvajes como domésticas. Cuando una población que requiere conservarse pura recibe aportes genéticos exógenos, el proceso ha de revertirse para recuperar el genoma original. Para ello, han de aplicarse todas las herramientas disponibles. En este trabajo se estudió, mediante simulaciones por ordenador, la eficiencia en el proceso de desintrogresión a partir de dos fuentes diferentes de información: la genealogía y los marcadores moleculares. En función de la información disponible se probaron varios métodos: i) utilizando el pedigrí: minimización del parentesco con los exógenos y minimización del parentesco parcial debido a los exógenos; ii) utilizando marcadores moleculares: selección de alelos exclusivos de población, minimización de las distancias genéticas con la población original, minimización del parentesco molecular con los exógenos, selección de haplotipos nativos y selección mediante un modelo mixto (GBLUP). Todos los métodos estudiados lograron recuperar parte del genoma nativo, siendo los más eficaces aquellos que utilizan información de genotipado masivo. La cantidad de información genética recuperada se vio limitada en todos los casos por el porcentaje total de introgresión, así como por el tiempo durante el que los genes exógenos se mezclaron en la población. Una consecuencia colateral adversa del proceso fue, en todos los casos, un gran incremento de la consanguinidad, debido a la selección como reproductores de los individuos más puros y por tanto un menor censo efectivo. Esta pérdida de variabilidad genética ha de ser controlada durante el proceso, aunque implique una menor efectividad en la eliminación de información exógena. Para ello hay que buscar una solución de compromiso entre la recuperación del genoma nativo y el mantenimiento de diversidad.

SUMMARY

Exchange of genetic material between populations happens frequently, and it has important adaptive and evolutionary implications. It has also been a common practice in domestic species management. Nevertheless, in some situations, an exogenous input of genetic material can be undesired, with negative consequences on both wild and domestic species. When a population that we want to maintain pure becomes introgressed by another, the original background must be recovered. In this study, computer simulations were used to analyse the ability to recover an introgressed genetic background from two sources of information: pedigree and different types of molecular markers. According to the available information, different strategies were studied: i) using the pedigree: minimum exogenous contribution and minimum partial coancestry, ii) using molecular markers: diagnostic alleles selection, minimum genetic distance with the native population, minimum exogenous molecular coancestry, native haplotypes selection and GBLUP. All the strategies were able to recover part of the native genome, and those that used genome wide information were the most efficient. The amount of native genome recovered was limited in all cases by the total percentage of introgression and the time elapsed from the introgression event until the management started. An increase of inbreeding was a by-product of the recovery process in all cases, as selection on the purest individuals led to a lower effective population size. This loss in genetic variability must be controlled during the process although it will decrease the efficiency in the removal of exogenous information. A compromise must be found between genomic recovery and the maintenance of genetic diversity.

INTRODUCCIÓN

APORTES EXÓGENOS DE MATERIAL GENÉTICO

La pérdida de diversidad genética reduce la habilidad de las poblaciones para adaptarse a nuevos ambientes y produce descensos en la eficacia biológica. Los beneficios del intercambio de recursos genéticos han sido ampliamente estudiados, considerándose que el aporte de material genético externo es un buen método para proteger una población (FRANKHAM et al. 2002).

En algunas ocasiones, los análisis sobre la viabilidad de poblaciones en grave peligro de extinción concluyen que éstas solo pueden ser recuperadas mediante la introducción de nuevo material genético, proveniente de otras razas o especies. En el proceso conocido como rescate genético (*genetic rescue*) se considera que una población sufre una severa depresión consanguínea de la que no puede recuperarse por sus propios medios, y su eficacia (y probabilidad de supervivencia) puede verse aumentada si un grupo de migrantes añade nueva variación genética al evitar la consanguinidad y/o a través de la heterosis de la descendencia (INGVARSSON 2001; TALLMON et al. 2004).

Dos ejemplos de rescate genético son la pantera de Florida (*Puma concolor coryi*) y la población sueca de víboras (*Vipera berus*) de Smygehuk. En ambos casos los efectos negativos de la consanguinidad afectaban gravemente a las poblaciones.

En el primer caso, ocho hembras de pantera de Texas (*Puma concolor stanleyana*) se transfirieron a la población de panteras de Florida (*Puma concolor coryi*) en un programa de introgresión intencional para restaurar su diversidad genética (HEDRICK 1995). En el análisis de seguimiento de la población se hallaron evidencias de vigor híbrido en los cruces F1 con influencia en la supervivencia de panteras (adultas y subadultas), considerándose exitoso en cuanto a la demografía de la población (BENSON et al. 2011).

En el caso de la población de víboras de Smygehuk, 20 machos de otra población se liberaron en la población Smygehuk causando una importante recuperación en su variabilidad genética y un incremento en la supervivencia de individuos juveniles (MADSEN et al. 1999).

Una estrategia similar al rescate genético es la de fusión de razas (*merging breeds*) en esquemas de conservación de especies domésticas. Si una raza es

importante para la diversidad de la especie, está en peligro de extinción y se encuentra en una situación en la que su conservación en pureza es muy difícil, podría ser beneficioso fusionarla con otra u otras razas también en peligro para mantener el máximo de diversidad posible (BENNEWITZ et al. 2008).

También en especies domésticas, es frecuente el cruce con razas comerciales aprovechándose no sólo de los efectos de heterosis y la eliminación de los problemas derivados de la consanguinidad de las razas puras, sino además, para aumentar la productividad y aprovechar las ventajas de razas locales mejor adaptadas a ambientes concretos (GOSEY 1991; SCHAEFFER et al. 2011)

Por otra parte, tanto en plantas como en animales, la introgresión de genes exógenos en diferentes poblaciones puede contribuir al proceso de especiación, dando lugar a nuevos taxones mejor adaptados, considerándose un positivo avance evolutivo (ALLENDORF & LUIKART 2007).

INTROGRESIÓN NO DESEADA

El mantenimiento de una población en pureza puede ser interesante tanto en el campo de especies domésticas como en poblaciones naturales.

Especies domésticas

En esta categoría las razones que nos llevan a la necesidad de un mantenimiento en pureza son mayoritariamente de tipo económico. Por ejemplo, existen poblaciones asociadas directamente a un producto de calidad. De algunos animales se obtienen distintos productos cuyas características dependen de su identidad genética. Los cerdos ibéricos y sus productos curados, las vacas lecheras *Reggiana* y el queso Parmesano Reggiano, y varios ejemplos más (DALVIT et al. 2007), ponen de manifiesto la importancia económica que la pureza del origen puede tener. En estos casos, una preocupación adicional es la de desarrollar herramientas de caracterización y trazabilidad genéticas en el producto final para evitar fraudes (BLOTT et al. 1999).

También algunas razas se encuentran ligadas a la práctica de actividades concretas. Por ejemplo, existen razas de caballo que se utilizan para competiciones deportivas y otros eventos específicos también con implicaciones económicas cuyo

mantenimiento en pureza es imprescindible (MAPA 2003). Existe todo un conjunto de reglamentaciones para registrar y controlar el origen de los individuos.

En último lugar, por motivos estéticos, las razas de perro se mantienen aisladas entre sí. Un perro sólo puede pasar a formar parte de una determinada raza si sus padres lo son, devaluándose su precio cuando se trata de un cruce (PARKER et al. 2004).

También existen razones no económicas por las que la mezcla de poblaciones domésticas pueda estar desaconsejada. Desde el punto de vista de la conservación, el flujo génico puede también ser dañino para alguna de estas especies o razas (tanto animales como vegetales). Las razas domésticas se consideran importantes componentes de la biodiversidad y su conservación requiere de su mantenimiento en pureza (HALL & BRADLEY 1995). Según datos de FAO (SCHERF 2000) la tercera parte de las 6400 razas domésticas documentadas se encuentran en peligro, extinguiéndose a una tasa del 1-2% por año.

La diversidad genética de las razas locales puede permitir desarrollar nuevas características en respuesta a cambios en el ambiente, enfermedades o las condiciones del mercado. Además, muchas razas locales poseen combinaciones genéticas y adaptaciones especiales a ambientes extremos (elevadas temperaturas, altitud, exposición a enfermedades, etc.) que no se encuentran en otras razas (MAUDET et al. 2002; SIMIANER 2005).

Existe un gran número de razas domésticas que han sido cruzadas con razas productivas (a fin de incrementar su valor económico) de modo que el número de razas locales disminuye (UGARTE et al. 2001). Sin embargo, las razas cruzadas no siempre están tan adaptadas como las originales. Caracteres como resistencia a enfermedades, forrajes pobres, etc. pueden perderse fácilmente, poniendo de manifiesto la importancia de mantener las razas locales en pureza o recuperarlas en caso de haber sido mezcladas (MORAIS et al. 2005; BARILLET 2007; TABERLET et al. 2008).

Poblaciones naturales

La percepción como positiva o negativa de la introgresión de genes en poblaciones naturales depende en gran medida del contexto de cada situación

particular. No obstante, desde el punto de vista de la conservación de la biodiversidad, en la mayoría de los ejemplos en los que se describe introgresión, este aporte de material genético externo se considera perjudicial. De este modo, la hibridación introgresiva aparece como una de las preocupaciones fundamentales con respecto a la pérdida de biodiversidad (RHYMER & SIMBERLOFF 1996; SUTHERLAND et al. 2006; RANDI 2008).

Las tasas de hibridación e introgresión (tradicionalmente considerados como eventos poco comunes) parecen mostrar en varios estudios una mayor frecuencia de la esperada (RHYMER & SIMBERLOFF 1996; ALLENDORF et al. 2001). El aumento del interés en estos aspectos y sus consecuencias se ve reflejado en el número de publicaciones relacionadas con el tema en los últimos años, como se muestra en la Figura I.1.

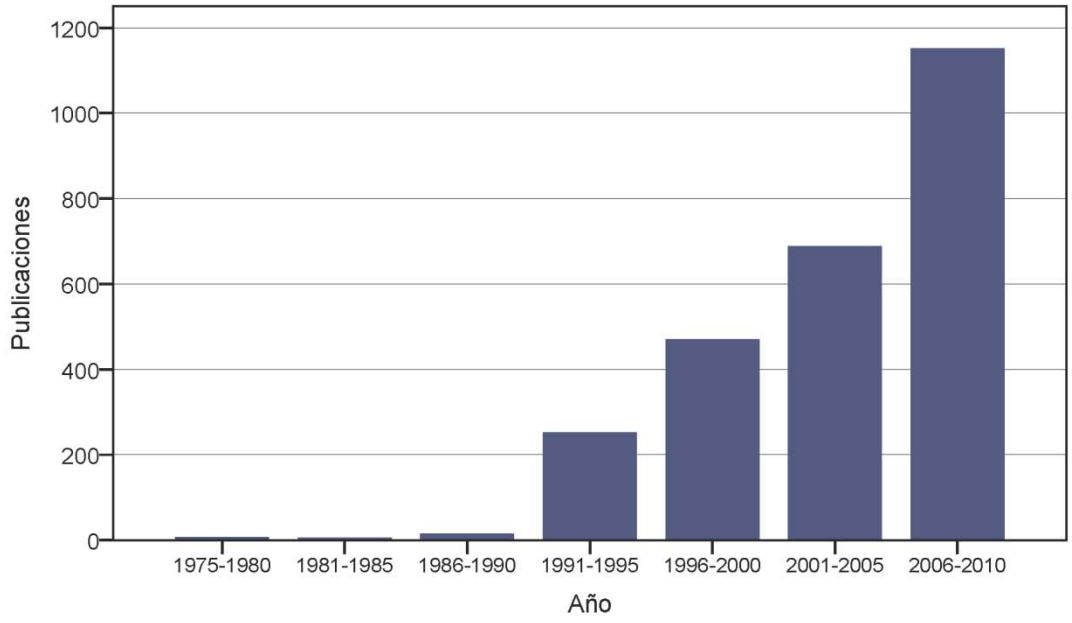

FIGURA I.1. Número de publicaciones por año que responden a la búsqueda en la Web of Knowledge del tema "Hybridization and introgression"[1].

Gran parte del incremento de procesos de hibridación se debe a causas antropogénicas. El impacto de las actividades humanas se observa en todos los niveles de la biodiversidad y parece ser responsable de este aumento,

[1] Modificado a partir de LARGIADÈR (2006)

convirtiéndose en un problema con graves implicaciones en la conservación de las especies involucradas, especialmente en aves, mamíferos y peces (VITOUSEK et al. 1997; LARGIADÈR 2006).

Las modificaciones en los paisajes son uno de los factores mediante los que la acción humana puede afectar a estos procesos. La alteración de los hábitats conlleva una modificación del flujo genético natural, que puede incrementarse entre especies que antes se encontraban separadas. Dicha modificación puede implicar un aumento en la capacidad invasiva de las especies a raíz de un aumento de su eficacia. También un movimiento directo de los individuos de distintas especies entre varias localizaciones, tanto intencional como accidentalmente, podrá conducir a hibridaciones e intercambios no deseados debidos a la intervención humana (CRISPO et al. 2011). Todas las acciones expuestas tienen graves consecuencias sobre la conservación de la biodiversidad ya que promueven la introgresión con los problemas que llevan asociados.

Existen muchos ejemplos de especies salvajes amenazadas por equivalentes domésticos o especies introducidas de manera artificial (RANDI 2008). De este modo, un nuevo influjo de material genético procedente de otra raza o subespecie puede amenazar a la población causando una *extinción genómica*. Dicha *extinción* no implica necesariamente la pérdida de alelos, sino que puede manifestarse como la pérdida de combinaciones de alelos en diferentes loci (haplotipos), que suelen ser la base de las adaptaciones locales. Así, linajes evolutivos completos podrían desaparecer (ALLENDORF & LUIKART 2007). Este proceso no tiene necesariamente que estar ligado a una mayor eficacia biológica de los individuos exógenos o la descendencia híbrida, sino que la polución genética que ocurre simplemente por deriva, puede provocar la desaparición de la conformación genética nativa.

Podemos ver varios ejemplos de introgresión entre distintos taxones en la Tabla I.1.

Otro motivo por el que los cruces en poblaciones naturales pueden ser perjudiciales es el fenómeno denominado *depresión híbrida*. En ocasiones el cruce entre poblaciones no produce un incremento en la eficacia, sino un descenso de la misma.

A NIVEL DE GÉNERO	
Vacuno doméstico (*Bos taurus*) y bisonte (*Bison bison*)	(FREESE et al. 2007)
A NIVEL DE ESPECIE	
Cangrejos de río (*Orconectes rusticus* y *O. propinquus*)	(PERRY et al. 2001)
Trucha arcoíris (*Oncorhynchus mykiss*) y degollada (*O. clarkii*)	(HOHENLOHE et al. 2011)
Salamandra de espalda roja (*Plethodon cinereus*) y de las *Big Levels* (*P. sherando)*	(BAYER et al. 2012)
Sapo de la costa del golfo (*Bufo nebulifer*) y de Fowler (*B. fowleri*)	(VOGEL & JOHNSON 2008)
Rana verde común (*Rana ridibunda*) y centroeuropea (*R. lessonae*)	(VORBURGER & REYER 2003)
Ánade real (*Anas playrhynchos*) y pato negro del pacífico (*A. superciliosa*)	(RHYMER et al. 1994)
Perdiz griega (*Alectoris graeca*) y roja (*A. rufa*)	(NEGRO et al. 2001)
Perdiz de Chukar (*Alectoris chukar*) y roja (*A. rufa*)	(BARBANERA et al. 2011)
Codorniz doméstica (*Coturnix japónica*) y salvaje (*C. coturnix*)	(BARILANI et al. 2005)
Liebre común (Lepus europaeus) y de montaña (L. timidus)	(REID 2011)
Turón (*Mustela putorius*) y visón europeo (*M. lutreola*)	(CABRIA et al. 2011)
Coyote (*Canis latrans*) y lobo rojo (*C. rufus*)	(MILLER et al. 2003)
Perro doméstico (*Canis lupus familiaris*) y lobo etíope (*C. simiensis*)	(GOTTELLI et al. 1994)
Vacuno doméstico (*Bos taurus*) y cebú (*B. indicus*)	(MACHUGH et al. 1997)
A NIVEL DE SUBESPECIE	
Gato doméstico (*Felis silvestris catus*) y gato silvestre (*F. silvestris* spp.)	(DRISCOLL et al. 2011)
Perro doméstico (*Canis lupus familiaris*) y dingo (*C. lupus dingo*)	(DANIELS & CORBETT 2003)
Jabalí europeo (*Sus scrofa scrofa*) y jabalí de Maremma (*S. scrofa majori*)	(VERNESI et al. 2003)
A NIVEL DE POBLACIÓN/RAZA	
Tímalo cultivado y salvaje (Adriático) (*Thymallus thymallus*)	(SUŠNIK et al. 2004)
Salmón común cultivado y salvaje (*Salmo salar*)	(HINDAR et al. 2006)
Visón americano doméstico y salvaje (*Neovison vison*)	(KIDD et al. 2009)
Zorro ártico doméstico y salvaje (*Alopex lagopus*)	(NORÉN et al. 2005)
Vacuno doméstico Charolais y Blanca Cacereña (*Bos taurus*)	(PADILLA et al. 2009)

Tabla I.1. Ejemplos de introgresión (población exógena y población amenazada) clasificados en función de los taxones implicados.

La reducción de la eficacia en los híbridos F1 puede deberse a incompatibilidades genéticas entre los taxones que hibridan, pérdida de las adaptaciones locales al ambiente, sobredominancia negativa o interacciones epistáticas. A menudo, esta reducción de la eficacia se retrasa a la F2 o generaciones posteriores cuando se ponen de manifiesto interacciones deletéreas entre loci homocigotos. Varios ejemplos de depresión híbrida han sido descritos en plantas, invertebrados y vertebrados (ALLENDORF & LUIKART 2007; EDMANDS 2007).

Poblaciones en cautividad

Cuando se trabaja con poblaciones en cautividad, tanto de especies salvajes como domésticas, la casuística de la introgresión no deseada presenta ciertas particularidades.

En primer lugar, la ventaja de las poblaciones en cautividad es que su gestión puede ser más intensa, controlando qué individuos generan descendencia y con quién se aparean. Incluso en esas condiciones, un manejo incorrecto de la población (especialmente en especies ganaderas) puede hacer que por descuido o intencionadamente, los animales no se mantengan en pureza y se apareen con individuos de otras poblaciones o razas.

El caso del rescate genético citado anteriormente podría ser un ejemplo de introgresión intencionada y conocida. Una vez que la población rescatada ha alcanzado la suficiente eficacia biológica (considerándose fuera de riesgo), podría ser interesante tratar de recuperar, al menos en parte, el genoma original ligado a sus características específicas.

También podríamos encontrarnos en circunstancias similares cuando se trata de reconstruir (completa o parcialmente) una raza extinta mediante recursos criogenéticos, como un banco de semen, empleando hembras de otra raza o subespecie (GANDINI & OLDENBROEK 2007). El fondo genético original en este caso sería el 50% del genoma procedente del banco de semen, mientras que el 50% restante, que provendría de las hembras exógenas, sería indeseado y debería eliminarse para completar la recuperación de la población original.

ESTRATEGIAS DE DEPURACIÓN

Una vez que se es consciente de que un proceso de introgresión ha tenido lugar y se pretende revertirlo, debemos aplicar estrategias de manejo que permitan eliminar la información genética exógena. Para ello, la idea general es detectar aquellos individuos portadores de variantes genéticas foráneas y restringir (o impedir) que tengan descendencia. Del mismo modo, favoreceremos la reproducción de los individuos con un porcentaje de información nativa alto, con lo que ésta aumentará progresivamente en la población. La identificación de los individuos deberá realizarse en base a la información disponible.

Genealogías

El conocimiento de las genealogías es muy común en animales de especies ganaderas, especialmente en aquellas en las que se ha implementado selección artificial durante generaciones (HALEY 2009). También es común la existencia de un libro de registros en algunas poblaciones en cautividad, en zoos o reservas (FRANKHAM et al. 2002). Sin embargo, la existencia de pedigrís en especies salvajes es inusual y muchas veces están incompletos o son incorrectos.

A través de la genealogía se puede calcular la probabilidad de haber recibido información genética de un fundador exógeno y, por tanto, el porcentaje promedio de genoma de origen no deseado. Dicha probabilidad puede calcularse vía el parentesco global o mediante el parentesco parcial debido a un grupo de individuos.

El caballo de *Przewalski* (*Equus ferus przewalskii)* es un ejemplo de raza en peligro de extinción debido al escaso número de ejemplares de la especie, con un pedigrí controlado y registrado desde el año 1899. En 1906 una hembra de caballo doméstico (*Equus ferus caballus*) se cruzó con un caballo de *Przewalski*. Sus genes perduran en la población y pueden ser rastreados gracias a la genealogía.

Para que el pedigrí nos permita detectar y eliminar introgresión, éste ha de encontrarse disponible y perfectamente registrado. En caso contrario debería disponerse de información adicional para su reconstrucción, por ejemplo a través de información molecular.

Marcadores moleculares

Analizando los alelos presentes en las dos poblaciones involucradas (nativa y exógena introducida) para un conjunto de marcadores moleculares, podremos obtener información que nos ayude a reconocer los individuos con genes exógenos. Del mismo modo que los marcadores moleculares se han empleado para detectar la existencia de introgresión y otras situaciones como aislamiento genético, consanguinidad o subdivisión (GROENEVELD et al. 2010), parece lógico que puedan ser usados en el proceso de depuración genética. En función del tipo de marcadores moleculares que tengamos disponibles, podremos utilizar la información que nos proporcionen bajo diferentes enfoques.

Tipos de marcadores

Marcadores multialélicos. Los marcadores multialélicos, en particular los del tipo microsatélite, han sido ampliamente utilizados en gran variedad de estudios tanto en animales domésticos como salvajes, debido a características como la codominancia, facilidad de genotipado y alta diversidad (VIGNAL et al. 2002). Aunque en la actualidad han sido desplazados por otros marcadores, aún se usan con frecuencia en poblaciones salvajes, en las que el desarrollo de SNP (polimorfismos de un solo nucleótido, *single nucleotide polymorphism*) aún no está tan extendido (SLATE et al. 2009). Existen paneles estándar con unos 30 microsatélites para un gran número de especies domésticas elaborados por FAO (1998) que se han empleado para identificación de individuos, análisis de paternidades, evaluación de la diversidad, comparación de razas, etc. (LENSTRA et al. 2012).

Genotipado masivo. En los últimos años el uso de marcadores de tipo SNP se ha convertido en la herramienta de referencia para caracterizar la variación genética de las poblaciones con varios fines: mejorar caracteres productivos mediante selección basada en evaluaciones genómicas, establecer prioridades en esquemas de conservación, trazabilidad de productos de calidad, estimas de diversidad genética, etc. Los SNP son abundantes y se distribuyen por todo el genoma (tanto en regiones codificantes como en no codificantes), son bialélicos y codominantes y su evolución puede describirse con modelos mutacionales sencillos (VIGNAL et al. 2002; MORIN et al. 2004). Los avances tecnológicos de los últimos años han

permitido que el número de marcadores disponibles alcance valores de hasta 770000 SNP como en el caso del ganado bovino (LENSTRA et al. 2012). Una información molecular tan densa permite en muchos casos reemplazar las genealogías aumentando la precisión en la estima de parámetros genéticos (HAYES et al. 2009) y en el mantenimiento de diversidad genética (DE CARA et al. 2011).

Haplotipos. La búsqueda de haplotipos (combinaciones de alelos en diferentes loci transmitidos conjuntamente) puede emplearse para agrupar los marcadores proporcionando otros métodos de análisis. Entre marcadores muy próximos el ligamiento es muy fuerte, de modo que no son independientes. Esto puede implicar información redundante, y el elevado número de marcadores (en sets muy densos) no se aprovecha al máximo. Combinando los marcadores en haplotipos se añade información adicional, en especial en lo que respecta a los fenómenos de recombinación pasados, pudiendo resultar una estrategia más potente que el uso de los marcadores de manera independiente (GATTEPAILLE & JAKOBSSON 2012).

Estrategias de manejo con marcadores moleculares

Cuando se emplea la información de marcadores moleculares con el fin de eliminar introgresión no deseada se pueden utilizar dos aproximaciones: i) basarse en la existencia de variantes alélicas propias (exclusivas) de las poblaciones nativa o exógena; ii) calcular medidas cuantitativas de diferenciación, para determinar si los individuos de la mezcla son más o menos parecidos a una población u otra.

Alelos indicativos de población. Independientemente del tipo (microsatélite o SNP), un marcador diagnóstico presenta alelos que originalmente se encuentran únicamente en individuos nativos o en individuos exógenos (lo que se conoce como alelos privados). La presencia de alelos exclusivos de individuos exógenos puede servir de indicador de la necesidad de penalizar la reproducción de su portador (o viceversa: portar alelos exclusivos de individuos nativos es motivo para ser favorecido en la reproducción). El número de marcadores diagnóstico es un factor limitante en su eficacia. Aunque se han encontrado ejemplos de marcadores diagnóstico a nivel de especie y subespecie (ROY et al. 1994; MACHUGH et al. 1997), no son muy frecuentes en poblaciones estrechamente relacionadas (VILÀ et al. 2003) como suelen ser las involucradas en fenómenos de introgresión.

En ocasiones, marcadores identificados como diagnóstico pueden no serlo, a raíz de errores en el genotipado o por un muestreo reducido. Por tanto todos los alelos, aunque no hayan sido detectados, están presentes en ambas poblaciones a diferentes frecuencias. El uso como exclusivo de alelos provenientes de marcadores no diagnóstico implicará una pérdida de eficiencia en la depuración que será dependiente de las frecuencias alélicas reales en las poblaciones puras. Cuanto más intermedias sean y, por tanto, más alejadas de la exclusividad, más ineficaces resultarán.

Medidas de parecido/diferenciación entre poblaciones. Existen varias medidas de diferenciación poblacional que podrían emplearse para minimizar el parecido con la población exógena y, consecuentemente, eliminar la introgresión no deseada.

Distancias genéticas. La diferencia en frecuencias alélicas entre poblaciones puede utilizarse para calcular distancias genéticas tales como la de Nei, Kullback-Leibler, etc. La gestión estará entonces dirigida a minimizar la distancia entre la población problema y la nativa.

Parentesco molecular. Los marcadores pueden emplearse para estimar un parentesco realizado entre los individuos de la población actual y los individuos exógenos (o los nativos). De forma análoga a como se actúa en el caso de las genealogías, la estrategia será detectar aquellos individuos que hayan recibido mayor proporción de información exógena. La precisión y por tanto la eficacia del método dependerá del número de marcadores (DE CARA et al. 2011).

Matriz de relaciones genómica. El cálculo de una matriz de relaciones genómica (*Genomic Relationship Matrix*, GRM) permite estimar las relaciones entre los individuos y por tanto estimar su grado de pureza. La GRM podrá utilizarse en un análisis de componentes principales así como en un modelo lineal para predecir la variable *raza nativa* y disponer de la proporción de genoma nativo en los individuos.

Origen de segmentos cromosómicos. Si calculamos la frecuencia con la que segmentos cromosómicos (o haplotipos) aparecen en las poblaciones nativa y exógena se puede estimar la proporción de genoma nativo y exógeno en la población problema (y en cada uno de los individuos particulares) según los haplotipos que presente cada uno de los individuos.

EVALUACIÓN DE LA EFICACIA DEL PROCESO

A la hora de evaluar la eficacia del proceso de desintrogresión llevado a cabo, la variable de mayor importancia es la proporción de alelos nativos en la población final. No obstante, la recuperación de la máxima cantidad de genoma original es el primer objetivo del proceso de desintrogresión. En los trabajos aquí presentados, dicho valor puede calcularse puesto que se trata de simulaciones, en las cuales podemos registrar el origen de los alelos, de modo que en un individuo cualquiera puede calcularse qué proporción de los mismos procede de un fundador exógeno y qué proporción procede de los fundadores nativos. En el caso de poblaciones reales, dicho valor puede ser una estima (genealógica o molecular) pero el porcentaje real será desconocido.

Además de la proporción de genoma nativo recuperada, es interesante monitorizar otros parámetros para evaluar cómo se ve afectada la estructura de la población. El proceso de desintrogresión implica una selección artificial de individuos y como tal lleva asociada una pérdida de variabilidad e incremento de la consanguinidad. Para evaluar dicho efecto, pueden monitorizarse variables como la heterocigosidad esperada en todo el genoma o la consanguinidad que se genera, tanto la genealógica como la molecular (heterocigosidad observada). Otros parámetros demográficos pueden ser interesantes, como el número de individuos reproductores o la varianza de las contribuciones. Estas variables serán útiles a la hora de indicar si el proceso conlleva un deterioro genético excesivo de la población. En última instancia habrá que buscar una solución de compromiso para poder llevar a cabo la depuración de la población sin una pérdida elevada de la diversidad.

OBJETIVOS

OBJETIVOS GENERALES

1. Analizar las distintas situaciones en que puede ocurrir un fenómeno de introgresión no deseada.

2. Evaluar las posibilidades de actuación para revertir introgresión no deseada en función de las características de la población y de la información disponible.

3. Evaluar las consecuencias de revertir introgresión no deseada de la población sobre parámetros de diversidad genética.

OBJETIVOS ESPECÍFICOS DE CAPÍTULO

Capítulo 1.

1.1. Evaluar, mediante simulación por ordenador, la capacidad de recuperación de una población que ha sufrido introgresión no deseada utilizando la información de genealogías.

1.2. Analizar, como ejemplo real, las consecuencias sobre la pureza de la población de la introducción en 1906 de una yegua doméstica en la población del caballo de Przewalski.

Capítulo 2.

2.1. Evaluar, mediante simulación por ordenador, la capacidad de recuperación de una población que ha sufrido introgresión no deseada utilizando la información de marcadores moleculares no densos.

2.1.1. Evaluar la capacidad de desintrogresión mediante marcadores con alelos exclusivos de población.

2.1.2. Evaluar la capacidad de desintrogresión mediante marcadores genéricos (sin alelos privados).

Capítulo 3.

3.1. Evaluar, mediante simulación por ordenador, la capacidad de recuperación de una población que ha sufrido introgresión no deseada minimizando el parentesco molecular con los individuos exógenos, calculado a partir de datos de genotipado masivo.

Capítulo 4.

4.1. Analizar las diferencias genómicas entre dos razas de ovino (Merino y Poll Dorset) a partir de la información de genotipado masivo mediante dos métodos: búsqueda de haplotipos específicos de raza y predicción de la raza a través de un modelo mixto (GBLUP).

4.2. Evaluar, mediante simulación por ordenador, la capacidad de recuperación de una población que ha sufrido introgresión no deseada a través de las dos aproximaciones anteriores.

CAPÍTULO 1

Removing exogenous information using pedigree data

INTRODUCTION

The benefits of interchanging genetic resources between populations have been deeply studied. New inputs of genetic material are usually considered a good way to protect species and biological diversity from threats such as the loss of ability to adapt to new environments and the decrease in fitness (FRANKHAM et al. 2002). However, gene flow can lead to an undesired introduction of genetic material into the population. Apart from the problems which could arise from outbreeding depression (ALLENDORF & LUIKART 2007) sometimes there is a need of maintaining the population genetic background pure. In that case, when an undesired introduction of genetic material happens, the original genetic conformation would have to be recovered.

There are numerous examples where keeping the original background would be interesting. In the field of domestic animals, a particular breed could be linked to a quality product with economical interest. Many breeds provide differentiated products directly related to the specific genetic information of the breed: *Iberian* pigs and the Jamón Ibérico, *Reggiana* dairy cows and the Parmigiano Reggiano cheese and some other examples pointed by DALVIT et al. (2007). There are also specific breeds connected to a particular activity, as the Spanish and other purebred horses which are involved in sport competitions and other entertainment events with high economic benefits (MAPA 2003). Just for aesthetical reasons dog breeds are kept separated. A dog can become an official member of a breed only if both parents belong to it, being economically devaluated when it is a mixed-breed (PARKER et al. 2004). The loss of value of crossed individuals is also common in other species.

From a conservationist point of view, gene flow can also be harmful for some animal and plant species or populations due to the replacement of native populations by invaders. Some examples of admixture and introgression have been described in birds (quail, BARILANI et al. 2005; partridges, NEGRO et al. 2001), fishes (grayling, SUŠNIK et al. 2004; trout, BOYER et al. 2008), bovine (cattle, PADILLA et al. 2009; bison, FREESE et al. 2007, HALBERT & DERR 2007) and carnivores (wolves, MILLER et al. 2003; cats, BEAUMONT et al. 2001).

Livestock breeds are considered important components of biodiversity, and its conservation is based on pure-breeding (HALL & BRADLEY 1995). Human activities contribute to increase rates of hybridization and introgression. The relevance of this problem has been underestimated, leading in the worst case scenario to populations going to extinction. This phenomenon is happening more frequently than expected (RHYMER & SIMBERLOFF 1996; ALLENDORF et al. 2001). In farm animals, microsatellite markers have revealed the occurrence of introgression (GROENEVELD et al. 2010).

Another scenario where undesired introgression occurs is when facing the task of reconstructing a completely or partially extinct breed by using cryoconserved semen on females from another breed or species (GANDINI & OLDENBROEK 2007). The 50% of the genome coming from the semen bank would be from the original background, but the other 50% coming from the females would be undesired and would have to be eliminated to achieve the complete recovery of the original species.

A good example of undesired introgression (in this case due to an incorrect management of the population) is the Przewalski's horse (BOUMAN & BOUMAN 1994). Przewalski's horse (*Equus przewalskii*) is an endangered species that was almost extinct after the Second World War. It was described in 1881 and its entire genealogy has been kept since 1899. In the Zoological Department of the Agricultural Institute of the University of Halle in Germany, counselor J. Kühn crossed a Przewalski's horse with a Mongolian domestic horse producing one colt (1906) that had more descendants with another Przewalski's. Domestic horse alleles introgressed the pure population and they still remain in the present Przewalski's horses. As the entire genealogy is kept and available (VOLF 1994) the expected percentage of foreign or pure genome of each individual can be calculated.

When the introgression process takes place, the crossed individuals can be detected or not by visual signs. When the admixture is not reflected morphologically, undesired information could be removed using recorded pedigree information, if available, or using markers that allow detecting the alleles coming from different populations.

The objective of the present study is to explore the efficiency of different methods to remove the exogenous genetic information from an introgressed population using exclusively pedigree information.

MATERIALS AND METHODS

Simulated data

A population with size (*N*) of 10 individuals (5 males and 5 females), kept constant over discrete generations was simulated. The pedigree was recorded all generations since the beginning, and used during the management. The genome of each individual was made up of 20 chromosomes of 1M each. A total of 100 multiallelic loci were simulated per chromosome. A Poisson distributed number ($\lambda = 1$) of crossing-overs with no interference were generated in random positions of each chromosome when creating the offspring.

In the initial generation individuals were not inbred and unrelated. They carried two different alleles at each locus (2*N* different alleles per locus in the base population) and, thus, loci were completely informative. This molecular information was used to calculate different parameters at the end of the management period to evaluate the efficiency of the methods.

Different scenarios with different degrees of introgression were simulated by varying the following factors:

Number of exogenous individuals. The percentage of introgression ranged from 10% to 50% by including 1 to 5 exogenous individuals (sex randomly set) as part of the base population. Native and exogenous individuals carried different alleles in all the 100 loci.

Number of generations without management. A number (1 to 5) of initial unmanaged generations (random contributions and mating) were simulated to obtain the admixture of the foreign information that was set in the base population and to create some relatedness among individuals.

Management

After the initial unmanaged phase, 10 generations of management were performed. Four different strategies were carried out in order to determinate

contributions of individuals to the next generation (i.e., offspring generated by each potential parent):

1. Random (R): The individuals mated randomly and parents were not selected.

2. Optimum contributions (OC): Optimal contributions from parents were calculated by minimising the global coancestry weighted by those contributions,

[1.1] $$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j f_{ij} \ ,$$

where c_i is the relative contribution of individual i to the next generation, and f_{ij} is the coancestry between individuals i and j (CABALLERO & TORO 2000). Strategy R corresponds to unmanaged populations and, thus, it provides the lower bound for the de-introgression process. Contrarily, OC (the standard management procedure recommended for conservation programmes) is directed to control the loss of genetic diversity and the rise of inbreeding. Therefore, both methods can be used as reference points for evaluating the performance of other strategies.

3. Minimum exogenous contribution (MEC): The contributions were obtained by minimising the total amount of information which came from the exogenous founders,

[1.2] $$\sum_{i=1}^{N} c_i f_{Ex,i} \ ,$$

where $f_{Ex,i}$ is the coancestry between individual i and all the exogenous founders.

4. Minimum partial coancestry (MPC): Contributions were calculated by minimizing the mean partial coancestry,

[1.3] $$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j f^{*}_{ij} \ ,$$

where f^{*}_{ij} is the partial coancestry between i and j. The partial coancestry represents the kinship between two individuals due to one specific ancestor, or, in other terms, it is the probability that an offspring from that couple is homozygous for an allele descending from a specific ancestor. Sum of partial coancestries due to each of the founders is the total coancestry. Partial coancestry can be determined by using a modification of the tabular method (LACY et al. 1996; LACY 1997). Under MPC strategy, partial coancestries included in the objective function were those due to the foreigner founders. Therefore, the sum across all the exogenous

founders provided the value for partial coancestry used to compare individuals and to decide their contributions.

MEC and MPC methods were also implemented in all scenarios adding a restriction on the maximum rate of inbreeding (ΔF). Three different restrictions were imposed: 5%, 10% and 15%. Also extra simulations were carried out with a population size of 100 individuals with 10, 20, 30, 40 or 50 exogenous individuals in the base population, followed by up to 5 generations without management and 10 generations under the four treatments.

All the optimizations were performed using *simulated annealing* algorithms (KIRKPATRICK et al. 1983; FERNÁNDEZ & TORO 1999). Once the optimum contributions were obtained, minimum coancestry matings were performed in all scenarios, except for the random management, by implementing the *Hungarian* algorithm (DANTZIG 1963). For each scenario and management method 20 replicates were simulated and results presented are averages over replicates.

Variables

In each generation several variables were calculated to evaluate the efficiency of the different strategies: (1) non-exogenous founder representation, calculated from genealogy, (2) non-exogenous founder representation, calculated from genomic molecular information, (3) average inbreeding coefficient, (4) mean coancestry, (5) mean partial coancestry and (6) observed homozygosity. Values for 3, 4 and 5 were calculated from pedigree.

Real data

The conditions of the captive breeding program of the Przewalski's horse have allowed keeping the complete pedigree of the horse since 1899, which is available in http://przwhorse.pikeelectronic.com/. The genealogy was analysed and the descendants of the Mongolian domestic mare introduced in 1906 were detected. All individuals not descending from the foreign introgressed horse were identified and the relationship between them and the whole population was studied from 1935 to 2009 to evaluate the potential of de-introgression at different times. The consequences of removing carriers of exogenous alleles, both on the levels of genetic diversity and inbreeding coefficient, were also evaluated.

RESULTS

Simulated data

Non-exogenous founder genealogical representation

Results for non-exogenous founder representation are shown in Figure 1.1. As expected, the R and OC strategies did not eliminate any exogenous representation but kept the values constant irrespective of the number of generations elapsed before management started, as only drift is affecting the frequency of foreign alleles. Note that native information observed at generation 10 under these strategies is near 90%, 80%, 70%, 60% and 50% of that present in the base population for 1, 2, 3, 4 and 5 exogenous founders, respectively.

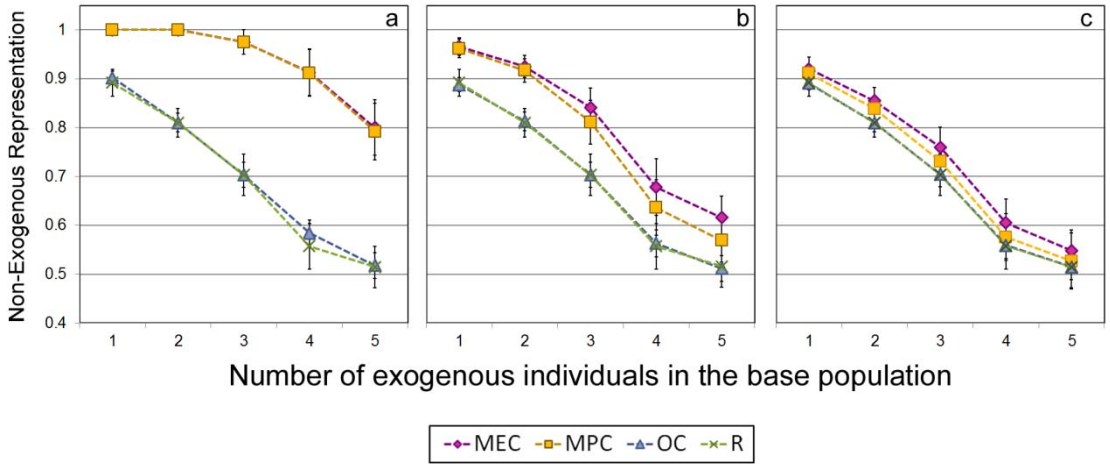

FIGURE. 1.1. Non-Exogenous Representation in the last generation of management with each of the four methods, according to the number of exogenous individuals. (a) One previous generation with no management, (b) Three previous non-managed generations, (c) Five previous non-managed generations.

The most effective method for removing undesired introgression was always MEC (Figure 1.1). Maximum values of non-exogenous founder representation were always obtained when using this strategy.

In some cases, MPC could reach the efficiency of MEC, particularly when the numbers of foreigners and non-managed generations were low. Differences between both methods became larger as these parameters increased. When the percentage of introgression was high or many generations elapsed till the

management, the chance for introgressed information being removed became too low irrespective of the strategy used (see right part of Figure 1.1c). Only in the scenarios with a small number of foreigners and a low number of generations of admixture, the original genomic information will be completely recovered.

It must be pointed out that the final value of the non-exogenous founder representation was reached in the first generation of management, and did not change afterwards under both MEC and MPC strategies (data not shown). This fact indicates that it would be enough one generation of management under MEC to obtain the best result.

Results about foreign representation calculated from the molecular information of each individual's genome follow the same pattern as the observed for the genealogical information in all cases (data not shown). This happened because the simulated *loci* are neutral, unlinked and completely informative in the base population, and thus genealogical and molecular coancestries are equivalent.

Inbreeding coefficient

The performance of both inbreeding and mean coancestry was similar. Consequently only the evolution of the inbreeding coefficient (F) is presented in Figure 1.2. The minimum inbreeding coefficient was obtained, as expected, with the OC method. MEC always led to the maximum values of inbreeding and mean coancestry even above the R method.

Under the R and OC strategies, the inbreeding levels reached were independent of the number of exogenous individuals in the base population but increased with the number of unmanaged initial generations. Contrarily, the number of exogenous founders affected the levels of inbreeding in MEC and MPC.

MPC led to similar results as MEC in scenarios with a low number of unmanaged generations and/or with a low number of exogenous founders (i.e., with little admixture). As the number of exogenous individuals increased the results of MPC differed from those of MEC and became more similar to those of OC. This performance is due to the fact that the greater is the number of exogenous, more founders are to be taken into account when calculating the partial coancestry and, therefore, approaching the global coancestry.

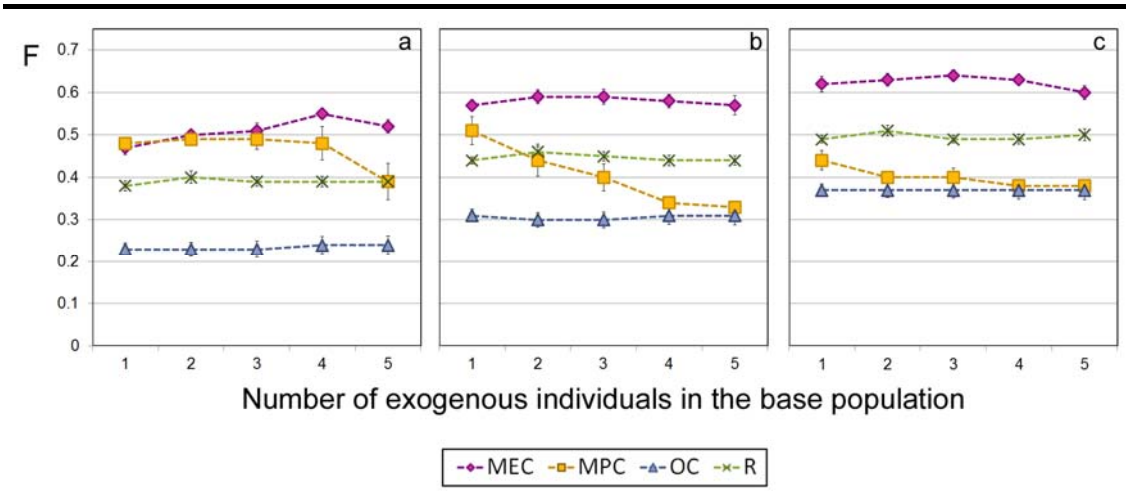

FIGURE 1.2. Inbreeding coefficient in the last generation of management with each of the four methods, according to the number of exogenous individuals. (a) One previous generation with no management, (b) Three previous non-managed generations, (c) Five previous non-managed generations.

The unpleasant performance of MEC regarding the levels of inbreeding or coancestry is a consequence of the importance that this strategy gives, by definition, to the elimination of foreign information, ignoring global genetic diversity. Thus, all individuals with the same percentage of their genomes coming from exogenous founders are equivalent, and MEC assigns the same value to solutions where they contribute equally, differentially or even when some do not contribute at all. Moreover, when foreign information is evenly distributed among individuals (i.e., exogenous representation is equal in all individuals) MEC turns into Random management, as the method lacks a criterion to prioritise individuals.

To alleviate this effect a modification of the method was implemented. It consisted in selecting from all the solutions with the same remaining proportion of exogenous information the one with the maximum number of individuals contributing to the next generation. This improved method yielded slightly lower levels of inbreeding, although still greater than the other methods (data not shown).

Another trend in the performance of the inbreeding coefficient under MEC could be observed. Cases with an intermediate number of foreigners showed larger values of *F* and mean coancestry. This could be explained by the fact that, when

there was little introgression, most available individuals to be used as parents for the next generation are completely free of exogenous influence and, thus, it is not necessary to reduce the number of used parents to perform the de-introgression. In cases with many exogenous founders, most individuals have undesired introgression, but they are almost equivalent and, consequently, all used as parents (there is no removal, but the genetic diversity is maintained). However, with an intermediate number of foreign founders just a few *pure* individuals (i.e., without exogenous ancestors) remain to be used as parents and, similar to a bottleneck effect, the mean inbreeding and coancestry increase.

When simulations using MEC and MPC included restrictions on the increase of inbreeding (5%, 10% and 15% of rate of inbreeding per generation) results were very similar to those obtained under OC method irrespective of the restriction imposed. While coancestry and inbreeding were kept low, there was no removal of exogenous representation in any of the performed scenarios (data not shown).

Partial coancestry

Results of the evolution of partial coancestry are shown in Figure 1.3. As expected, the minimum value of partial coancestry was always obtained by minimising the partial coancestry (MPC), since this strategy was developed to do so. Populations under the R strategy yielded a partial coancestry value which increased with the number of exogenous individuals, and it was always larger than that with OC. As it happened with inbreeding, MPC became similar to OC when the number of foreigners (and also the number of unmanaged generations) was large. This is due to partial coancestry representing a part of the kinship that it is greater the larger is the number of foreigners founding the population.

When the number of unmanaged generations or external individuals was small, MEC led to a lower partial coancestry than R and OC, and to similar partial coancestry than MPC. As soon as the number of unmanaged generations or exogenous founders increased, the efficiency of MEC in keeping low levels of partial coancestry decreased. That could be explained by the fact that the individuals, despite they had lower exogenous percentage of information, were more related through the remaining foreign lineages. It has to be taken into

account that the minimisation of the partial coancestry is not a specific objective of the MEC method.

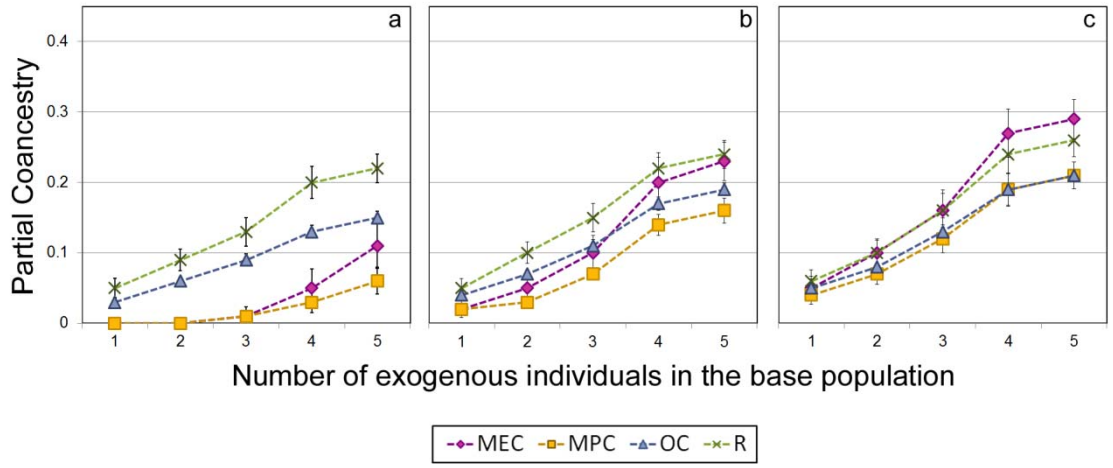

FIGURE 1.3. Partial coancestry in the last generation of management with each of the four methods, according to the number of exogenous individuals. (a) One previous generation with no management, (b) Three previous non-managed generations, (c) Five previous non-managed generations.

Observed homozygosity. The performance of the observed homozygosity calculated from the simulated loci in all methods is equivalent to the pedigree inbreeding coefficient. Its value is not genealogical but represents the realisation at the genomic level, and represents the identity by descent because the 2000 loci of each individual are neutral and completely informative in the base population.

Large population

Table 1.1 shows the non-exogenous founder genealogical representation and the inbreeding coefficient achieved after ten generations of management for MEC and MPC strategies, N = 10 and N = 100 individuals, and a level of introgression from 10% to 50%. Results for R and OC were similar to those obtained under these strategies in the 10 individuals population, with no de-introgression and F levels only dependent on the number of unmanaged generations (data not shown). Comparisons between the two population sizes showed that the efficiency of both methods increased with a larger number of individuals for the same percentage of introgression. In all cases lower levels of exogenous information were achieved in the 100 individuals population, and, of course, with a lower increase of inbreeding.

The MEC strategy proved to be better than the MPC strategy especially with a large degree of introgression. MPC performed similarly to MEC when the percentage of exogenous founders was 10% (with similar values of native representation and *F*) but it got worse with more than 20% of exogenous individuals.

NON-EXOGENOUS REPRESENTATION						
		Level of introgression				
	N	**10%**	**20%**	**30%**	**40%**	**50%**
MEC	**10**	0.920	0.855	0.760	0.605	0.548
	100	0.990	0.921	0.876	0.766	0.676
MPC	**10**	0.911	0.839	0.731	0.576	0.527
	100	0.984	0.864	0.784	0.649	0.537
F						
		Level of introgression				
	N	**10%**	**20%**	**30%**	**40%**	**50%**
MEC	**10**	0.615	0.634	0.644	0.632	0.599
	100	0.167	0.240	0.264	0.258	0.271
MPC	**10**	0.435	0.401	0.399	0.379	0.376
	100	0.159	0.066	0.051	0.043	0.041

TABLE 1.1. Comparison of non-exogenous representation and inbreeding coefficient after ten generations of management (under MEC and MPC) on populations of size *N* = 10 and *N* = 100 individuals, with 10 to 50% of introgression and 5 generations of admixture.

As in the scenario with *N* = 10, the whole effect of MEC on the removal of foreign information in populations with 100 individuals was achieved in the first generation of management. This did not happen when managing with MPC where several generations where needed to remove the undesired information, inducing, thus, an increase in inbreeding. In this case, the efficiency of MEC is higher because the information is removed quickly and the increase of inbreeding is lower in the first generation.

Real data

Results from the Przewalski's horse studbook analysis are shown in Table 1.2 which gives the total number of reproductive individuals with their correspondent mean coancestry (*f*) as well as the number of reproductive individuals with no

relationship with the undesired mare (i.e., individuals with no introgressed information) with their global *f* from 1935 to 2010. The total number of Przewalski's horse has quickly increased since 1980, but the influence of the Mongolian domestic horse introduced is still maintained, as no particular management strategy has been implemented to remove it.

From a total of 1800 theoretical reproductive individuals currently alive, just 182 are no related to the introgressed mare. The *f* of this group of individuals is twice the *f* of the total population, reflecting a huge reduction of the genetic diversity harboured by the *pure* subset. Currently, just a small part of the population remains pure and with a high inbreeding level. But similar levels are found when looking at the beginning of the recorded genealogy. The influence of the exogenous horse was quickly spread into the population so just a little percentage of highly related individuals were not descendant of the foreign mare when a few generations since the introgression elapsed.

YEAR	TOTAL		NATIVE	
	N_{ind}	f	N_{ind}	f
1935	38	0.053	26	0.093
1940	32	0.075	22	0.120
1950	23	0.084	13	0.132
1960	35	0.154	17	0.234
1970	100	0.175	42	0.246
1980	201	0.155	65	0.266
1990	506	0.145	111	0.269
2000	1023	0.144	129	0.254
2010	1800	0.140	182	0.257

TABLE 1.2. Results of the Przewalski's horse analysis.

DISCUSSION

Gene flow between populations is usually considered beneficial because it can protect the biological diversity and increases fitness by avoiding the rise of inbreeding depression. Some population analysis point out that extremely endangered populations can only be restored by *introgressing* new genetic

material, looking for an increase in fitness due to the reduction of inbreeding that introducing exogenous alleles can produce. This process is called *genetic rescue*. It has been proved that very low levels of migration are enough to recover most of the genetic variation lost in a small population (INGVARSSON 2001; TALLMON et al. 2004).

However, some disadvantages of the admixture of genetic information have been also pointed out, when dealing with wild or domestic species, such as economical devaluation or replacement of native populations by invaders. When undesirable introgression has occurred there is a need of developing methodologies to remove the exogenous genetic information in order to recover the original background related to the economic interest or the biodiversity component. In the present study, methods based in genealogical information were tested for their accuracy in the depuration, and the effects on other genetic measures of the population, mainly *F* and *f*, were determined.

For the whole range of parameters evaluated in the simulations, it was proved that the best strategy to remove undesired information is to avoid the contribution to the next generations of those carrying the highest proportion of genetic information coming from the exogenous founders. The MEC method achieved the best results regarding the elimination of exogenous alleles in all scenarios studied. Results were the same when measuring genealogical or genomic representation of the native founders.

Notwithstanding, the power of the strategy is limited by the total number of individuals and foreigners in the population, as well as the number of generations of mixing. However, even considering a long period of introgression, when the population is large enough the ability of finding individuals completely unrelated to the exogenous founders become higher. Therefore, the size of the population is also a very important factor that affects the possibility of recovery, and obviously, the levels of inbreeding reached after the management.

The principle of the MEC method implies to select among the available candidates those individuals which keep the highest percentage of native background (ideally without foreign information). Therefore, it reduces considerably the number of animals contributing to the next generation and

provokes the large increase of inbreeding reflected in Figure 1.2 (similar results were observed in observed homozygosity for the genomic data). The high level of coancestry induced by the MEC strategy denotes a large loss of diversity, being this phenomenon a drawback of the method. Actions taken to avoid this side effect (i.e., a restriction in the increase of the inbreeding coefficient) led MEC to perform like OC with slow increases of *F*, but also with no removal of undesired information. It is clear, then, that the level of de-introgression (and the speed of the process) is directly related to this increase in inbreeding. Less stringent restrictions on the acceptable ΔF can be tested to look for an equilibrium between the degree of removal and the loss of diversity.

As mentioned above, the removal of introgression under MEC is accomplished in the first generation of management. The method selects among all the available individuals, those with less exogenous influence. When these individuals mate, they produce all the descendants with the exact same proportion of relationship with the exogenous founders and, therefore, in the next generation MEC cannot choose among them for a second generation of management. If at least one male and one female can be found unrelated to all exogenous founders these individuals will be selected and all descendant from that moment will be also unrelated to foreigners. Consequently, no possibility of de-introgression will exist in later generations as all candidates would be equally valuable. This redistribution of the exogenous genetic information makes the method ineffective after just one generation. Thus, after implementing MEC in the first generation in order to achieve the maximization of the native genetic background, the sensibly strategy would be to keep on managing the population with OC strategy to minimise the further increase in *F* and the loss of genetic diversity.

General results point out the importance of the time passed away between the introgression and the starting of the management. They show that with few generations without management, a small amount of introgression can spread into the population and turn out almost impossible to recover. Therefore, it is very important to act as soon as possible to keep introgression controlled. Results of the simulations are confirmed with real data from the Przewalski's horse Studbook, where the influence of a single foreigner greatly influences the whole pedigree and

remains noticeable in the current population, a hundred years later. The important efforts made in the past to save the Przewalski's horse and the current programs to reintroduce it in wild have been very successful in achieving the preservation of a species that was almost extinct (less than 50 horses after the World War II). Currently it is a potentially viable population living in the border between Mongolia and China (BOUMAN & BOUMAN 1994). However, after 100 years, the influence of the Mongolian domestic horse introduced in the population is still visible and the efforts to remove it would imply a large loss of viability making the idea of recover original Przewalski's horse background unfeasible.

The concern about conserving populations pure, especially livestock breeds (HALL & BRADLEY 1995), lead us to warn about the drawbacks of admixture. The results of this study enhance the importance of keeping recorded genealogies of the populations, and show the importance of a quick reaction against introgression to keep it under control. The efficiency of the information provided by genealogies depends on the total amount of individuals and the percentage of undesired introgression and admixture. The genealogy has to be completely available to be useful to our de-introgression purposes, which could be relatively common in livestock breeds, but not feasible in natural populations threatened for other invasive species or breeds where keeping track of individuals is difficult or impractical (ALLENDORF & LUIKART 2007). In this case, there is a need to incorporate extra information. Molecular information can provide a tool to identify the breed to which animal belongs to. As molecular markers were capable of reveal introgression (GROENEVELD et al. 2010), they will be useful for accomplishing the opposite effect, i.e., helping in the identification of the undesired individuals that are more related to the exogenous genetic background. Molecular markers can also be helpful for reconstructing the pedigree (BUTLER et al. 2004) allowing to apply the strategy suggested in this study. Also the molecular coancestry can be used to trace those individuals which have a foreign origin and select among the population individuals less related to them.

The conclusion from the present study is that even small undesired introgression can lead a population to quickly get mixed and lose the genetic conformation that it is intended to be preserved. In that case actions should be

taken as soon as possible to recover most of the original genetic background. The importance of keeping pedigree records, or the possibility of reconstructing it through molecular markers, was also pointed out.

CAPÍTULO 2

Molecular markers allow removing introgressed genetic background: a simulation study

INTRODUCTION

Mixing populations is beneficial because of the improvements on fitness related to the avoidance of inbreeding depression (FRANKHAM et al. 2002). However, there can also be negative consequences of mixing genetic resources. Outbreeding depression may occur due to hybridization and break up of co-adapted gene complexes (when local adaptation exists) and this may lead to extinction in both wild and domestic populations (RHYMER & SIMBERLOFF 1996; ALLENDORF et al. 2001). Moreover, invasive species are a great threat, often irreversible, and with major implications to human health, economic losses or disruption of ecosystems. Introgression can be particularly risky for endangered populations like wild animals threatened by their domestic relatives. This situation, more frequent than expected, increases because of human-mediated actions (VITOUSEK et al. 1997).

Crossing farmed animals can also be undesired. Several domestic breeds are associated to quality products with economic interest (*Iberian* pigs, *Reggiana* dairy cows…) and they are meant to be kept with a pure genetic background, because the economic benefits may disappear if purity is not assured (DALVIT et al. 2007).

Introgression examples have been documented between some species: wolves (*Canis rufus*) and coyotes (*Canis latrans*) (MILLER et al. 2003), wild and domestic partridges (*Alectoris rufa* and *A. graeca*) (NEGRO et al. 2001), bisons (*Bison bison*) and domestic cattle (*Bos taurus*) (FREESE et al. 2007), European mink (*Mustela lutreola*) and polecat (*Mustela putorius*) (CABRIA et al. 2011), rainbow trout (*Oncorhynchus mykiss*) and cutthroat trout (*Oncorhynchus clarkii*) (HOHENLOHE et al. 2011). In some of these cases the introgression does not imply any risk to the population, and the studies are merely descriptive or point out to uncommon events. In many others, the introgression is a threat to the species, which are vulnerable to this phenomenon. A biased gene flow can imply that important species in some ecosystems become more endangered suggesting that some actions should be taken.

In a previous study (Chapter 1), different scenarios trying to cover the complexity of the introgression events were simulated, varying the number of

foreign individuals entering the population and the number of generations elapsed before the recovery management started. In that study, the information of the pedigree (completely recorded since the introgression took place) was used to select which individuals should contribute to the next generation in order to remove non-native alleles. Among the tested methods, minimisation of the coancestry with the foreign founders provided the best results regarding the amount of exogenous genetics eliminated. This strategy allowed removing part of the exogenous alleles in most scenarios. However, even small introgression phenomena (i.e., few foreign individuals and few generations of admixture) could lead to irrecoverable situations, encouraging the strict control of the populations and the rapid action in case of undesired introgression. The study also pointed out the problem of increased inbreeding and coancestry associated to the removal process.

As many studies have shown, molecular markers can help in the detection of hybrids and in the discovery of introgression events (GROENEVELD et al. 2010). Therefore, it would be expected that they can also be used to accomplish the opposite task (i.e., identifying the purest individuals, helping in the removal of exogenous alleles from the population).

The objective of this study was to analyse, through computer simulations, the efficiency of several methods based on markers information on the removal of undesired exogenous alleles from a mixed population. The study assumes that the population is kept in captivity and, thus, there is a tight control on the reproductive process. For this purpose, several scenarios where exogenous genetic information was admixed in a native population were simulated. Then, molecular markers based techniques were used to recover the native background.

MATERIALS AND METHODS

Population structure

Populations with two different sizes, 100 individuals and 20 individuals (50% males and 50% females), were simulated with constant size and sex ratio along generations. One hundred individuals could represent a typical population size for local breeds of domestic animals. Twenty individuals is a more realistic scenario

when dealing with conservation programs of endangered wild species, which usually have smaller population sizes.

The populations ran during two periods. A first period (admixture), with a variable number of random discrete generations, and a second one (management), with ten discrete generations. Two factors determined the different introgression scenarios.

Number of exogenous individuals. In the 100 individuals population, 10 to 50 exogenous individuals (sexes randomly assigned in each replicate) were included as part of the base population (the rest of individuals, up to 100, were native to complete the base population), implying an introgression percentage of 10 to 50%. In the 20 individuals scenarios, the percentage of introgression simulated was the same (10-50%) by including 2, 4, 6, 8 or 10 exogenous individuals in the base population and completing up to 20 with native individuals.

Number of generations without management (admixture period). One to five generations with random contributions and mating were simulated prior to management, to simulate the admixture of the foreign alleles that were included in the base population into the native genetic pool.

The genome of each individual was made up of one chromosome of 20M, with a total of 2000 multiallelic loci (non-marker loci). Individuals in the base population (all non inbred and unrelated) carried two different alleles at each locus and, thus, were all heterozygous and different from each other (the number of alleles per locus in generation 0 is 200 or 40, for N = 100 or N = 20, respectively). This situation is completely informative. Besides, the origin of each allele (native or exogenous) can be determined, and it was used for evaluation. These 2000 multiallelic loci were used in the evaluation of the efficiency of the methods eliminating exogenous alleles and to measure the maintenance of genetic diversity. When creating gametes, a Poisson distributed (λ = 20) number of crossing-overs (one crossover is expected on average on each Morgan) with no interference were generated in random positions over the chromosome.

Additionally, markers were simulated (evenly spaced along the genome) to be used in the removal of the foreign alleles. Different situations were considered:

Diagnostic markers. Five to 20 biallelic markers were simulated. In the base population, all native individuals were homozygous for allele 1 in all markers, and all foreigners were homozygous for allele 2. Therefore, alleles were private for native or foreign individuals (see Table 2.1).

MARKER TYPE	NUM. MARKERS	POPULATION	*ALLELE 1* FREQ.	*ALLELE 2* FREQ.	*ALLELE 3* FREQ.	*ALLELE 4* FREQ.
Diagnostic	5 10 15 20	*Native*	1	0	—	—
		Exogenous	0	1	—	—
Diagnostic-like	5 10 15 20	*Native*	0.80	0.20	—	—
		Exogenous	0.20	0.80	—	—
Non-Diagnostic	5 10 15 20	*Native*	0.80	0.07	0.06	0.07
		Exogenous	0.07	0.80	0.06	0.07

TABLE 2.1. Combinations of frequencies in the native and the exogenous population of each possible allele in each type of marker simulated.

Diagnostic-like markers. Five to 20 biallelic markers were simulated. The two alleles in each marker were present in both populations with very different frequencies. In the native population, allele 1 was present at frequency 0.8 and allele 2 at frequency 0.2, while in the foreign population, the frequencies for allele 1 and 2 were 0.2 and 0.8, respectively. This distribution of frequencies was simulated to mimic a scenario where alleles are thought to be private but they are not. To further investigate the consequences of this erroneous assumption, extra simulations were run with different sets of frequencies.

Non-Diagnostic. Five to 20 markers with four alleles each were simulated. Frequencies of the alleles in the original native and foreign populations are shown in Table 2.1 and were assumed to be known without error.

Management

Diagnostic and diagnostic-like markers. In every generation of management individuals with the highest number of *native alleles* were chosen to be parents of the next generation. In the *diagnostic* markers scenarios, *native alleles* were those exclusive of the native population, and in the *diagnostic-like* scenarios, *native alleles* were those at the highest frequency (0.8) in native individuals. Both strategies consider the markers as diagnostic, but in the diagnostic-like scenario, this is an incorrect statement, and allows investigating the consequences of assuming a marker as being diagnostic when it is not.

Thus, among all individuals available each generation, only those with the maximum number of native alleles contributed to the next generation (at least one male and one female should be selected to allow for contributions from both sexes to be the same). For example, if just one female has the maximum number of *native alleles*, this female will be the mother of all the offspring. According to this assumption, the number of individuals contributing offspring was not the same in each generation of management. When more than one individual had the same (and maximum) number of *native alleles* (the most likely situation), the contribution of each individual was randomly decided. No explicit restriction on the number of contributing individuals was imposed.

Non-Diagnostic. To recover the native background, the contributions to the next generation were decided by minimising the expected genetic distances, between the original native population (with frequencies assumed known without error) and the current population. Three genetic distances were considered:

Cavalli-Sforza and Edwards Chord Distance (1967)

[2.1] $$D_{ch} = \sum_{m}^{M} \sqrt{2(1 - \sum_{a}^{A_m} \sqrt{p'_{am}\, p_{am}})},$$

Nei's Minimum Distance (NEI 1973; NEI 1987)

[2.2] $$D_m = \frac{J_X + J_Y}{2 - J_{XY}},$$

where

[2.3] $$J_X = \sum_{m}^{M} \sum_{a}^{A_m} \frac{p'^2_{am}}{M},$$

[2.4] $$J_Y = \sum_{m}^{M} \sum_{a}^{A_m} \frac{p^2_{am}}{M},$$

[2.5] $$J_{XY} = \sum_{m}^{M} \sum_{a}^{A_m} \frac{p_{am} p'_{am}}{M},$$

Kullback-Leibler Divergence (1997)

[2.6] $$KL = \sum_{m}^{M} \sum_{a}^{A_m} p'_{am} \log\left(\frac{p'^2_{am}}{p_{am}}\right),$$

In all cases, A_m is the total number of alleles in locus m, M is the total number of marker loci, p_{am} is the frequency of allele a at locus m in the native population, and p'_{am} is the expected frequency in the next generation due to a particular scheme of contributions (i.e., number of offspring per parent). This can be calculated as:

[2.7] $$p'_{am} = \sum_{i=1}^{N} c_i g_{iam},$$

where N is the number of individuals, c_i is the relative contribution of individual i to the next generation, and $g_{i\,am}$ is the probability of gametes from individual i carrying allele a of marker m (1 for homozygotes aa, 0.5 for heterozygotes and 0 for individuals not carrying allele a).

All the optimizations were solved using *simulated annealing* algorithms (KIRKPATRICK et al. 1983; FERNÁNDEZ & TORO 1999). Once contributions were decided, minimum coancestry matings were arranged in all cases using the *Hungarian* algorithm (DANTZIG 1963). Twenty replicates per scenario were simulated and results presented are averages across replicates.

The pedigree of the populations was recorded during the two periods for evaluations, but never used in the management.

Variables

Every generation, several variables were calculated to evaluate the efficiency of the strategies in the de-introgression (removal of foreign alleles): native founder

representation (i.e., the proportion of alleles coming originally from native founders, calculated from non-marker loci), average inbreeding coefficient, mean coancestry and observed homozygosity. Inbreeding coefficient and mean coancestry values were calculated from the pedigree data.

RESULTS

Native representation

Results for native representation (i.e., the proportion of alleles coming from native founders, *NR*) are shown in Figure 2.1 for five to 20 *diagnostic*, *diagnostic-like*, and *non-diagnostic* markers in the 100 individuals population. As expected, the results obtained with the *diagnostic* markers were the most efficient of the three types, with the native representation being higher as the number of markers increased. In all cases using *diagnostic* markers, some recovery was achieved, reaching a complete recovery of the native background in those cases where the introgression spread in the population during a low number of generations. When using *diagnostic-like* markers the usefulness decreased as a consequence of assuming them being *diagnostic*, but still, there was some recovery that must be highlighted. The degree of recovery of the native background was always lower than when using the *diagnostic* markers, but the differences were small, particularly in scenarios with a larger admixture period (5 generations). The influence of the number of markers used in the management is also remarkable, with higher levels of information (more markers) yielding better results.

More detailed results for 20 *diagnostic-like* markers with several *native allele* frequencies and intermediate frequencies of the alleles in the foreign population (0.5/0.5), in scenarios with an admixture period of five generations and after ten generations of management (100 individuals population), are shown in Table 2.2. As expected, the greater the frequency of the *native allele*, the higher the removal of undesired introgression as we were approaching the *diagnostic* scenario, even when the frequency of the *native allele* in the exogenous population was high.

In all cases, the maximum *NR* reached was achieved after three to four generations of management and, in cases with a little admixture, the recovery was complete in just one generation of management (data not shown).

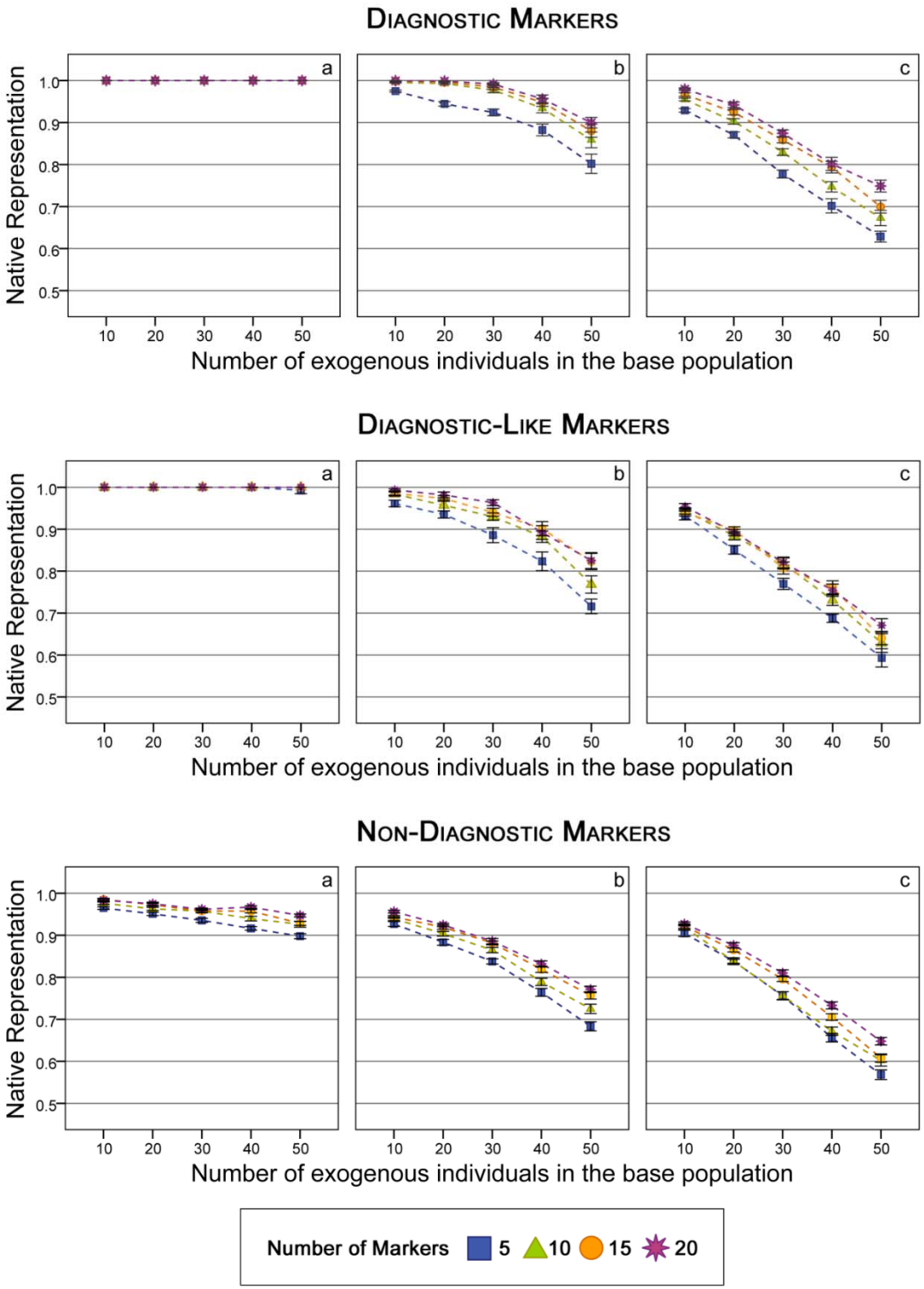

FIGURE 2.1. Native representation under the different management strategies in the 100 individuals scenarios (N = 100). Values shown are those obtained at the 10th generation of management. a) 1 generation of admixture, b) 3 generations of admixture, c) 5 generations of admixture.

NATIVE ALLELE FREQ.		NUMBER OF EXOGENOUS INDIVIDUALS				
		10	20	30	40	50
0.70	***NR***	0.929 ± 0.008	0.840 ± 0.012	0.735 ± 0.012	0.649 ± 0.015	0.584 ± 0.019
	F	0.417 ± 0.013	0.443 ± 0.009	0.456 ± 0.008	0.462 ± 0.009	0.483 ± 0.010
0.80	***NR***	0.930 ± 0.008	0.866 ± 0.012	0.785 ± 0.019	0.714 ± 0.017	0.588 ± 0.017
	F	0.334 ± 0.015	0.365 ± 0.009	0.395 ± 0.009	0.412 ± 0.011	0.466 ± 0.010
0.90	***NR***	0.950 ± 0.007	0.867 ± 0.011	0.792 ± 0.017	0.722 ± 0.017	0.620 ± 0.017
	F	0.245 ± 0.012	0.300 ± 0.012	0.345 ± 0.009	0.355 ± 0.014	0.369 ± 0.011
0.95	***NR***	0.952 ± 0.006	0.893 ± 0.011	0.803 ± 0.015	0.732 ± 0.018	0.638 ± 0.016
	F	0.208 ± 0.013	0.249 ± 0.011	0.309 ± 0.010	0.336 ± 0.012	0.371 ± 0.012
0.99	***NR***	0.954 ± 0.004	0.903 ± 0.010	0.814 ± 0.012	0.729 ± 0.016	0.640 ± 0.016
	F	0.099 ± 0.005	0.191 ± 0.013	0.266 ± 0.009	0.313 ± 0.014	0.346 ± 0.012

TABLE 2.2. Results obtained for Native Representation (*NR*) and inbreeding coefficient (*F*) after managing during 10 generations with 20 *diagnostic-like* markers with different *native allele* frequencies (5 generations of admixture scenarios). In all cases the frequency of the *native allele* in the foreign population is 0.5.

Besides, the initial amount of alleles introgressed in the native population is also a key factor to determine the potential of success, cases with high number of exogenous individuals leading to an irrecoverable situation. It must be pointed out that in scenarios with many generations of admixture, the possibilities of recovery are quite low, even when using information of many markers (see Figure 2.1 and Table 2.2).

Results obtained minimising any of the three genetic distances (using the information of the *non-diagnostic* markers) were similar and, consequently, only the results for the minimisation of the *Kullback-Leibler Divergence* (*KL*) are presented (Figure 2.1, lower panel). As in the other scenarios, the ability of recovery of the original background using the minimisation of *KL* was greatly dependent on the number of generations of admixture. In cases with a short

admixture period, a good percentage of native representation could be recovered, even with a high proportion of individuals introgressed (40-50%). Notwithstanding, in those scenarios where individuals mixed for many generations the restoration was minimal. Again, the recovery increased with the number of markers.

It is remarkable that *diagnostic-like* markers performed better (i.e., leads to higher levels of *NR*) than the *non-diagnostic* markers in all situations irrespective of the number of generations without management or the number of exogenous individuals. Results for both types of markers are comparable despite the fact that the number of alleles was not the same, because the frequency of the most common allele is the same in both simulations. Hence, if markers had four alleles at frequencies 0.8/0.07/0.06/0.07 (as in the *non-diagnostic* scenario), managing with the *diagnostic-like* method would imply to choose as *native* allele the most frequent and taking no action for the rest of alleles, no matter if there is one or three more.

Some other simulations were carried out with different combinations of frequencies and number of alleles (not shown). Results were always linked to the differences in the frequencies between both populations. Cases with a large number of alleles gave lower values of recovery because allele frequencies become more similar between native and foreign populations, making impossible to differentiate the origin of alleles.

The results for native representation in the 20 individuals population (Figure 2.2) showed the same pattern as the 100 individuals population. The *diagnostic* markers obtained the better results, and *diagnostic-like* markers performed better than minimising the genetic distances in all scenarios, confirming the ranking in the efficiency of the methods. However, the percentage of native genetics recovered in the 20 individuals simulations was slightly lower in all cases, especially when the percentage of introgression was high (40-50%). Scenarios with five generations of admixture were almost irrecoverable (irrespective of the method) pointing out the importance of acting soon especially in small populations.

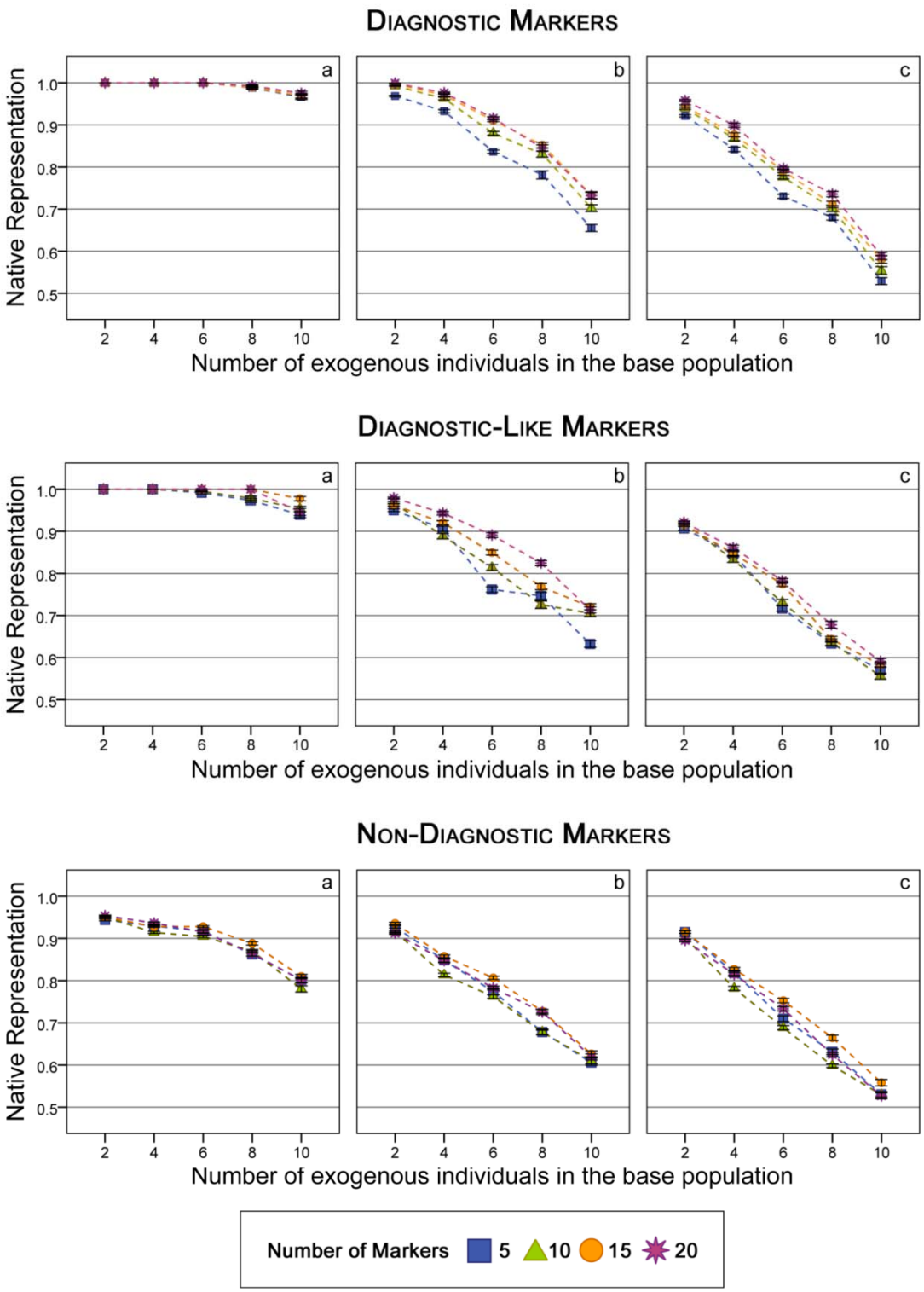

Figure 2.2. Native representation under the different management strategies in the 20 individuals scenarios (N = 20). Values shown are those obtained at the 10th generation of management. a) 1 generation of admixture, b) 3 generations of admixture, c) 5 generations of admixture.

Inbreeding coefficient

Trends for mean coancestry and inbreeding coefficients were similar. Consequently, only the levels of inbreeding (*F*) after ten generations of management are presented in Figure 2.3 for five to 20 markers of the three types. The increase of inbreeding is a clear side-effect of the de-introgression process when managing with diagnostic or diagnostic-like markers. The values of *F* were higher in cases with more introgression to remove. This is a consequence of the restriction on the number of individuals contributing to the next generation posed by the method. The higher the levels of introgression, the fewer individuals are expected to be pure, and thus, when selecting individuals to reproduce, inbreeding will rise faster.

When evaluating scenarios with *diagnostic-like* markers, the increase of inbreeding was more pronounced than that obtained with the *diagnostic* markers. This can be observed for all scenarios (Figure 2.3 and Table 2.2). The increase of inbreeding (consequence of the method) was higher with a lower frequency of the *native allele*. The number of individuals contributing was lower than when using the *diagnostic* management (data not shown). Having different frequencies makes fewer individuals bearing the same number of *native* alleles and thus, fewer individuals will contribute offspring. Moreover these individuals are not the purest, because the alleles are not really private, and the Native Representation results are worse.

As mentioned before, the maximum *NR* is reached after 3 or 4 generations of management when differences in *NR* are no longer found between individuals. Therefore, the method selects contributions randomly in subsequent generations. From that point, management could be switched to a method devoted exclusively to keep diversity and avoid the increase of inbreeding, like minimum coancestry contributions (CABALLERO & TORO 2000).

No significative differences were found between using any of the three genetic distances for de-introgression purposes (nor for inbreeding results either) so only the *KL* results are presented in Figure 2.3 (lower panel). The increase of the inbreeding due to the ten generations of management was small in all cases. Lower

values of *F* were obtained when using non-diagnostic than using diagnostic or diagnostic-like markers.

The increase of inbreeding after ten generations of management in the 20 individuals population (under the three strategies) is shown in Figure 2.4. As happened in the 100 individuals population, the increase of inbreeding is a side-effect of the method, reaching higher values of *F* due to the smaller population size, but following the same pattern as in the previous simulations.

Observed homozygosity

The evolution of the observed homozygosity calculated from the non-marker multiallelic loci (representing the realisation at the genomic level of the pedigree inbreeding) in all scenarios agreed with the results obtained from the inbreeding coefficient (data not shown). As the 2000 loci were completely informative in the base population, observed homozygosity also measures the identity by descent leading to the same results.

DISCUSSION

Undesired introgression of genetic material into a population may be a situation to avoid and, when it happens, it would be essential to take actions to recover the original native genetics and to remove the exogenous genetics. This could happen in livestock populations with an economic interest linked to the genetic background, or in natural populations, endangered by the admixture with exogenous invaders. To achieve a successful recovery all the available information could be useful. The use of a pedigree has been shown to provide good results (Chapter 1) but with several limitations, including the requirement of a perfectly recorded pedigree.

In the absence of a pedigree, molecular markers have been used to detect introgression (GROENEVELD et al. 2010) and to differentiate between native and exogenous origin when dealing with admixed populations (OLIVEIRA et al. 2008). Here, it has been proven that they can also be helpful in removing exogenous alleles from a population. Results show that the efficacy of markers is clear and that the efficiency increases with the number of markers, as expected.

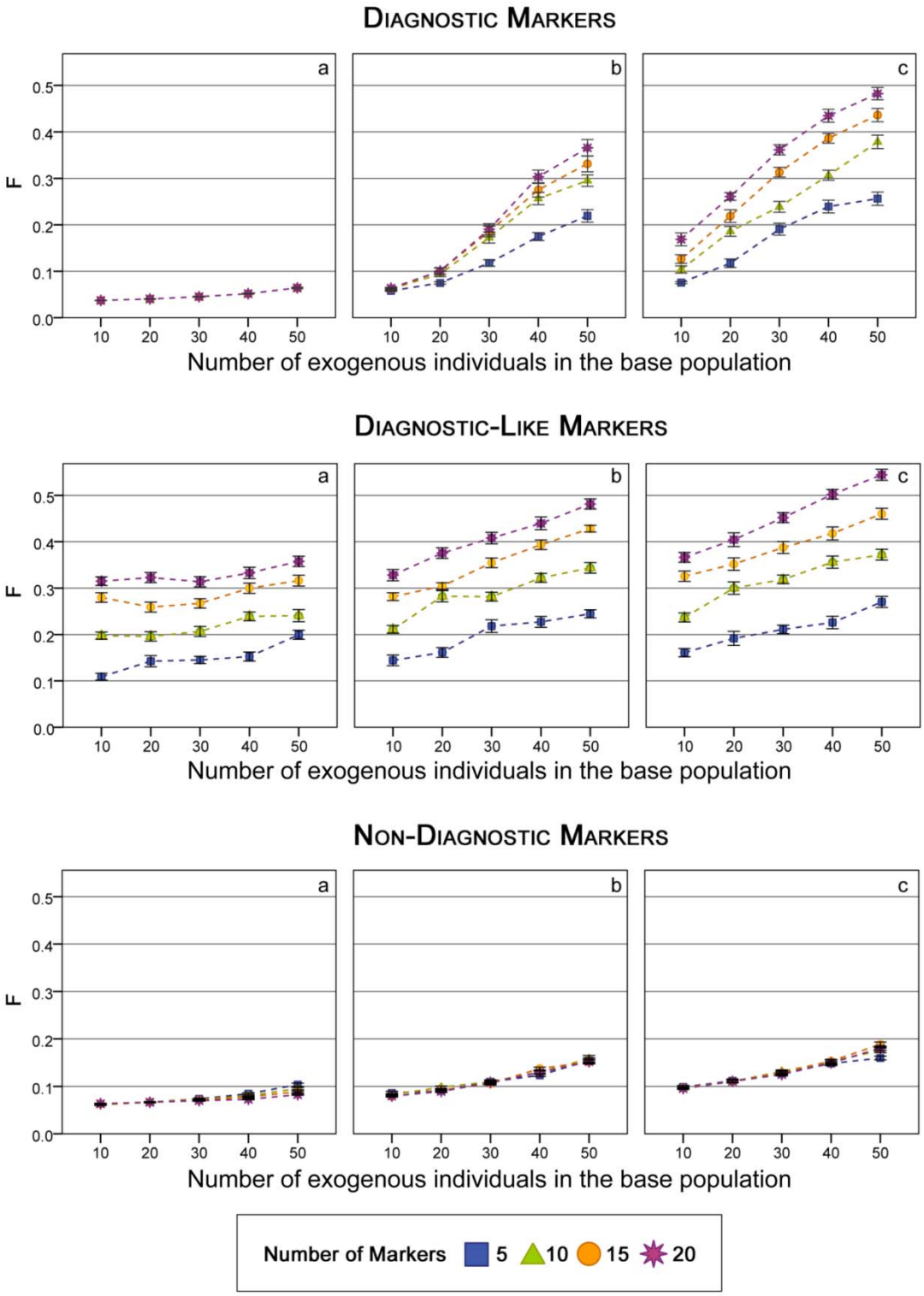

FIGURE 2.3. Inbreeding coefficient under the different management strategies in the 100 individuals scenarios (N = 100). Values shown are those obtained at the 10th generation of management. a) 1 generation of admixture, b) 3 generations of admixture, c) 5 generations of admixture.

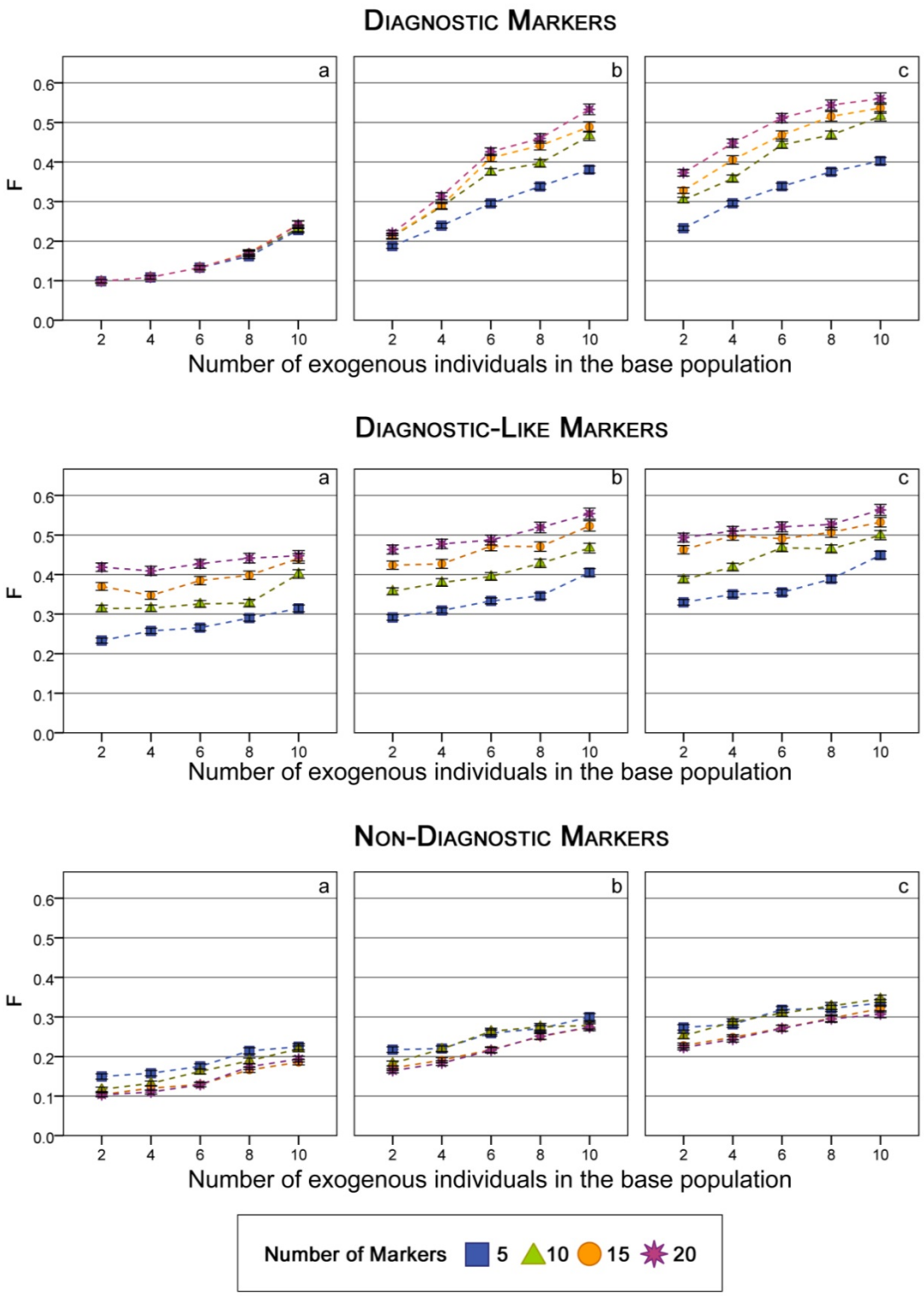

FIGURE 2.4. Inbreeding coefficient under the different management strategies in the 20 individuals scenarios (N = 20). Values shown are those obtained at the 10th generation of management. a) 1 generation of admixture, b) 3 generations of admixture, c) 5 generations of admixture.

The greatest percentage of recovery was achieved with *diagnostic* markers, because with their private alleles it is possible to identify clearly native and foreign origins of alleles in the candidate individuals. Despite some examples of private alleles that have been found at the level of species or subspecies (ROY et al. 1994; MACHUGH et al. 1997; GOODMAN et al. 1999) most of the times private alleles may be uncommon in closely related populations (VILÀ et al. 2003). Unfortunately, this is the most likely situation when dealing with introgression.

The impact of the population size on the recovery of native genetics reflects that the number of individuals in the population is crucial in the de-introgression process. Having a larger number of individuals allows the methods to have a higher power of recovery, because the probability of finding purer individuals increases. Anyway, the diagnostic markers were able to remove some exogenous genetics even in the smaller population, except for the most introgressed scenarios (5 generations of admixture).

On the other hand, the diagnostic nature of a particular allele may be false and only due to deficient information (e.g., when not enough number of animals has been genotyped). Results from simulations using *diagnostic-like* markers (i.e., assumed to have private alleles but not having so) showed that, even with this incorrect information, some percentage of native background can be restored. This is true even in the worst case scenario, in which the allele considered as *native* can be present in the exogenous population at a frequency as high as 0.5 (Table 2.2). Finding enough number of markers with relatively extreme frequencies can lead to an acceptable recovery.

When we are aware of markers not being diagnostic, relying on genetic distances may also lead to the recovery of high levels of the original genetic background. The higher the differences between allele frequencies in both populations, the greater the recovery. The kind of markers required to apply the genetic distances strategy are more common; the key factor is having allele frequencies different enough between the pure admixed populations. Nevertheless, treating them as *diagnostic* by selecting the presence of the more frequent allele has been proved to be more effective than minimising the *KL* divergence (or any of the other two distances, which yielded equal results). This

happens in the two population sizes tested, although the differences between *diagnostic-like* and *no-diagnostic* in the 20 individuals population are no so high.

Using markers as if they were diagnostic (as in the *diagnostic-like* approach) instead of using genetic distances has also another advantage. While genetic distances require good estimates of frequencies in *pure* populations, and being sure that the population used as a reference has the same genetic origin as that in the process of de-introgression, selecting the presence of an allele requires no assumptions about frequencies.

In addition to the amount of information provided by the markers, the probability of success in de-introgression is clearly related to the percentage of undesired background introgressed in the population and to the length of the admixture period.

The values of *NR* obtained when using pedigree information to recover the native background (Chapter 1) were very similar to those achieved with the *diagnostic* markers and slightly higher than managing with *diagnostic-like* and *non-diagnostic*. Therefore, a reasonable number of informative markers are enough to achieve the same recovery without the requisite of a total knowledge of the genealogy, which could be a highly unlikely situation in many cases.

Some extra simulations were carried out to test the efficiency of the methods in a smaller genome (number of crossovers simulated through a Poisson distribution with $\lambda = 1$ representing one chromosome of 1M). The efficiency of the methods becomes higher in this case (Figure 2.5) because the markers are now more informative, due to the higher linkage disequilibrium between markers and the rest of loci. The *diagnostic* markers are still the best approach to remove the exogenous genetics, but the other two methods allow recovering almost 100% of the native genome in the scenarios tested. Notwithstanding, this genome length is very unrealistic for species in conservation programs.

Similar to management based on pedigree (Chapter 1), the de-introgression process using *diagnostic* and *diagnostic-like* markers implies an increase of inbreeding due to the inherent reduction in the number of contributing individuals, in both genome sizes (Figure 2.6).

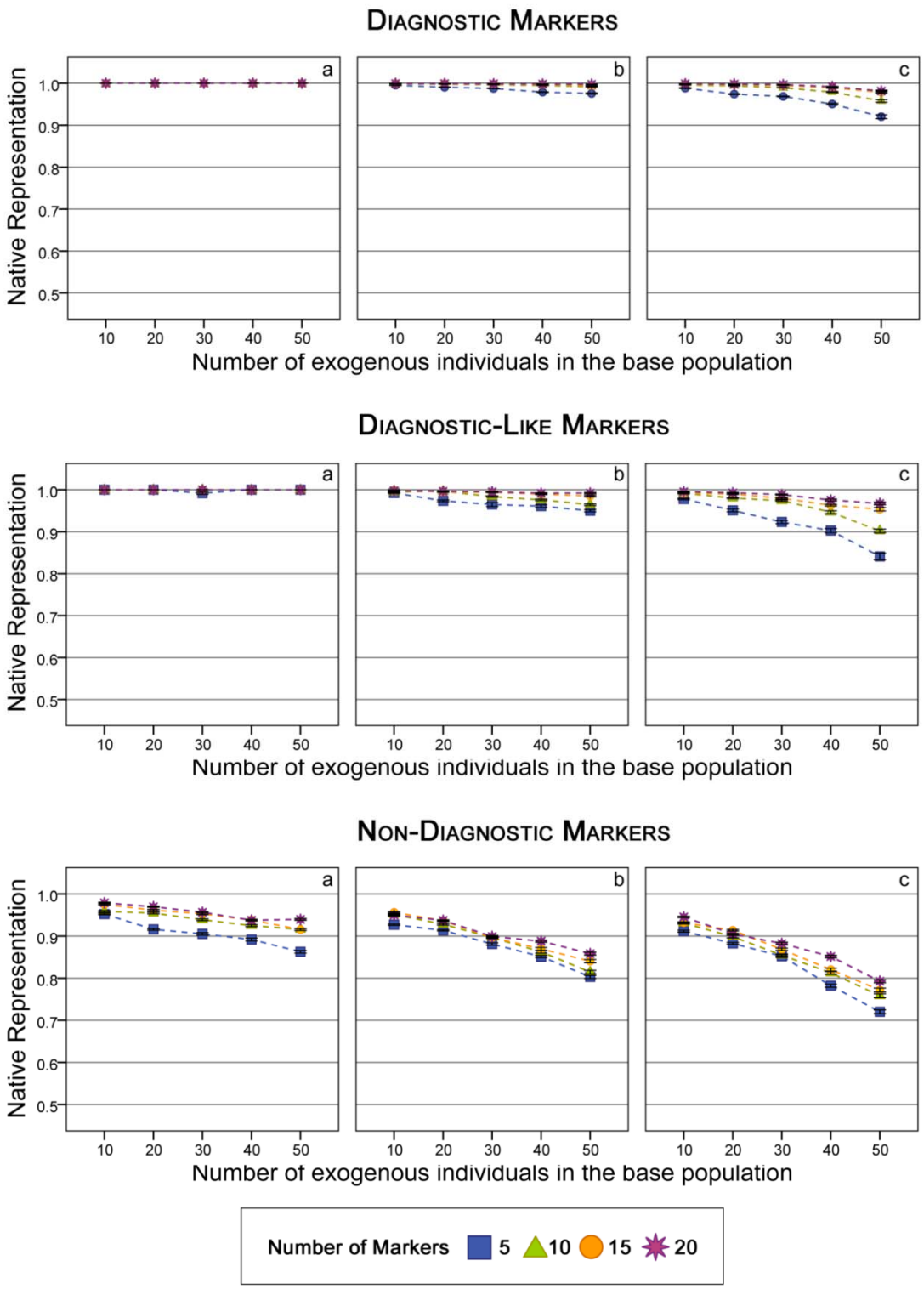

FIGURE 2.5. Native representation under the different management strategies in the 1 Morgan scenarios (N = 100). Values shown are those obtained at the 10th generation of management. a) 1 generation of admixture, b) 3 generations of admixture, c) 5 generations of admixture.

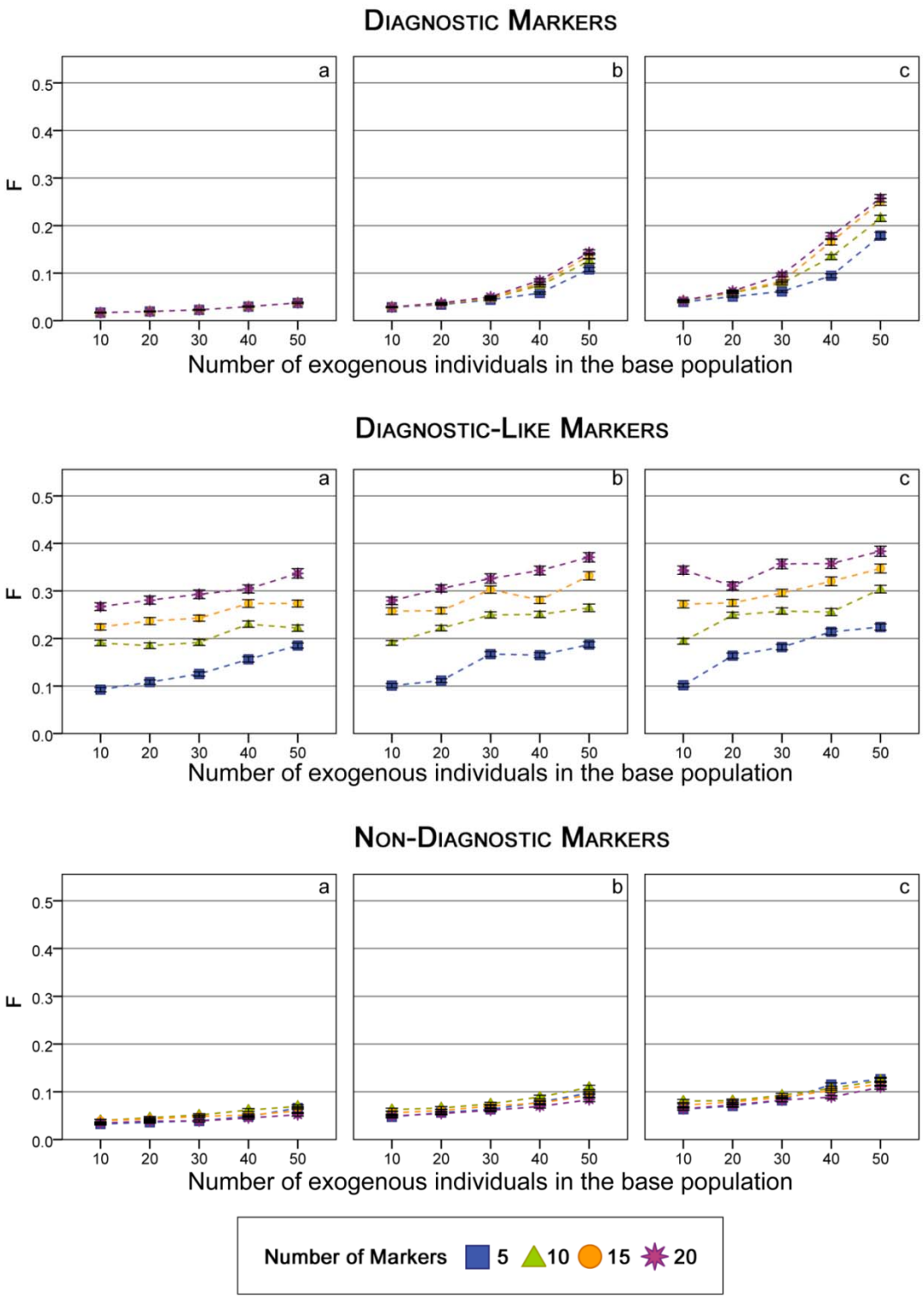

Figure 2.6. Inbreeding coefficient under the different management strategies in the 1 Morgan scenarios (N = 100). Values shown are those obtained at the 10th generation of management. a) 1 generation of admixture, b) 3 generations of admixture, c) 5 generations of admixture.

The effective population size (N_e) of individuals in each generation of management varies during the generations, being relatively small at the beginning of the management (N_e = 3-20 in the *diagnostic*, N_e = 2-15 in the *diagnostic-like* markers, N_e = 30-50 in the *non-diagnostic* markers) and becoming higher (around 100) when the methods stop working. It is not a realistic scenario managing a population using one female and one male, and, of course, some kind of control on the loss of diversity should be incorporated. However the restriction in the inbreeding (or assuring a minimum number of individuals contributing each generation) cannot be generalised, because the particular value will depend on the characteristics of the species and the genetic structure of the population.

This side effect of the methods must be taken into account when planning the management by deciding what rate of inbreeding (ΔF) we are willing to accept in the process of recovery. Explicit restrictions on the minimum number of contributing parents or on expected molecular coancestry of the next generation may be implemented to avoid a too much rapid increase of *F*. Additionally, the number of generations that the method is applied should be limited in order to replace the management for de-introgression with the classical strategy to control the increase of inbreeding. Enlarging the period of removal may lead to little extra recovery with a large rate of *F*, which may be unacceptable. From the moment that the maximum *NR* has been reached, the objective of the management would be to maximise diversity. However, as it is not possible to predict the time at which maximum *NR* has been reached, it could be advisable to include the restriction on ΔF from the start of the management.

Whereas increasing the number of markers leads to a slight improvement in the recovery of native background, it also leads to an increase in inbreeding particularly when using *diagnostic* and *diagnostic-like* markers. Values of *F* reached after managing with a large number of these markers were high, suggesting another variable to take into account in each situation to get the highest recovery but losing the lowest amount of genetic diversity.

On the other hand, the inbreeding reached by minimizing the *KL* divergences was lower than it was with the other methods (irrespective of the number of markers), which implies that the number of individuals contributing descendants

to the next generation is higher. This must be also taken into account, especially when dealing with populations where the control of *F* is needed.

The values of *F* obtained in the pedigree management (Chapter 1) were lower than those obtained with the *diagnostic* and *diagnostic-like* markers, excluding those cases with a small number of markers and particularly in cases with a lot of admixture.

As mentioned before, the acceptable value of ΔF depends on each situation and the information available must be assessed before starting the process. In any case, we must be aware that a control over the inbreeding during the management would always imply a loss of efficiency. It should be analysed every generation, deciding which amount of exogenous alleles and increase of inbreeding is acceptable for each particular situation.

The present results apply only to captive population, where reproductive control is high. If our interest is to de-introgress a wild population, a good alternative would be to establish an *ex-situ* population, where the proposed methods are to be applied, and this will provide purer individuals to be released into nature.

The conclusion from the present study is that a not too large set of markers can provide a good tool for removing undesired introgression from a population. The use of this information can lead to a substantial recovery, especially having *diagnostic* markers or alleles much more represented in the population of interest than in the exogenous one. The importance of acting soon to avoid irrecoverable admixture of the exogenous genetics is a main concern common to all the methods (as it was in the pedigree approach) and it highlights the importance of prevention by controlling these populations as far as possible.

CAPÍTULO 3

Advantages of using molecular coancestry in the removal of introgressed genetic information

INTRODUCTION

Interbreeding can be considered as a positive or negative strategy to population's management depending on the situations. Many studies analysed the benefits of a new genetic input: gene flow between populations can restore a loss of genetic diversity and avoid the disadvantages of inbreeding (FRANKHAM et al. 2002; TALLMON et al. 2004). Nevertheless, disadvantages of the interchange of genetic material have been also observed. Introgression can lead populations to extinction and nowadays this is more likely to occur because the number of invasive species threatening wild populations has increased noticeably due to human activities (RHYMER & SIMBERLOFF 1996; ALLENDORF et al. 2001). In the field of domestic animals, the maintenance of pure populations can be essential to assure some quality products (DALVIT et al. 2007), because of other economic reasons such as horses involved in competitions and other activities (MAPA 2003) or for aesthetical reasons like dog breeds (PARKER et al. 2004).

Livestock breeds are recognized as important components of world biodiversity (HALL & BRADLEY 1995). Local breeds have been selected to fit a wide range of environmental conditions and human needs and their genetic diversity could help developing new characteristics in response to changes in environment, diseases, or food quality or quantity demand. The selection of a few highly productive breeds has caused the decline of numerous breeds which often possess special adaptations (to harsh conditions, disease resistance, etc.) not found in the former (SCHERF 2000; MAUDET et al. 2002). In many cases, crossbreeding with a more productive breed leaded to the disappearance of the specific features and adaptive traits of local breeds. Therefore, they should be recovered to avoid population extinction (UGARTE et al. 2001; MORAIS et al. 2005; TABERLET et al. 2008).

In a previous study the pedigree was used as source of information to recover an introgressed genetic background (Chapter 1). Different introgression events were simulated with a varying number of exogenous individuals entering the population and different number of generations in which the information was admixed. Based on the information of a completely recorded genealogy, the minimization of the coancestry of the current population with the foreign founders proved to be the best method to remove the largest amount of undesired

introgression. Notwithstanding, the method had some disadvantages, like an extra increase in the inbreeding of the population. Moreover, in cases where the amount of introgression was too high or uncontrolled for many generations the method was relatively inefficient.

In most realistic scenarios a reliable pedigree is lacking and the use of molecular information is the only option. In a simulation study (Chapter 2) the information of a few microsatellite-like markers was used for the same purpose of removing undesired introgression through the calculation of genetic distances between admixed population and the pure ones or by direct selection of carriers of private alleles exclusive of the native population. In these cases the success was related to the differences in the frequencies between the exogenous and the native population. In situations with a few markers with very similar allele frequencies in both populations the efficiency decreased considerably.

The increase in the number of markers that have become available through the new developments of genotyping techniques, allows for the replacement of pedigree information by marker information in many tasks. Moreover, dense marker information could provide a more precise picture of genetic conformation, being more powerful than pedigree information, which corresponds to an expectation along the whole genome. Hence, when genealogical data is not available, molecular markers can be used directly through the calculation of molecular coancestry or used to estimate the genealogical coancestry.

Some studies have analysed the ability of molecular markers to substitute the genealogical information. HAYES et al. (2009) demonstrated that replacing the relationship matrix derived from pedigree with a realized matrix (calculated through genome-wide information) in BLUP analysis, the accuracy of the breeding values can be increased. DE CARA et al. (2011) proved that with high marker density, molecular information improves results over pedigree data when used to manage populations to maintain genetic diversity via the minimum coancestry contributions methodology.

The objective of this study was to analyse, through computer simulations, the consequences of substituting the pedigree coancestry with the molecular

coancestry, calculated from genome-wide information, in the task of removing exogenous genetic background from an introgressed population.

MATERIALS AND METHODS

The computer simulations comprised three parts: first, the two original populations (native and exogenous) were generated; second, exogenous individuals introgressed native population and resulting offspring mated randomly for a varying number of generations; and third, the mixed population was managed to recover the native background.

Native and exogenous population

Two populations (native and exogenous) of 100 individuals each (50 males and 50 females) were created. The genome of each individual in both populations was made up of 20 chromosomes one Morgan in length with two types of biallelic loci: 2500 markers and 25000 non-markers loci in each chromosome. All loci were equidistant and markers were evenly spaced between the non-marker loci.

Initially, the frequencies of the two alleles for each locus (markers and non-markers) were 0.5/0.5, and alleles were randomly assigned. To create offspring, a Poisson (λ = 1) distributed number of crossing-overs were generated (with no interference) at random positions over the chromosome. The native and exogenous population mated separately at random during 100 discrete generations with constant population size and sex ratio. This generates linkage disequilibrium between markers and the rest of the loci with a different pattern for each of the two populations.

The 2500 markers per chromosome were used in the management for the removal of exogenous background. The non-marker loci were used for evaluation, as alleles coming from the native population were distinguishable from those coming from the exogenous population. Therefore, looking at the non-marker loci we could evaluate which percentage of genome of each individuals comes from native or foreign ancestors, and thus, measuring the efficiency of the de-introgression process.

Exogenous introgression

Once the native and exogenous populations were created, the introgression was simulated. Two types of introgression processes were simulated:

One introgression event. A number (10, 20, 30, 40 or 50) of exogenous individuals joined the native population creating a mixed population of 100 individuals (constant over the generations). This mixed population mated without management (randomly) for one to five discrete generations to carry out the admixture of the exogenous genetic background (Figure 3.1).

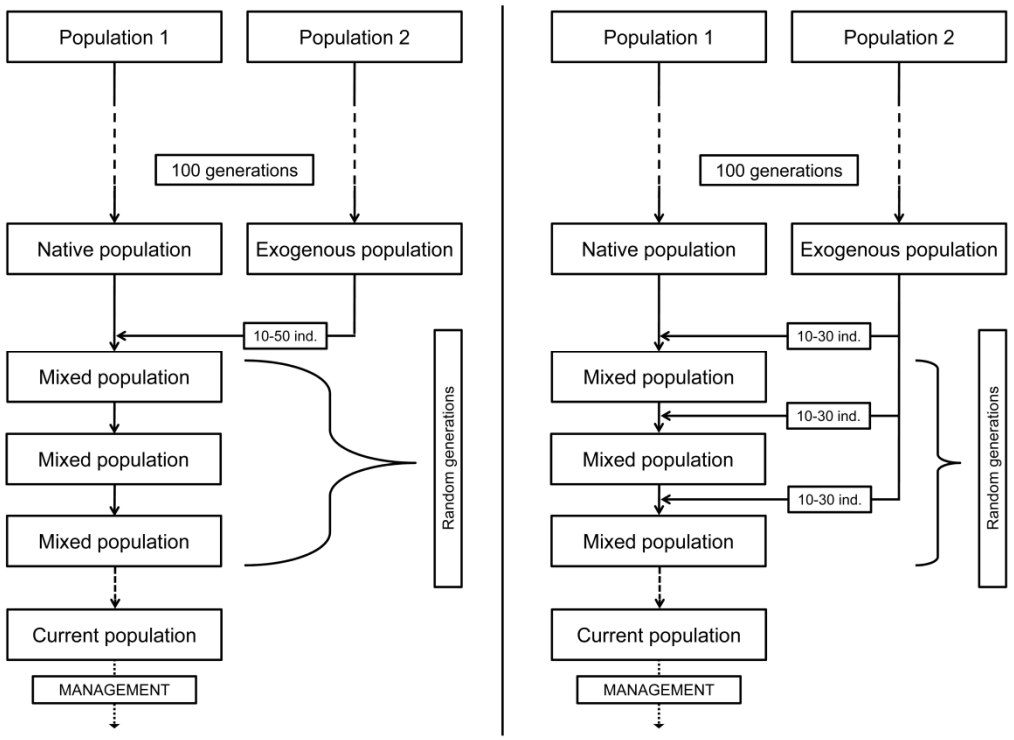

FIGURE 3.1. Design of the two types of simulations. Left: One introgression event. Right: Several introgression events

Several introgression events. A number (10, 20 or 30) of exogenous individuals was added to the population in each of the one to five discrete generations without management (always the same number of exogenous individuals). The population size was kept constant (N = 100) over generations. The individuals required to reach 100, apart from the exogenous, were obtained from the native population in generation one and from the already mixed population thereafter (see Figure 3.1).

Management

After the unmanaged generations of admixture (in both scenarios) 10 generations of management were simulated. To eliminate the exogenous information, the molecular coancestry (calculated from the marker genotypes), between the current and the exogenous individuals introduced in any of the generations was calculated. A correction in the calculation of these coancestries was implemented to eliminate the marker similarity between individuals due to the original frequencies of the two alleles in the base population of both the native and the exogenous population, as follows. Let g_{ij} be the genotype of individual i at SNP j with the values "0", if the individual is homozygote for allele 1, "1" if the individual is heterozygote and "2" if the individual is homozygote for allele 2. Then, the standardized genotype of individual i at SNP j (x_{ij}) can be calculated as in SHEPPERD et al. (2010):

[3.1] $$x_{ij} = \frac{g_{ij} - 2p_j}{\sqrt{2p_j(1-p_j)}},$$

being p_j the frequency of allele "1" at marker j in the base population. A matrix X, composed by the x_{ij} values of the individuals, can be constructed for the current population, as well as for the exogenous individuals. A matrix of genomic relationship between current (c) individuals and the exogenous (ex) can be calculated as:

[3.2] $$A' = \frac{X_c - X'_{ex}}{N_{mark}},$$

where N_{mark} is the total number of markers. To eliminate the exogenous information, on each generation of management, contributions of individuals to the next generation (i.e., percentage of offspring generated by each potential parent) were calculated by minimizing an objective function which includes the relationship between current individuals and exogenous:

[3.3] $$\sum_{i=1}^{N} c_i a'_{i,Ex},$$

where c_i is the relative contribution of individual *i* to the next generation and $a'_{i,Ex}$ is the genomic relationship between individual *i* and all the exogenous individuals obtained from (3.2). As the genomic relationship $a'_{i,Ex}$ is calculated through many markers, it becomes impossible to get two individuals with the same value. For this reason, minimizing the expression (3.3) leads the method to pick up just one male and one female, those with the minimum values of $a'_{i,Ex}$. To avoid this result, a restriction was implemented: each possible parent in the population could only contribute with 10 offspring (of any sex). This entails 20 equally contributing parents each generation, which implies a theoretical rate of inbreeding (ΔF) of 0.025 (assuming random selection and mating). Once the 20 parents were selected, random matings were arranged to create the next generation.

Variables

In every generation several variables were calculated to evaluate the efficiency of the strategies: native founder representation (i.e., the proportion of alleles coming from native founders, based on genomic non-marker information) and average inbreeding coefficient (based on pedigree information) calculated from the beginning of the simulations and, therefore, including the generation of native and exogenous populations and the admixture and management periods. Also the values of ΔF were calculated as:

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}, \quad [3.4]$$

where *t* is the number of the current generation (one to five).

Twenty replicates per scenario were simulated. Results presented are averages across replicates.

RESULTS

One introgression event

Native representation. Results for Native Representation (*NR*) obtained after one or ten generations of management in the one introgression event simulations are shown in Figure 3.2 (upper panel). It can be observed that a noticeable recovery of the native genetic background was obtained by minimizing the

coancestry with the exogenous individuals calculated from the genotype for the markers.

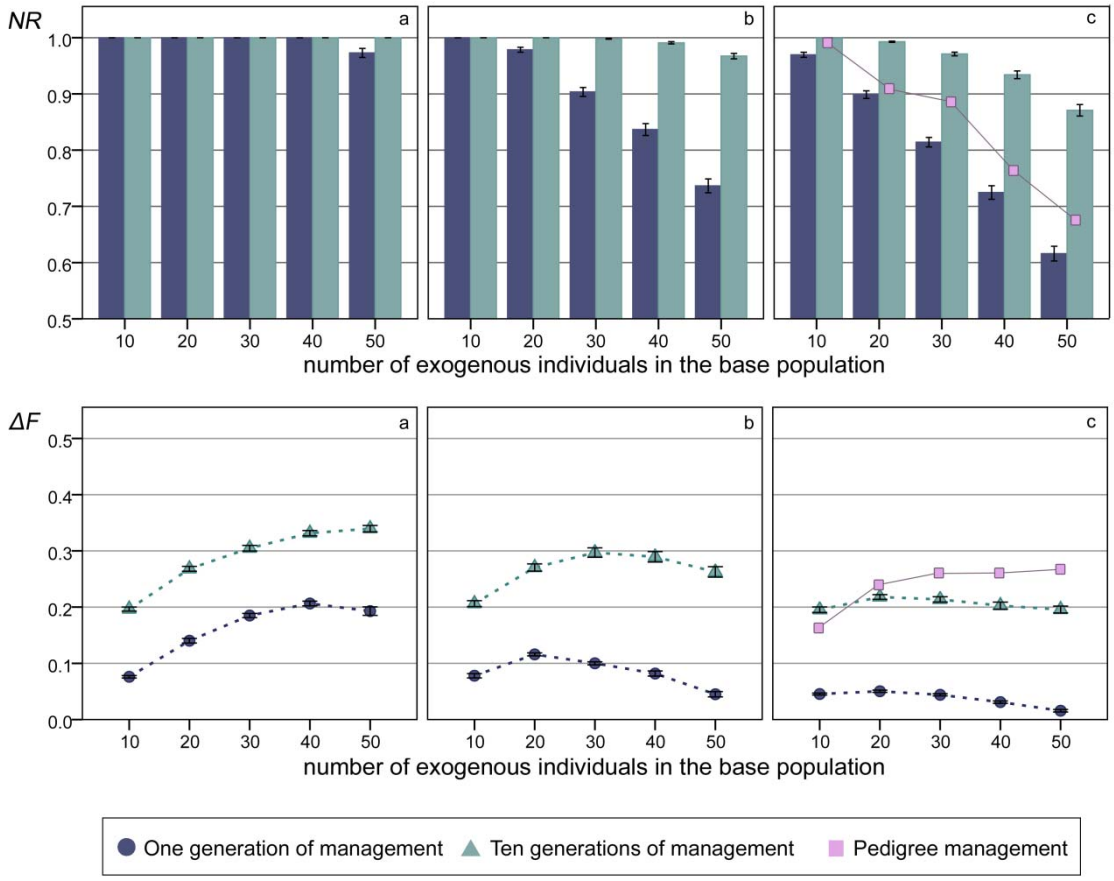

FIGURE 3.2. Native representation and *ΔF* in the one introgression event simulations obtained after one or ten generations of management. a) 1 non-managed generation, b) 3 non-managed generations, c) 5 non-managed generations.

In the cases with just one generation of admixture and those with minimal introgressed information, almost all the removal is achieved in just one generation of management. As the percentage of introgression (and/or time of admixture) increases, the method requires more generations of management to achieve the maximum.

When comparing these results to those obtained when using pedigree data (only in the 5 generations without management scenarios, see Figure 3.2) the use of genetic markers performs markedly better than the pedigree in the task of removing the introgressed genetic material. Notwithstanding, the use of pedigree

data achieves all the removal of exogenous genetic information in just one generation of management (Chapter 1). In this manner, when looking at the results for a single generation of management the use of the pedigree would yield better results. The same levels of removal as using pedigree are reached by the molecular strategy at the second generation of management (not shown), and in subsequent generations marker based management continues to remove exogenous alleles. This eventually results in more removal of exogenous alleles than pedigree based management at the end of the ten generations period.

Inbreeding coefficient. The values of observed ΔF in the one introgression event simulations are shown in Figure 3.2 (lower panel) for the first generation of management, and after ten generations of management. The higher increase of F above the expected for unmanaged populations is a general consequence of the removal methods due to the reduction in the number of contributing individuals. Notwithstanding, the restriction imposed on the maximum contribution per breeder allowed the method to somehow control the increase of F.

The values of ΔF obtained are higher than the theoretical value of 0.025 in the first generations of management, when the removal is larger. This is due to the fact that the 20 contributing individuals can be more related than choosing them completely at random. This fact also explains the observation of higher ΔF for scenarios with a large recovery of native background. After a few generations the maximum removal is almost achieved and the population is more homogeneous regarding the coancestry with the exogenous individuals. At this time the values of ΔF are close to the theoretical value.

The results of F obtained in the pedigree management cannot be directly compared with those obtained in the present study. In the pedigree management the algorithm chose solutions with the largest number of contributing parents when several solutions with the same value for the global coancestry exist. There was an unspecific limitation of the inbreeding, but no explicit restriction imposed on the increase of F (Chapter 1).

Nevertheless, the values of F were similar under both management systems after ten generations, at least for medium and high levels of introgression.

Several introgression events

Native representation. In Figure 3.3 (upper panel) values obtained for *NR* are shown for the scenario involving several introgression events. The recovery of native background in this case was also substantial reaching 100% of recovery in the less introgressed scenarios.

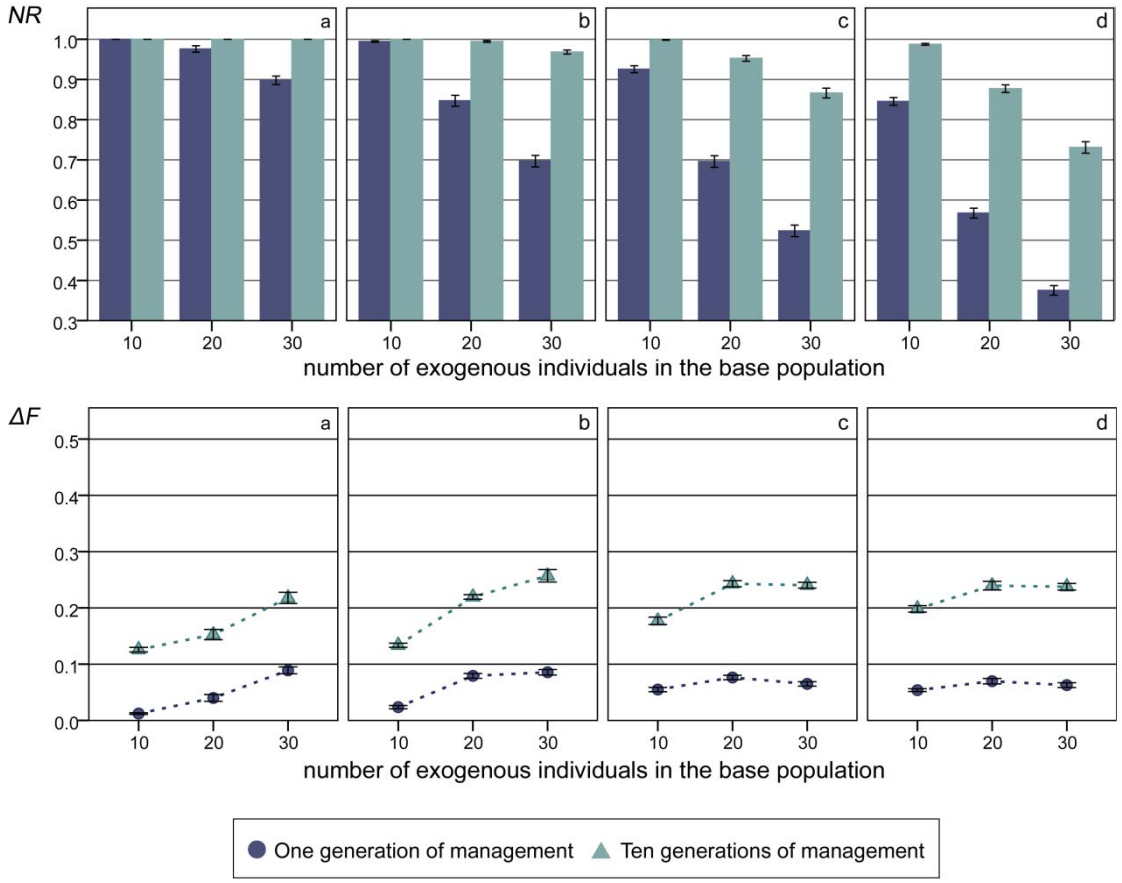

FIGURE 3.3. Native representation and *ΔF* in the several introgression event simulations obtained after one or ten generations of management. a) 2 non-managed generations, b) 3 non-managed generations, c) 4 non-managed generations, d) 5 non-managed generations

As observed in the one introgression event simulations, a great part of the recovery was reached after one generation of management, especially in cases with not much introgression. When the level of introgression is higher, it took more time for the method to reach the maximum de-introgression.

When comparing equivalent cases of one and several introgression event simulations, (i.e., same total percentage of introgression and same number of generations of mixing, see Table 3.1) it is observed that, although the values are similar, in the several introgression scenarios the method recovered more of the native genome.

	NUMBER OF GENERATIONS			
EXOGENOUS PER GENERATION	**2**	**3**	**4**	**5**
10	19.0	27.1	34.4	41.0
20	36.0	48.8	59.0	67.2
30	51.0	65.7	76.0	83.2

Table 3.1: Total percentage of introgression in the current population under the *several introgression events* scenarios, according to the number of generations of introgression and the number of exogenous individuals per generation.

Inbreeding coefficient. As before, the method implied a ΔF larger than the theoretical value in the early generations, which reduced later to the value of 0.025, which was expected (Figure 3.3). The values of ΔF were similar to those obtained in the one introgression event scenarios when the maximum removal was achieved (Figure 2.3).

DISCUSSION

Disadvantages of crossbreeding have been pointed out for economic and conservational reasons highlighting the benefits of maintaining the purity of some populations (RHYMER & SIMBERLOFF 1996; ALLENDORF et al. 2001; DALVIT et al. 2007). Many local breeds have become endangered or extinct because of crossbreeding with more productive breeds (TABERLET et al. 2008). The disappearance of these breeds would be a great loss to the genetic basis of livestock production, specifically regarding their particular adaptations and the possibility to response to changes in the environment or market (TABERLET et al. 2008; WINDIG & ENGELSMA 2010). Actions to preserve these breeds are taken worldwide, but if an undesired

introgression event happens it will be necessary to recover the original background, and develop methods to cope with this situation.

The ability of molecular markers to replace pedigree information to perform different tasks has been proven when the density of markers is large enough (HAYES et al. 2009; DE CARA et al. 2011). The number of markers available in livestock and many other species has increased up to 770000 SNP (cattle). This new technology allows reconsidering the advantages of the use of pedigree information, apart from the necessity of using markers when pedigree information is absent.

In a previous study, the efficiency of recovery of a native background after being introgressed by exogenous individuals was evaluated using the information of a completely recorded pedigree. That study showed that small inputs of exogenous genetics can rapidly spread in the population becoming very difficult to completely recuperate the original genetic background. Pedigree information allowed the recovery in some situations but at the cost of a high increase of inbreeding (Chapter 1).

In the present study, simulated genomic data was used to test molecular marker based methods for their ability for the removal of exogenous genetic material. Marker information was used by replacing pedigree based coancestry with molecular coancestry. The removal of exogenous information in the admixed population using the genome-wide data was successful, particularly in the cases where introgression was limited. The strategy based on molecular markers obtained a higher de-introgression level than the pedigree since it was better at detecting the exogenous genetic material. But advantages appeared after the second generation of management (Chapter 1).

As happened when managing with the pedigree, the values of ΔF showed that each generation of removal involved an extra increase of inbreeding. This shows that we should use the de-introgression method as few generations as possible to avoid this inbreeding effect. In the scenarios with limited introgression, the method required just a few generations to achieve the maximum removal of exogenous genetical material. After this, the management should change towards

minimizing the inbreeding rate through, for example, Optimum Contributions management (CABALLERO & TORO 2000; SONESSON & MEUWISSEN 2001).

When the same percentage of exogenous alleles was mixed in the population progressively (i.e., in several introgression events) the recovery was higher than in the case of a single introgression event. This showed the importance of the time elapsed from the moment the introgression took place to the moment the removal began, as already pointed out. In the multi stage introgression scenario part of the introgression happened more recently than in the case of a single introgression event and, thus, it is easier to remove.

Molecular information also proved to be useful to recover an introgressed population through other approaches based on marker genotypes (Chapter 2). The use of private alleles allowed achieving a substantial recovery of the native background, but required the existence of a large number of molecular markers with alleles exclusive from one population, which is not usually the case. On the other hand, genetic distances were useful when dealing with markers with several alleles, but only in cases where the frequencies were sufficiently different between the native and exogenous populations. The strategies in the present study obtained equal or even better results (regarding the *NR* levels) than the use of private alleles or genetic distances in all comparable scenarios. Improvements can reach up to 15% (relative to using genealogical information) proving that genome-wide information can be more useful and effective for recovering from an introgression event.

Our conclusion is that genome-wide information can be used to remove an introgressed genetic background and to completely recover the native information when the contribution of the exogenous population is limited to 30-40% and the number of generations of admixture is not too high (1-3 generations). The use of molecular coancestry to perform this task proved to be an effective tool to recover the native genome, and the recovery was higher than when using pedigree information (Chapter 1) or a small number of markers (Chapter 2). The availability of genome-wide information in natural populations is not as high as it currently is in farmed animals, so it is expected that our method can be most easily applied in the latter. The characteristics of each particular situation must be studied and,

even with relatively few SNP, the molecular coancestry can be calculated and used in the absence of a proper pedigree for de-introgression. Regardless of the possibilities for recovery when dealing with introgression, it is always essential to avoid as much as possible undesirable exogenous inputs of genetic material, because the de-introgression process involves an increased rate of inbreeding that can result in a significant cost for an endangered breed.

CAPÍTULO 4

Estimation of genomic breed proportions to remove introgressed genetic material: an example in sheep

INTRODUCTION

Crossbreeding is a strategy commonly used in livestock and many wild populations. The increase in fitness that a new genetic input can infuse has been widely studied, and mixing breeds has been used to increase genetic variability and the performance for productive traits (FRANKHAM et al. 2002; SCHAEFFER et al. 2011). Nevertheless, crossbreeding can imply some disadvantages as outbreeding depression, loss of adaptability and finally, lead to the extinction of some populations (ALLENDORF & LUIKART 2007).

Undesired introgression of genetic material occurs in both livestock and wild populations. In livestock, crossbreeding with more productive breeds has led local breeds to lose their specific singularities and adaptive traits such as disease resistance, adaptation to a specific climate or harsh conditions (TABERLET et al. 2008). Some of these breeds are endangered because of this, and they should be recovered in order to avoid extinction (UGARTE et al. 2001; MORAIS et al. 2005; TABERLET et al. 2008). In wild populations, admixture happens more frequently than expected and it is increasing due to human activities (RHYMER & SIMBERLOFF 1996; ALLENDORF et al. 2001). Some wild populations are endangered for the widespread of their domestic relatives, that causes loss of biodiversity through introgressive hybridization (RANDI 2008). In some cases, populations have become endangered because of the genetic inputs of exogenous individuals, and some actions should be taken to recover their native backgrounds. Examples have been described in a variety of species such as cattle (PADILLA et al. 2009), partridge (BARBANERA et al. 2011), trout (HOHENLOHE et al. 2011), mink (CABRIA et al. 2011), salamander (BAYER et al. 2012), etc.

In previous studies, different sources of information were used to recover the genetic background of a population that suffered undesired introgression. This process was called de-introgression. Several introgression scenarios were simulated where the objective was to remove the exogenous genetics by minimising the genealogical coancestry (Chapter 1), the molecular coancestry calculated through genome wide information (50000 SNP) (Chapter 3) and identifying the probability of origin with multiallelic markers (Chapter 2)

Removing exogenous material using only genealogical coancestry obtained good results. However, it required a completely recorded pedigree, which is unlikely to exist in wild populations. Minimising the molecular coancestry calculated through genome wide information provided the best results regarding the amount of native genomics recovered. The success in the recovery when using other kind of markers depended on the number of markers and allele frequencies. Hence, exclusive markers (i.e., having private alleles occurring only in native or exogenous populations) obtained good results even with a medium number of markers (5 to 10), and more markers were required when alleles were segregating in both native and exogenous populations at intermediate frequencies (i.e., more similar populations). The number of markers available was a major limiting factor and essential to choose the appropriate method when facing the task of recovering a native background (Chapter 2). Moreover, this approach requires a perfect knowledge of the genetic conformation (the true allelic frequencies) of the original pure populations. A side effect of all three methods is the high increase of inbreeding due to the restriction in the number of individuals contributing to the next generations.

In this study, a real data set of 6000 sheeps was used. The genotypes (OvineSNP50 BeadChip) of individuals of two pure breeds (Merino and Poll Dorset) and F1 crosses of these breeds were used to evaluate the ability of two alternative methods to determine the level of similarity of the crosses to the pure breeds. The first method was originally developed in cattle to detect segments of zebu and taurine origin by BOLORMAA et al. (2011). In the present study, we applied the method to classify observed haplotypes in the F1 crosses as coming from Merino or Poll Dorset. The second method, described in VANRADEN et al. (2011), used a linear model to predict breed identity through the genomic relationship matrix between the pure and crossed individuals. Several scenarios of admixture were simulated using the real genotypes as a based population. Then, the simulated genotypes were used to test the ability of the two mentioned methods to remove the foreign genetic information from a mixed population. Additionally, to avoid a large increase of inbreeding, the strategies included a restriction on the number of individuals contributing.

The first objective of this work was to analyse the efficiency of these two methods to detect Merino/Poll Dorset segments in real F1 crossbred individuals. The second objective was to evaluate the ability of the methods in restoring the Merino background after an introgression process by removing the maximum Poll Dorset genetics in the individuals of the population, as well as comparing these results with those obtained in previous studies.

MATERIALS AND METHODS

The data used was obtained from the Australian Cooperative Research Centre for Sheep Industry Innovation (Sheep CRC). Individuals from three populations were used: 4964 pure Merino, 188 pure Poll Dorset and 811 crosses (all of them 50% Merino and 50% Poll Dorset). The pedigree of the individuals was available and used to calculate the genealogical coancestries.

The individuals were genotyped using the OvineSNP50 BeadChip (Illumina, San Diego, CA), which reacts to 54977 single nucleotide polymorphisms (SNP). The following quality control measures were applied: SNP were removed if they had a call rate of < 95%, an Illumina Gentrain (GC) score of < 0.6, a minor allele frequency of < 0.01, were out of Hardy–Weinberg equilibrium (a *P*-value cut-off of 1^{-15}), had no genome location or were in > 0.99 r^2 with another SNP on the chip. After these measures were applied, 48599 SNP were used. Data for genotyped animals were removed if their genotype call rate was < 0.9 or if their mean heterozygosity was > 0.5, which would indicate sample contamination. Sporadic missing SNP were imputed using Beagle (BROWNING & BROWNING 2009). The data was phased using ChromoPhase (DAETWYLER et al. 2011).

Breed origin description

A principal component analysis of the genomic relationship matrix (G), calculated as in YANG et al. (2010), was performed using the R version 2.11.1 (R DEVELOPMENT CORE TEAM 2010) to prove the ability of the genomic information to differentiate between the three groups of individuals. After that, two methods were developed to predict breed proportions in the crossed individuals:

Haplotypes approach. The haplotypes approach is described in BOLORMAA *et al*. (2011). Each chromosome is divided into nonoverlapping segments consisting of

10 consecutive SNP. We estimated the probability of a segment *i* being of Merino origin (b_{Mi}) for each of the up to 2^{10} possible haplotypes for a specific segment as:

[4.1] $$b_{Mi} = \frac{p_{Mi}}{p_{Mi} + p_{PDi}},$$

where p_{Mi} is the frequency of the *i*th haplotype in the pure Merino animals and p_{PDi} is the frequency of the *i*th haplotype in the pure Poll Dorset animals. Haplotypes with a *b* value lower than 0.4 were classified as Poll Dorset and those with a *b* value higher than 0.6 were classified as Merino. The remaining haplotypes were left unassigned.

GBLUP approach. The GBLUP approach is described in VANRADEN et al. (2011). A genomic evaluation was performed using ASReml (GILMOUR et al. 2002) to predict Merino proportions in the individuals using the model:

[4.2] $$\mathrm{y} = \mu 1 + \mathrm{Zg} + \mathrm{e},$$

where y is a vector with the proportion of merino of the animals (as phenotypic record), μ is the intercept, Z is a incidence matrix relating animal effects to phenotypes, g is a vector of additive genetic effects, and e is the vector of residuals. The following distributions were assumed: $\mathrm{g} \sim N(0, \mathrm{G}\sigma^2_g)$ and $\mathrm{e} \sim N(0, \mathrm{I}\sigma^2_e)$. G was a genomic relationship matrix, calculated as in YANG et al. (2010).

The pure Merino and pure Poll Dorset animals were used as training set with phenotypes coded 1 and 0 respectively. Then, the Merino proportion was predicted in the three populations (pure Merino, pure Poll Dorset and crossbred individuals). A heritability of 99% was assumed for the trait *Merino proportion.*

De-introgression process

The pure Merino and pure Poll Dorset genotype data set were used to simulate several introgression scenarios from which the Merino genetic background was intended to be recovered (Figure 4.1). A population of 100 individuals was created using the real individuals, with a variable number of Poll Dorset (10, 20, 30, 40 or 50) and the rest (up to 100) being Merino. Sex was randomly assigned to the genotypes used to create the individuals (50 males and 50 females). The mixed

population mated randomly during 1, 3 or 5 generations to produce different levels of admixture of the Poll Dorset genetic information into the Merino background. Five generations of management were then simulated using two different approaches to remove the Poll Dorset genetics. The real origin of the alleles was known in all generations and used for evaluation of the efficiency in the recovery. The pedigree was also known and used for calculating coancestries and inbreeding. Twenty replicates per scenario were simulated.

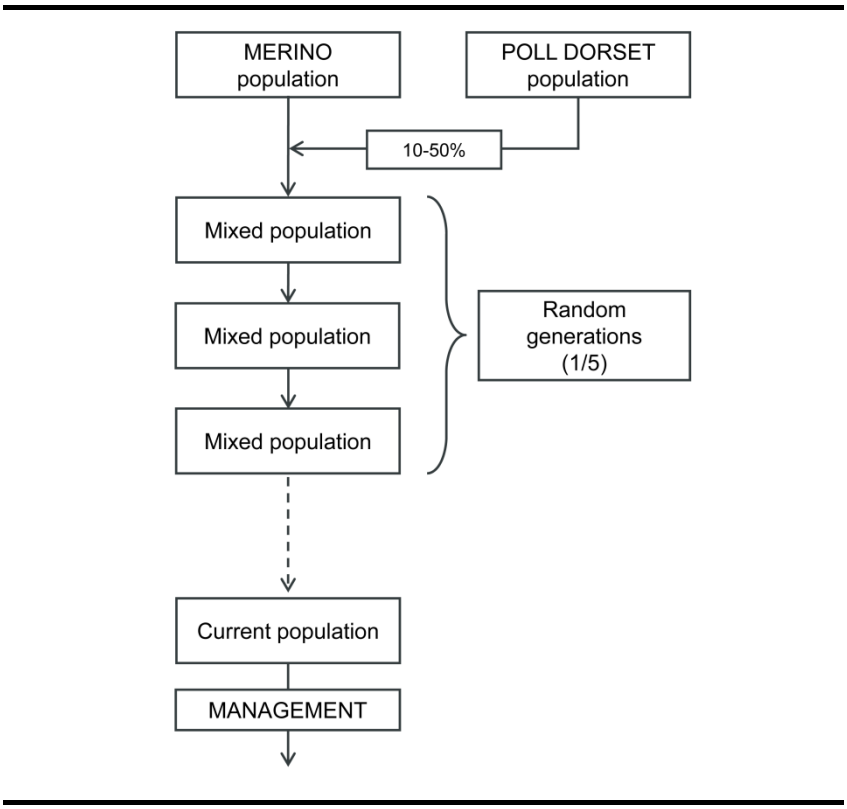

FIGURE 4.1. Diagram of the introgression simulation. The real individuals were used to create a mixed population of 100 individuals that mated randomly during 1 to 5 generations. Afterwards, five generations of management started.

Sheep genome is made up of 26 chromosomes. The number of SNP per chromosome is showed in Figure 4.2. Each chromosome was considered to be 1M and, to create gametes, a Poisson distributed ($\lambda = 1$) number of crossing-overs (one crossover is expected on average on each Morgan) with no interference were generated in random positions over the chromosome.

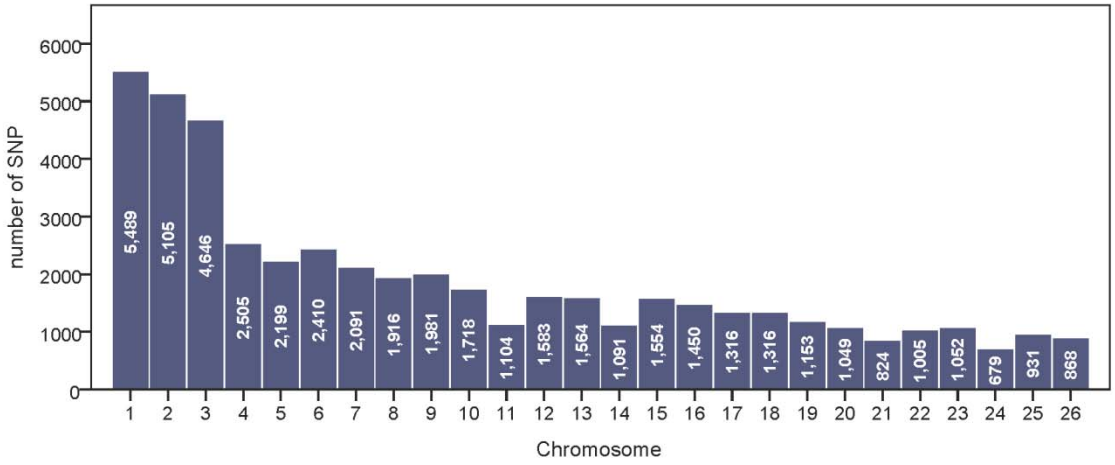

FIGURE 4.2. Number of markers per chromosome.

Management

To remove the Poll Dorset information two approaches were used:

Haplotypes approach. A random sample of 188 pure Merinos and the 188 pure Poll Dorset genotypes were used for training the haplotypes approach. In each generation, haplotype *b* values were computed for all individuals using equation [4.1], and a mean *b* value per individual was calculated. Those individuals with largest mean *b* values were assumed to be the purest Merinos.

GBLUP approach. A random sample of 188 pure Merinos (the same subset as the haplotypes approach each replicate) and the 188 pure Poll Dorset genotypes were included with phenotypes (Merino phenotype coded as 1, Poll Dorset phenotype coded as 0), together with the genotypes for the selection candidates in each generation. Breed proportions were then predicted using model [4.2]. The predictions were used to identify the purest Merino animals in each generation of management.

Exogenous information was eliminated in each generation of management by choosing the 10 purest animals per sex (i.e., the 20 individuals with highest number of Merino haplotypes, or highest Merino proportion) to equally contribute to the next generation (10 offspring each). Individuals were mated randomly. The above procedure implies a theoretical rate of inbreeding (ΔF) of 0.0125 (assuming random selection and mating).

Monitoring

In every generation two variables were calculated to evaluate the efficiency of the strategies: 1) percentage of Merino (i.e., the real proportion of alleles coming from Merino founders) and 2) inbreeding coefficient (*F*). The *F* values were calculated considering the real coancestries between the original individuals (from the real pedigree) and the genealogy of the admixture and management periods. From these, values of ΔF were calculated as:

[4.3] $$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}},$$

where *t* is the number of the current generation (one to five).

Extra simulations

Extra simulations were carried out to cover more extreme situations, including a longer period of admixture and lower selection pressure (i.e., forcing more individuals contributing). The number of individuals actually contributing was increased to 40 (i.e., 20 males and 20 females, 5 offspring each) in scenarios with 5 and 20 generations of admixture, to check how the removal of Poll Dorset genetic material was accomplished in more extreme conditions.

RESULTS

Breed origin description

Principal component analysis. Figure 4.3 shows a plot of the first two principal components (PC, those with larger eigenvalues). It can be observed that the first PC already separates the two breeds and the crosses. The crosses are situated half way between Merino and Poll Dorset groups. A further level of division can be observed within the Merino breed using the second PC. This differentiation is due to two large half-sib groups of Merino selection lines.

Haplotypes. The distribution of the haplotypes in chromosome 1 of 25 sampled pure Merinos, 25 sampled pure Poll Dorset and 25 sampled crossed individuals is shown in Figure 4.4. In the figure, each line shows one chromosome and, thus, each individual is represented by two consecutive lines. The proportion of unassigned haplotypes was 0.049 ± 0.001 in all the groups. Most of the haplotypes in the pure

Merino individuals were classified Merino and in the pure Poll Dorset most were classified Poll Dorset. The distribution of the type of segments in the crossed individuals shows that all of them are F1 crosses because they have one entire chromosome coming from Merino and the other one coming from Poll Dorset.

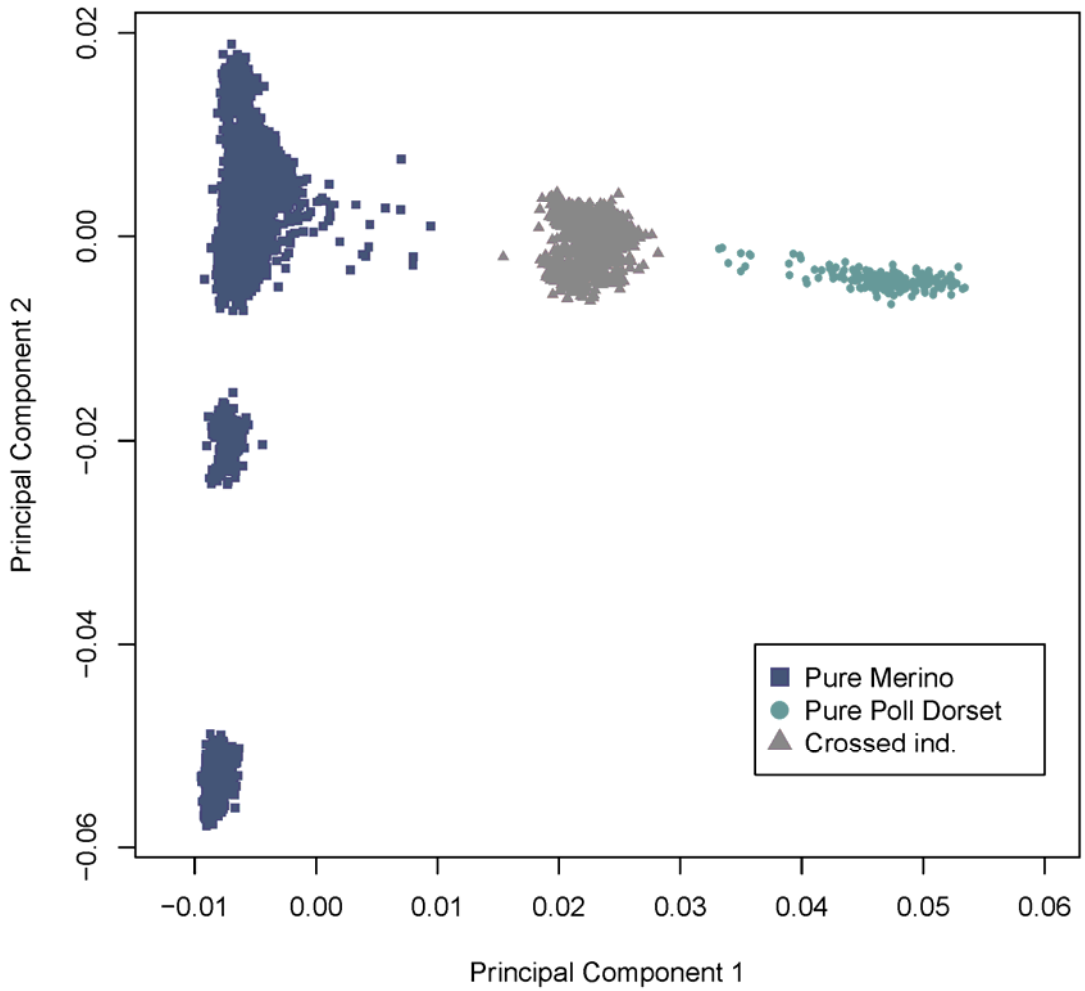

Figure 4.3. Plot of the different individuals according to PC 1 and PC 2.

The distribution of the *b* values of the 10 SNP segments in the three groups of individuals (all over the genome) is shown in Figure 4.5. As stated before, most of the segments in the Merino group are recognized as Merino (90% of the segments with a *b* value > 0.6) and most of the segments in the Poll Dorset group are recognized as Poll Dorset (91% of the segments with a *b* value < 0.4).

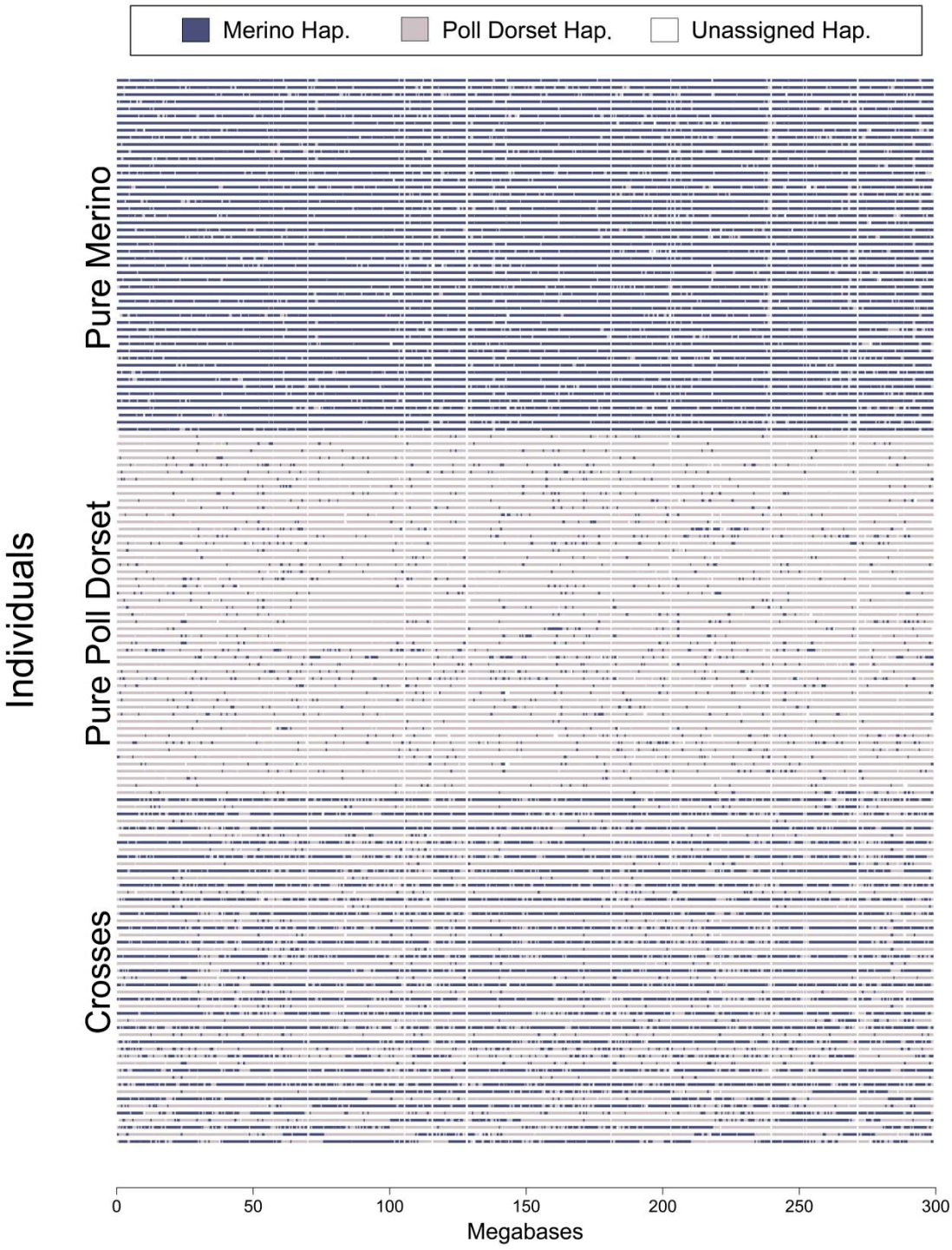

FIGURE 4.4. Plot of the origin of the 549 haplotypes of Chromosome 1 in a sample of 25 random pure Merino, 25 random pure Poll Dorset and 25 random crossbred animals. Each line represents an individual's chromosome.

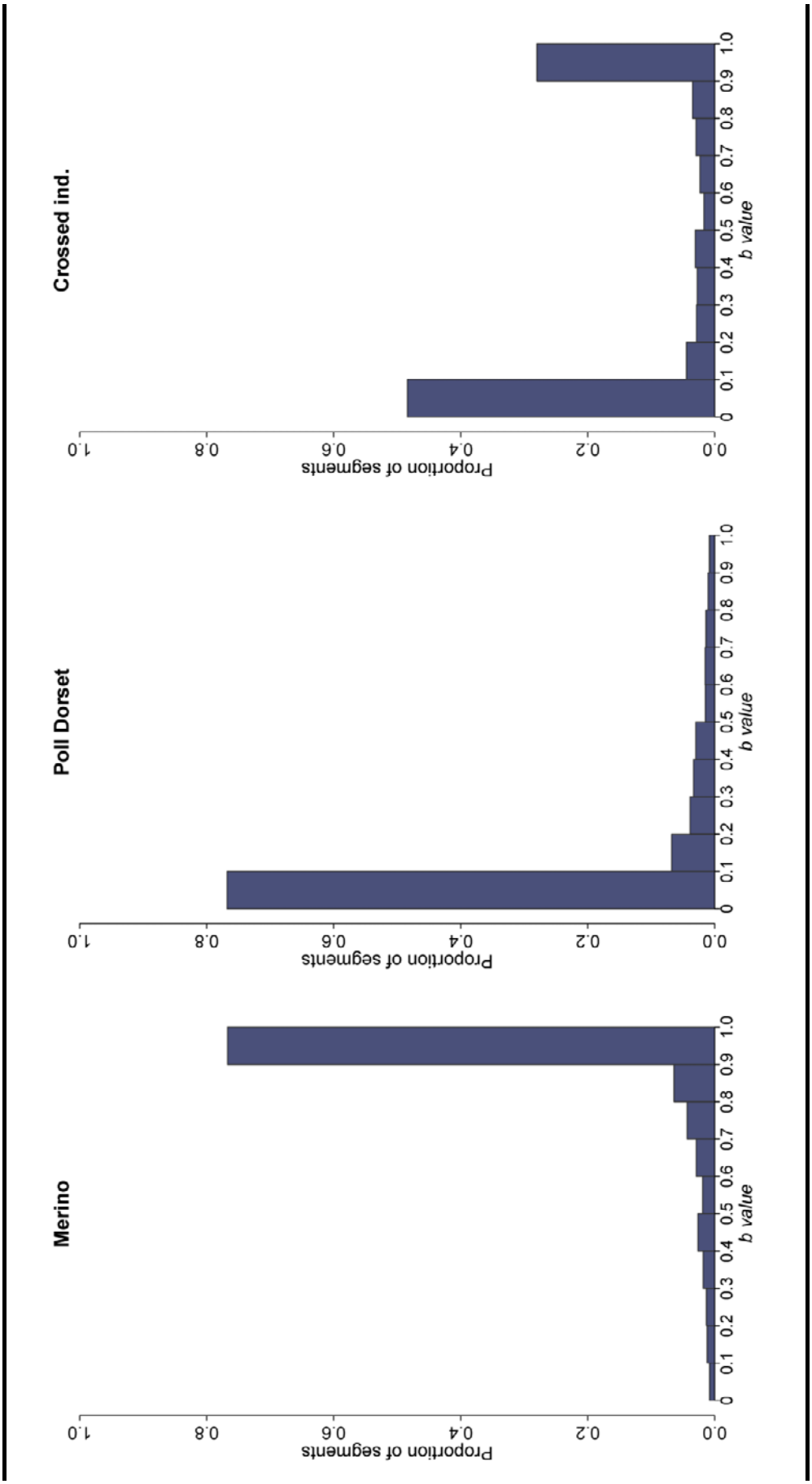

FIGURE 4.5. Distribution of the *b* values for 10 SNP segments across the 26 ovine chromosomes in pure Merino, pure Poll Dorset and F1 crossed individuals.

The results in the crossed animals showed a mixed pattern with 58% of the segments considered Poll Dorset and 37% of the segments considered Merino. The higher number of Poll Dorset segments in the crosses could suggest that the Poll Dorset individuals are not as pure as expected but crossed with Merino. Thus, some Merino segments, present in Poll Dorset individuals are recognized as Poll Dorset.

GBLUP. The classification of individuals from the GBLUP genomic prediction of Merino proportion is shown on Figure 4.6 compared with the mean *b* value of each individual. Data show that the GBLUP method is also able to separate the three groups of individuals. The correlation between the GBLUP solution and the *b* values was 0.99 considering the whole set.

De-introgression process

Results for the proportion of Merino recovered (*MR*) for the haplotype approach and GBLUP approach were very similar in all scenarios, consequently, only the results for the haplotype approach are presented.

The values of *MR* obtained after one or five generations of management in the different introgression scenarios are shown in Figure 4.7 (upper panel). The Merino background was almost completely recovered in all the scenarios after five generations of management. The recovery was nearly completed in the first generation of management when the percentage of Poll Dorset introgression was small and only mixed for one or three generations, but it took longer in the scenarios with more complex introgression.

The results of inbreeding obtained after managing one or five generations with the haplotypes approach are shown in Figure 4.7 (lower panel). The values of ΔF obtained were close to the expectation (ΔF_1 = 0.01) in the first generation of management, but they became higher than the expectations in the subsequent generations ($\Delta F_{2, 3, 4, 5}$ = 0.035). This is because ΔF was calculated assuming random selection and mating in the parents. Selecting the purest individuals resulted in individuals that were more related than if they were selected at random after the second generation of management.

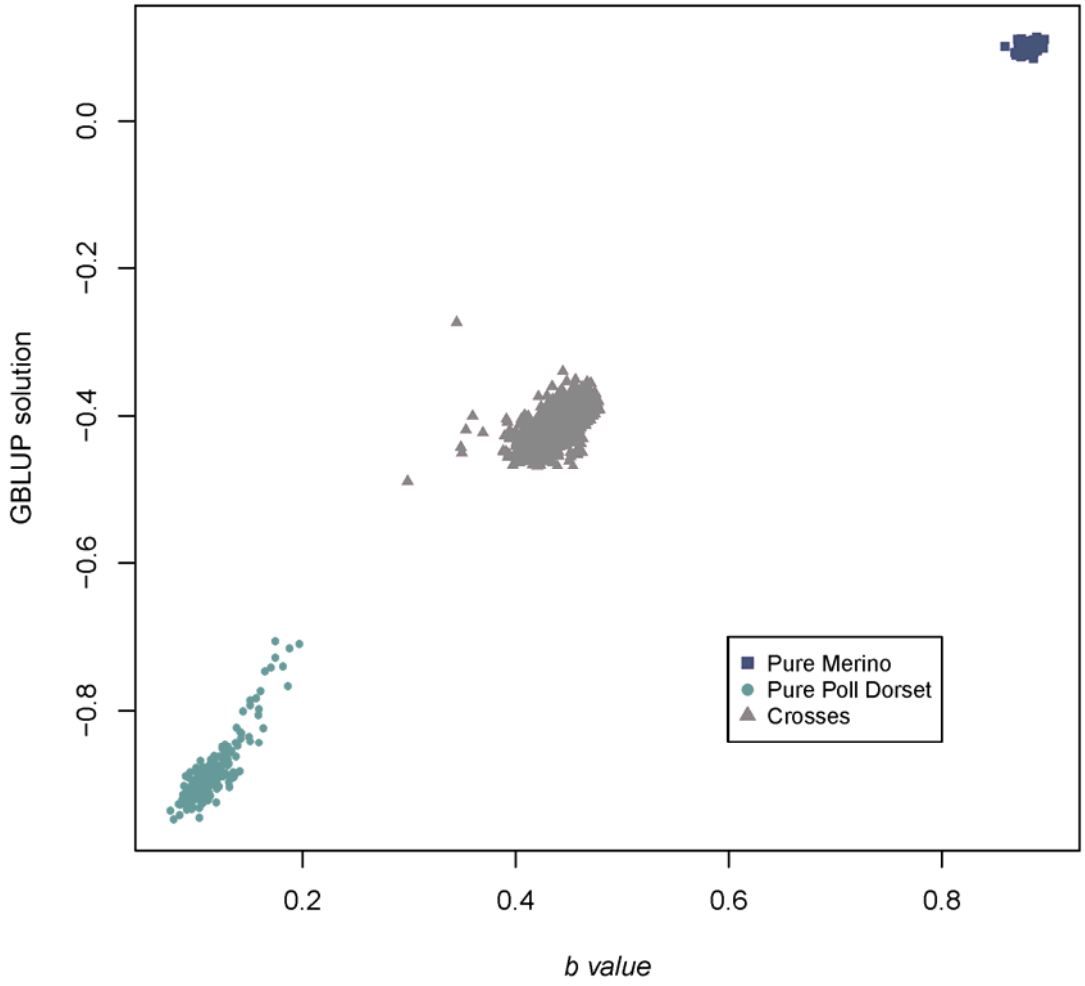

FIGURE 4.6. Comparison of the GBLUP solutions and the haplotypes results.

The results of inbreeding were slightly different when using the GBLUP approach (not shown). The accumulated ΔF after five generations of management is lower when using GBLUP (ΔF_{hap} = 0.15, ΔF_{GBLUP} = 0.13). The differences in ΔF appeared in the final generations, once the maximum Merino proportion was achieved, but the method is still being applied. If estimated proportions are similar for all available individuals, the selected breeding animals are chosen at random not involving a higher level of relationship between them and, thus, ΔF is closer to expectations. In contrast the haplotype based method still detected differences between candidates related to common origin leading to higher ΔF.

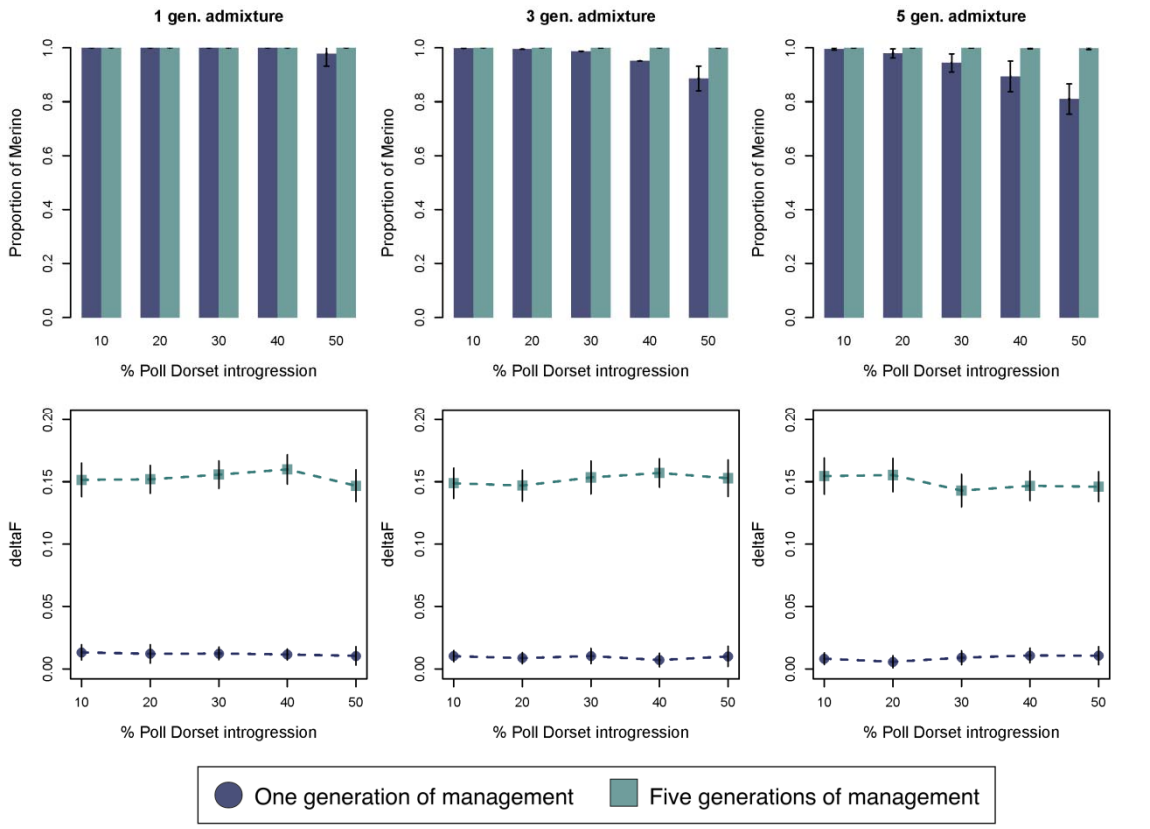

FIGURE 4.7. Proportion of Merino recovered (upper panels) after one or five generations of management and ΔF (lower panels) after one or five (cumulative ΔF) generations of management using the haplotypes approach, for 1, 3 or 5 generations of admixture (20 individuals contributing).

Results obtained when doubling the number of contributing individuals to 40 (haplotypes approach) for 5 generations of admixture are shown in Figure 4.8 (left). The *MR* obtained in these simulations was considerably lower for one generation of management, due to the smaller selective pressure. Consequently, the method requires the five generations of management to achieve the maximum recovery but with a smaller increase of inbreeding.

A more extreme situation was tested in which exogenous material was admixed for a longer period of 20 generations, but trying to maintain more genetic diversity by using 40 contributing individuals. The results (Figure 4.8, right) prove that the method can use the haplotypes information even when the admixture time is long, and reach up to 80% of Merino in five generations of management with an acceptable level of inbreeding. As it happened before, the *ΔF* in the first generation

of management is lower than the expectation, but it becomes more similar to the expected values in the following generations of management.

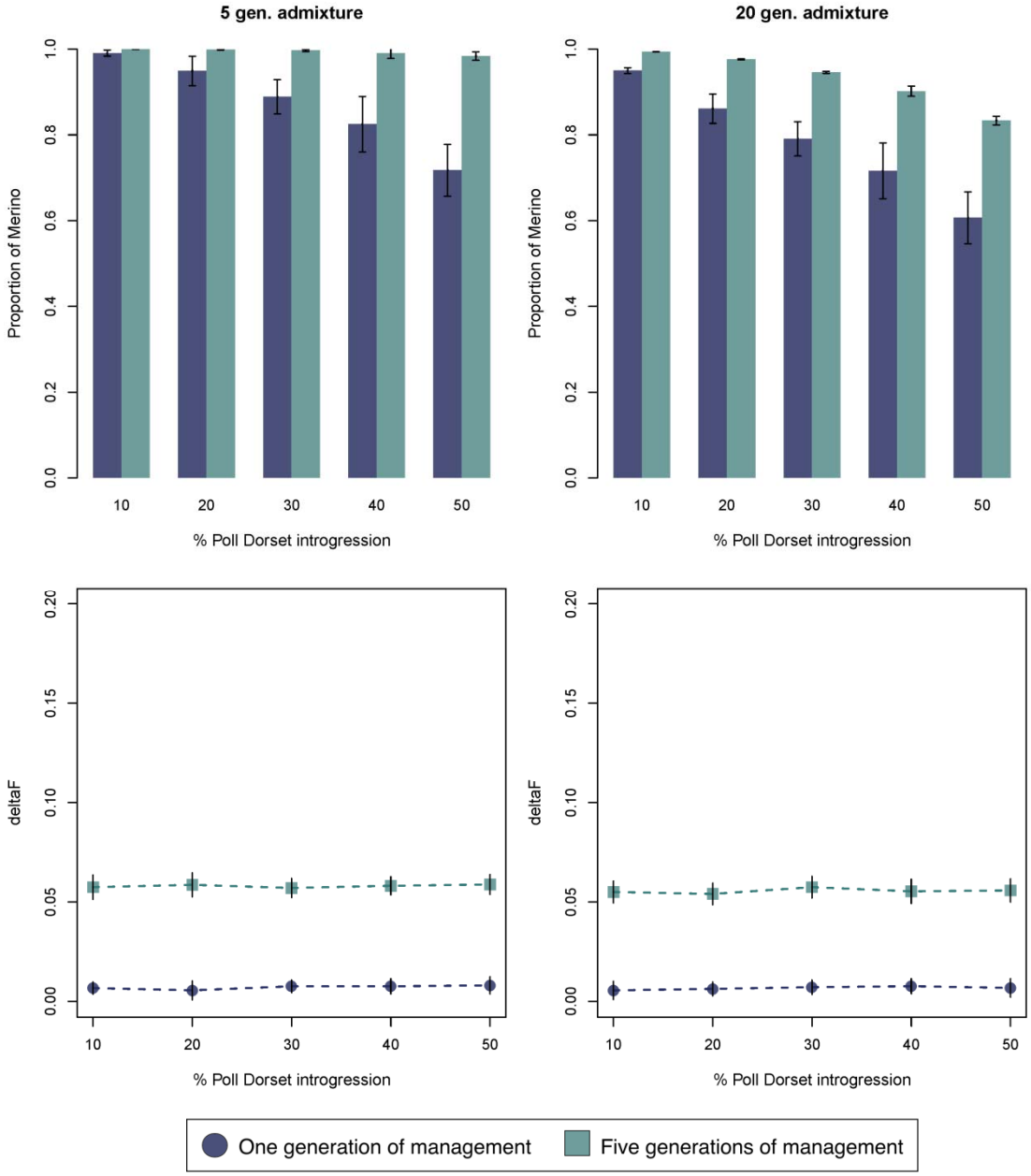

FIGURE 4.8. Proportion of Merino recovered (upper panels) after one or five generations of management and *ΔF* (lower panels) after one or five (cumulative *ΔF)* generations of management using the haplotypes approach for 5 or 20 generations of admixture (40 individuals contributing).

DISCUSSION

Disadvantages of crossbreeding have been studied in livestock and wild populations (ALLENDORF et al. 2001; DALVIT et al. 2007). Many populations are close to extinction due to hybridization and genetic introgression, and some of this introgression could be reversed if tools for detecting the purest individuals are available.

In previous studies, the ability to recover native genetic information after introgression events of pedigree based methods (Chapter 1), and different types of molecular markers (Chapters 2, 3), was analysed in several simulated scenarios. In all cases it was shown that small inputs of exogenous genetics rapidly spread in the native background. Besides the amount of exogenous alleles introgressed, the time elapsed from the introgression until the management started was crucial to determine the probability of recovery. When it was long (5 generations of admixture) it became very difficult to completely recuperate the original genetic background.

In the present study we used a robust data set to prove the ability of two methods to find out the origin of genetic information in an admixed population from two different breeds, even after several generations of introgression. Results showed that both methods were able to detect the Merino and the Poll Dorset genetic fragments.

Both the GBLUP solutions and *b* values differentiated between pure and crossed individuals (Figure 4.6). The correlation between the prediction of the Merino content using GBLUP and *b* values was 0.99 in the whole population, but just 0.54 when calculated for the crossed individuals. Despite this medium correlation, Figure 4.9 shows a clear tendency of higher *b* values implying higher GBLUP solutions. This was confirmed when evaluating their performance in the de-introgression process. The *MR* was the same for both methods, suggesting that they chose similar individuals according to their Merino proportion.

Regarding the speed of recovery, the performance of the methods was very similar to what was observed in previous studies (Chapters 1, 2, 3). Most of the recovery was achieved in the first generation of management, but the most

introgressed scenarios required one or two more generations. On the other hand, the strategy implied an increase of inbreeding due to the number of individual contributing, and their relationships.

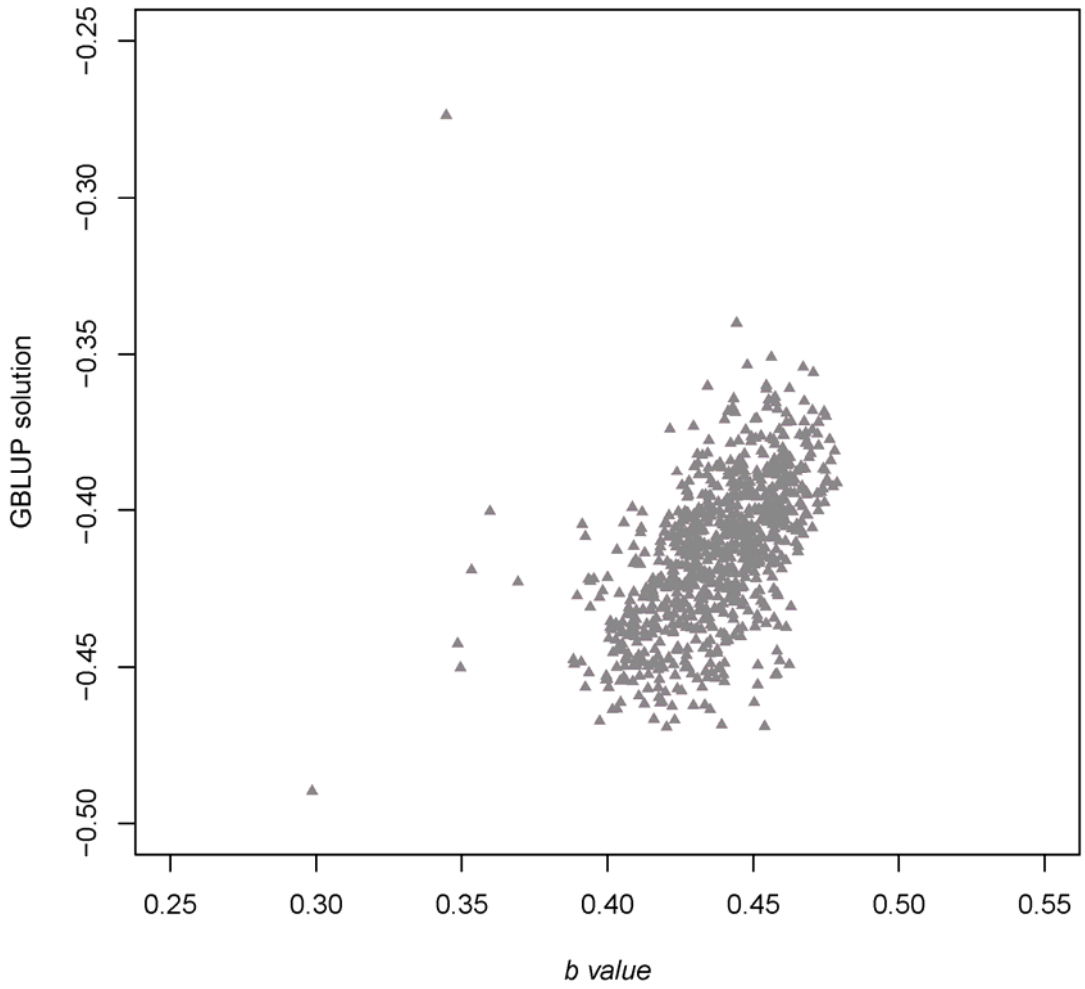

FIGURE 4.9. Comparison of the GBLUP solutions and the haplotypes results in the crossed individuals.

The GBLUP and haplotypes approaches obtained better results than using pedigree, or microsatellite-like markers (Chapter 1, 2). It is expected that using genome wide information outperforms the use of the pedigree because the latter gives average expected values while the former provides the particular realisations in every region of the genome in linkage disequilibrium with the markers, as it has been revealed in different studies (HAYES et al. 2009; DE CARA et al. 2011).

Nevertheless, the results obtained in the present study were also much better than those obtained by minimising the molecular coancestry calculated from 50000 SNP on simulated genomes (i.e., also using genome-wide information) and presented in Chapter 3. This suggests that the GBLUP and haplotypes approaches could be better than minimisation of molecular coancestry. However, the datasets differed between the two studies. While the molecular coancestry approach was tested on simulated data, the GBLUP and haplotypes approach here have been tested on real data.

A key to the success in these kind of studies is the differentiation among populations. If the populations that mixed are too similar, it will be very difficult to differentiate between both genetic backgrounds, and, thus, de-introgression will be more difficult to achieve. If both populations are clearly different, the task will be easier to accomplish.

The genetic differences between Merino and Poll Dorset populations became clearer in the second set of scenarios of introgression. Even when a 50% of Poll Dorset introgression occurred, and after 20 generations of admixture, the haplotypes and GBLUP approaches were capable of recovering up to 80% of Merino genetics (Figure 4.8), even when 40 individuals were force to contribute. This was not a very strict selection pressure, however, the recovery was successful in five generations of management. This can happen because both populations are different enough, and even after 20 generations of admixture the methods are still capable of recognizing the Merino or Poll Dorset segments. This does not mean that the simulated data in Chapter 3 is not realistic, but it represents a more restrictive situation (more similar populations and more difficulty to recover). Populations susceptible of undergoing undesired introgression could be more similar to each other than Merino and Poll Dorset, or than the simulated ones. But certain levels of recovery are to be expected when using the methodology presented in this study, being the particular results depending on the similarity between populations.

The current methods heavily rely on a group of genotyped pure individuals to train *b* values and GBLUP predictions. The haplotypes approach cannot be applied without some native and exogenous pure information. In contrast, the GBLUP

approach could still be applied with a sample of just one group, native or exogenous, where the proportions of pure genetics was known. While its efficiency would decrease, it would still be able to remove exogenous genetic material.

The methods here described are designed for populations on which a SNP chip has been developed. Its availability in domestic species increases every day, especially due to genomic selection studies (MEUWISSEN et al. 2001), and panels of up to 770 K are available in cattle (LENSTRA et al. 2012). This density of markers is not available in wild species, nevertheless, new approaches for discovering SNP are being developed thanks to next generation sequencing (DAVEY et al. 2011). Methods such as reduced-representation libraries (RRLs) and genotyping-by-sequencing (GBS) can be used in species without a reference genome, for SNP discovering on a small scale, being cheaper and more feasible methods than developing SNP arrays (VAN TASSELL et al. 2008; ELSHIRE et al. 2011)

The increase of inbreeding every generation of management is a side effect of the method that cannot be avoided, but it can be controlled. As shown, a bigger number of individuals contributing decreases ΔF and still removes exogenous information. Increasing the number of individuals selected to contribute, or setting an explicit restriction on ΔF could lead to an acceptable recovery without losing too much genetic diversity.

DISCUSIÓN

La introgresión de genes exógenos en diferentes poblaciones es un fenómeno habitual en la naturaleza y que puede contribuir a los procesos de adaptación y especiación de muchos taxones con importantes consecuencias evolutivas (FRANKHAM et al. 2002). También en ganadería, las razas locales se cruzan comúnmente con otras más productivas para incrementar su valor económico (UGARTE et al. 2001).

Sin embargo, existen muchos ejemplos en los que un aporte exógeno de material genético se considera negativo. Las modificaciones en los paisajes, así como cambios directos en la localización de individuos, ambos debidos a la acción humana, aumentan el porcentaje de mezcla de poblaciones y tienen graves repercusiones sobre la biodiversidad (VITOUSEK et al. 1997; CRISPO et al. 2011). Muchas especies salvajes se encuentran amenazadas por sus equivalentes domésticos o especies introducidas (RANDI 2008). La pérdida de combinaciones alélicas (base de adaptaciones locales) y otros procesos implican la aparición de depresión híbrida que puede llevar también a la desaparición de poblaciones (ALLENDORF & LUIKART 2007; EDMANDS 2007).

En especies domésticas, muchas razas tienen intereses económicos ligados a la preservación de su fondo genético en pureza (DALVIT et al. 2007). Además la introducción de individuos exógenos puede ser perjudicial para una población, perdiéndose sus adaptaciones particulares como resistencia a enfermedades o a ambientes extremos (TABERLET et al. 2008).

Como se ha visto, la conservación de poblaciones en pureza es justificable tanto por motivos económicos como para evitar la pérdida de diversidad genética. Sean cuales sean los motivos, la introducción indeseada de material exógeno en una población cuyo fondo genético sea interesante preservar puro, implica la necesidad de recuperarlo mediante un proceso de depuración o *desintrogresión*. Para ello es necesario analizar el contexto particular de las poblaciones involucradas, así como la información disponible para poder llevar a cabo dicha recuperación.

En la Tabla D.1 se muestra un resumen de las estrategias de manejo estudiadas a lo largo de esta Tesis en función de la información disponible, así como el capítulo en el que se describen.

GENEALOGÍAS	
Mínimo parentesco con los exógenos (MEC)	Capítulo 1
Mínimo parentesco parcial debido a los exógenos (MPC)	Capítulo 1
INFORMACIÓN MOLECULAR NO DENSA	
Selección por alelos diagnóstico	Capítulo 2
Distancias genéticas	Capítulo 2
INFORMACIÓN MOLECULAR DENSA	
Mínimo parentesco molecular con los exógenos (MECG)	Capítulo 3
GBLUP	Capítulo 4
Determinación del origen de haplotipos	Capítulo 4

Tabla D.1. Estrategias de manejo estudiadas en función de la información disponible.

Como se ha visto en todos los estudios, la posibilidad de recuperación de los genes nativos se ve limitada tanto por la cantidad de información exógena que entra en la población, como por el tiempo que transcurre hasta que se actúa contra dicha introgresión (es decir: el tiempo durante el que la información exógena permaneció mezclándose). Los casos con porcentajes de introgresión elevada (a partir del 30%) y largos períodos de mezcla (5 generaciones) son irrecuperables con la mayoría de los métodos utilizados en este trabajo. El ejemplo del caballo de Przewalski en el que una sola hembra de caballo doméstico se introdujo en la población hace cien años, muestra como una mínima cantidad de material exógeno puede extenderse a casi todos los individuos de la población.

Cuando el mismo porcentaje de introgresión ocurre en varias generaciones la posibilidad de recuperación es mayor (comparando escenarios con la misma proporción total de genoma exógeno introducido) consecuencia de un tiempo menor durante el cual la información exógena se mezcla.

El tamaño de la población problema es también un elemento clave para el éxito de la desintrogresión. La posibilidad de encontrar individuos puros o con un porcentaje de genes exógenos pequeño es escasa en poblaciones de censo reducido (Capítulos 1 y 2). Un problema añadido en estas poblaciones es el incremento de consanguinidad que se deriva del proceso, mucho más grave cuanto menor sea el número de individuos disponibles.

El tamaño del genoma de la especie también influye sobre el resultado de la desintrogresión cuando se realiza a partir de información molecular. Poblaciones de especies con genomas pequeños son más sencillas de recuperar mediante desintrogresión con marcadores moleculares, como se vio en el Capítulo 2, ya que el ligamiento entre los genes y los marcadores, y con ello la informatividad de estos últimos, es mayor.

La cantidad de información disponible con la que se intente revertir la mezcla, así como su fiabilidad es el último punto clave en la eficacia del proceso. El éxito del proceso dependerá de ambas, independientemente del método aplicado. Las genealogías mal anotadas o los datos moleculares con un gran número de errores de genotipado pueden implicar que el reconocimiento de los individuos más puros no se lleve a cabo correctamente, de modo que la efectividad del proceso sea muy baja o nula.

La elección de la metodología se realizará en función de la disponibilidad de información. Cuando haya una genealogía disponible, minimizar el parentesco de los individuos disponibles actualmente con los exógenos introducidos (MEC) será el método más eficaz. La estrategia logró recuperar gran parte del fondo genético nativo en una población de 10 individuos, en los escenarios con menos introgresión (10-30% exógenos, 1-3 generaciones de mezcla). Mientras que en la población de 100 individuos la recuperación fue alta en todos los escenarios, incluso tras cinco generaciones de mezcla.

Sin embargo, en este caso la existencia de un pedigrí completo es imprescindible para poder aplicar el método. Las genealogías suelen estar disponibles en especies ganaderas, en las que existe un control específico sobre los apareamientos por motivos productivos, o en poblaciones en cautividad (FRANKHAM et al. 2002; HALEY 2009). Si tan solo disponemos de un pedigrí parcial, éste puede completarse o reconstruirse mediante marcadores (BUTLER et al. 2004) y así, el método genealógico podrá también aplicarse.

Varios estudios han probado que la información molecular, cuando es lo suficientemente densa, puede remplazar a las genealogías en diversas tareas, como calcular la precisión de las estimas de parámetros genéticos, o mantener diversidad genética (HAYES et al. 2009; DE CARA et al. 2011). En este trabajo se ha

demostrado que lo mismo ocurre en el caso de desintrogresión, en el que la información de 50000 SNP usados para calcular el parentesco molecular logró mejores resultados que el método genealógico (MEC).

La información de genotipado masivo ha resultado ser la más útil consiguiéndose recuperar la mayor cantidad de fondo genético nativo a través de los diferentes métodos estudiados. Los paneles de SNP permiten disponer de marcadores a lo largo de todo el genoma, lo cual proporciona la información necesaria para detectar y seleccionar todos los fragmentos con origen nativo (debido al desequilibrio de ligamiento). Al contrario de lo que ocurre cuando se utilizan las genealogías, que suponen el valor esperado promedio del parentesco entre individuos, el valor que se obtiene mediante los marcadores moleculares es el parentesco realizado (BAUMUNG & SÖLKNER 2003; ENGELSMA et al. 2011) lo que le confiere una mayor potencia.

Una consecuencia práctica de esta característica es que el método MEC (basado en información genealógica) sólo es eficiente la primera generación en la que se implementa. Al escoger como padres los menos emparentados con los exógenos, los individuos de la población resultante tienen exactamente la misma relación genealógica con los exógenos, y por tanto no hay más capacidad para discriminar. Sin embargo, aunque el pedigrí muestre un parentesco igual para todos los individuos, habrá todavía diferencias entre ellos para los marcadores, que permitirán más generaciones de depuración cuando se utilice información molecular.

Respecto de las estrategias estudiadas en los Capítulos 3 y 4, si bien es cierto que los resultados obtenidos difieren sensiblemente (mostrando una mayor recuperación del genoma nativo con el manejo mediante haplotipos y GBLUP), dichos resultados no pueden ser comparados de manera directa debido a que las simulaciones no fueron realizadas con los mismos parámetros. Mientras que las poblaciones en el caso de minimización del parentesco molecular (Capítulo 3) fueron totalmente simuladas (dos poblaciones independientes con las mismas frecuencias de partida), en el caso del manejo mediante haplotipos y GBLUP se utilizaron los genotipos reales de dos razas de ovino (Merino y Poll Dorset) para simular los escenarios de introgresión. Una consecuencia de esto es que el nivel de

diferenciación entre las poblaciones nativa y exógena no es el mismo en los dos grupos de simulaciones. Esta diferenciación es precisamente la base para reconocer los individuos portadores de la mayor proporción de alelos nativos y en los tres métodos implicados es esperable una efectividad diferencial. En el caso del Capítulo 4, los haplotipos procedentes de los dos orígenes son perfectamente distinguibles en los individuos cruzados F1. Las poblaciones Merino y Poll Dorset son genéticamente muy diferentes lo que se demuestra en las simulaciones con 20 generaciones de mezcla. Incluso después de un período de mezcla tan largo, los métodos son capaces de diferenciar entre ambas poblaciones y recuperar más de un 80% de genoma nativo en casos dónde la introgresión fue del 50%. El escenario simulado en el Capítulo 3 para estudiar la eficiencia del método MECG es mucho más restrictivo. Esto no implica que no sea realista, sino que representa un caso de poblaciones más próximas y parecidas genéticamente.

Los métodos MECG y GBLUP, aunque se basan ambos en matrices de parentesco calculadas a partir de SNP, difieren en las correcciones que se aplican al calcular dichas matrices. El primer método calcula el parentesco molecular como identidad en estado de los alelos (CABALLERO & TORO 2002) y posteriormente corrige para las frecuencias iniciales simuladas, que son conocidas sin error (SHEPHERD et al. 2010). En el método de GBLUP la matriz de relaciones genómicas se calculó como se describe en YANG et al. (2010), donde el parentesco es corregido por las frecuencias actuales de los SNP. Para tener en cuenta que estas frecuencias no son conocidas, sino estimadas, se hace otra corrección por el tamaño de muestra.

La mayor limitación en cualquiera de estos métodos es la disponibilidad de la información. La utilidad de los SNP en razas ganaderas está aumentando considerablemente debido al creciente interés en la selección genómica (MEUWISSEN et al. 2001) y la disponibilidad de paneles con un gran número de SNP para estas poblaciones se incrementa cada día. Sin embargo, en poblaciones salvajes, la existencia de información masiva no es común, y aún está por desarrollar (SLATE et al. 2009). Por tanto, la aplicación de los métodos mencionados queda restringida a casos concretos, como por ejemplo los de razas

locales cruzadas con otras más productivas cuyas características adaptativas quisiéramos recuperar (MORAIS et al. 2005; TABERLET et al. 2008).

No obstante, siguen desarrollándose técnicas para descubrimiento de SNP utilizando nuevas aproximaciones, que resultan cada vez más económicas, e incluso aplicables en especies que no tengan disponible la secuencia del genoma. VAN TASSELL et al. (2008) describen un método en el que se descubren nuevos SNP a la vez que se validan y caracterizan en un solo paso, dónde el precio de descubrimiento de un SNP es aproximadamente 0.48$. De un modo similar, ELSHIRE et al. (2011) proponen cómo identificar un gran número de marcadores mediante técnicas de genotipado por secuenciación (GBS) en un solo experimento. Los chips de SNP, que resultan más baratos de desarrollar para consorcios que los empleen en diferentes poblaciones, podrán evitarse en poblaciones salvajes, gracias a las técnicas de secuenciación de nueva generación (*next generation sequencing*) que permiten obtener pequeños paneles de SNP a precios más competitivos (DAVEY et al. 2011).

Cuando no haya disponible información molecular de alta densidad, otro tipo de información molecular puede servir para llevar a cabo la desintrogresión de la población problema, ya sean microsatélites o un número reducido de SNP. Los marcadores microsatélites han sido ampliamente utilizados para múltiples fines tanto en conservación como para caracterización de razas en poblaciones ganaderas y salvajes (LUIKART & ENGLAND 1999; TORO et al. 2009).

Como se muestra en el Capítulo 2, la información molecular no densa también puede emplearse para la reversión de la introgresión. Los resultados logrados en este caso dependerán, además del porcentaje de introgresión y del tiempo de mezcla, del número de marcadores disponibles y de la frecuencia de los alelos de dichos marcadores en las poblaciones nativa y exógena. Es decir, de nuevo la diferenciación entre ambas poblaciones es el factor crucial para el éxito, puesto que cuanto más diferentes sean las frecuencias de los alelos en las dos poblaciones, mayor capacidad tendrán los métodos para reconocer fragmentos genómicos procedentes de los nativos/exógenos. Así, los marcadores diagnóstico, es decir, aquellos con alelos exclusivos de al menos una de las poblaciones, son los que obtienen mejores resultados a la hora de revertir introgresión. Por el contrario, los

alelos a frecuencias similares en ambas poblaciones, tienen poca potencia de recuperación, con resultados aceptables cuando la introgresión fue baja pero prácticamente ineficaces tras varias generaciones de mezcla.

En poblaciones próximas, como suelen ser las implicadas en procesos de introgresión, los marcadores diagnóstico no son frecuentes (VILÀ et al. 2003). Aun así, pueden encontrarse algunos ejemplos a nivel de especie y subespecie (ROY et al. 1994; MACHUGH et al. 1997) que, combinados con otros marcadores, podrían servir para llevar a cabo la desintrogresión de la población problema.

La desintrogresión mediante distancias genéticas (Capítulo 2) ha demostrado ser el menos eficaz de los métodos probados. En casos de poca mezcla, la recuperación es notable, sin embargo, tras dos o tres generaciones de mezcla los resultados fueron nulos.

Un último punto a tener en cuenta respecto a la utilidad de los métodos basados en información molecular, es la necesidad de conocer las frecuencias alélicas en las poblaciones originales, o la existencia de individuos puros (nativos o exógenos) genotipados como referencia para utilizar los métodos estudiados.

La minimización de la distancia genética con la población nativa original asume que las frecuencias alélicas en la población nativa pura son conocidas, sin embargo, éstas pueden estar mal estimadas. Al igual que los errores de genotipado, una mala estimación de las frecuencias alélicas implicará que la población obtenida de la desintrogresión no sea más parecida a la población nativa original, y por tanto un fracaso del proceso.

Si la población de referencia es demasiado pequeña para calcular las frecuencias originales y aplicar los métodos expuestos en el Capítulo 2, una solución alternativa podría ser tratar de inferir la genealogía utilizando los genotipos de los marcadores (BUTLER et al. 2004). Así, aunque sólo dispongamos de un número reducido de individuos exógenos (de los cuales no podrían estimarse correctamente las frecuencias alélicas), podría calcularse su relación con los individuos actuales, y aplicando el método MEC aumentar la proporción de genoma nativo en la población.

Por su parte, la estrategia de minimización de parentesco molecular con los exógenos implica tener cierto número de individuos exógenos puros genotipados. Aunque en el Capítulo 3 se trabajó con los mismos exógenos que entraron en la población, el método MECG podría funcionar también con una muestra aleatoria de individuos exógenos, puesto que se trata de una medida de parentesco o parecido. En cualquier caso dicha información ha de estar disponible. Este método tiene una estrategia equivalente y opuesta, si la información genómica disponible es de los individuos nativos puros, en cuyo caso podría maximizarse el parentesco molecular con los nativos (MACG).

Las estrategias de GBLUP y haplotipos fueron testadas con muestras aleatorias de nativos y exógenos. En el caso de GBLUP, el método podría llevarse a cabo si la muestra consta sólo de una de las dos poblaciones, aunque, evidentemente, cuanta más información disponible, más eficaz será el método. Sin embargo necesitamos individuos de los dos tipos si queremos implementar la estrategia basada en la detección de haplotipos.

En la Figura D.1 se muestra un diagrama de flujo que resume las posibilidades de actuación para llevar a cabo la desintrogresión de una población en función de la información disponible. Aunque la representación está muy simplificada y existan casos más complejos, el procedimiento a llevar a cabo depende de la información disponible (genealógica o molecular) y de la diferenciación entre poblaciones.

Una posibilidad que no se ha tratado en este estudio, es la de desintrogresión basada en fenotipos. En ocasiones, el interés por una determinada población se manifiesta en la expresión de un carácter concreto, lo que justificaría usar como criterio de depuración el fenotipo. La eficiencia en la depuración dependerá de la arquitectura genética del carácter de interés. Aquellos controlados por uno o pocos genes (cualitativos) serán fáciles de fijar, pero no proporcionarán información sobre el resto del genoma. Por el contrario, en caracteres de tipo infinitesimal (cuantitativos), se espera que puedan actuar a lo largo de todo el genoma. El problema en este último caso es determinar cuál es el valor para el carácter que queremos conservar.

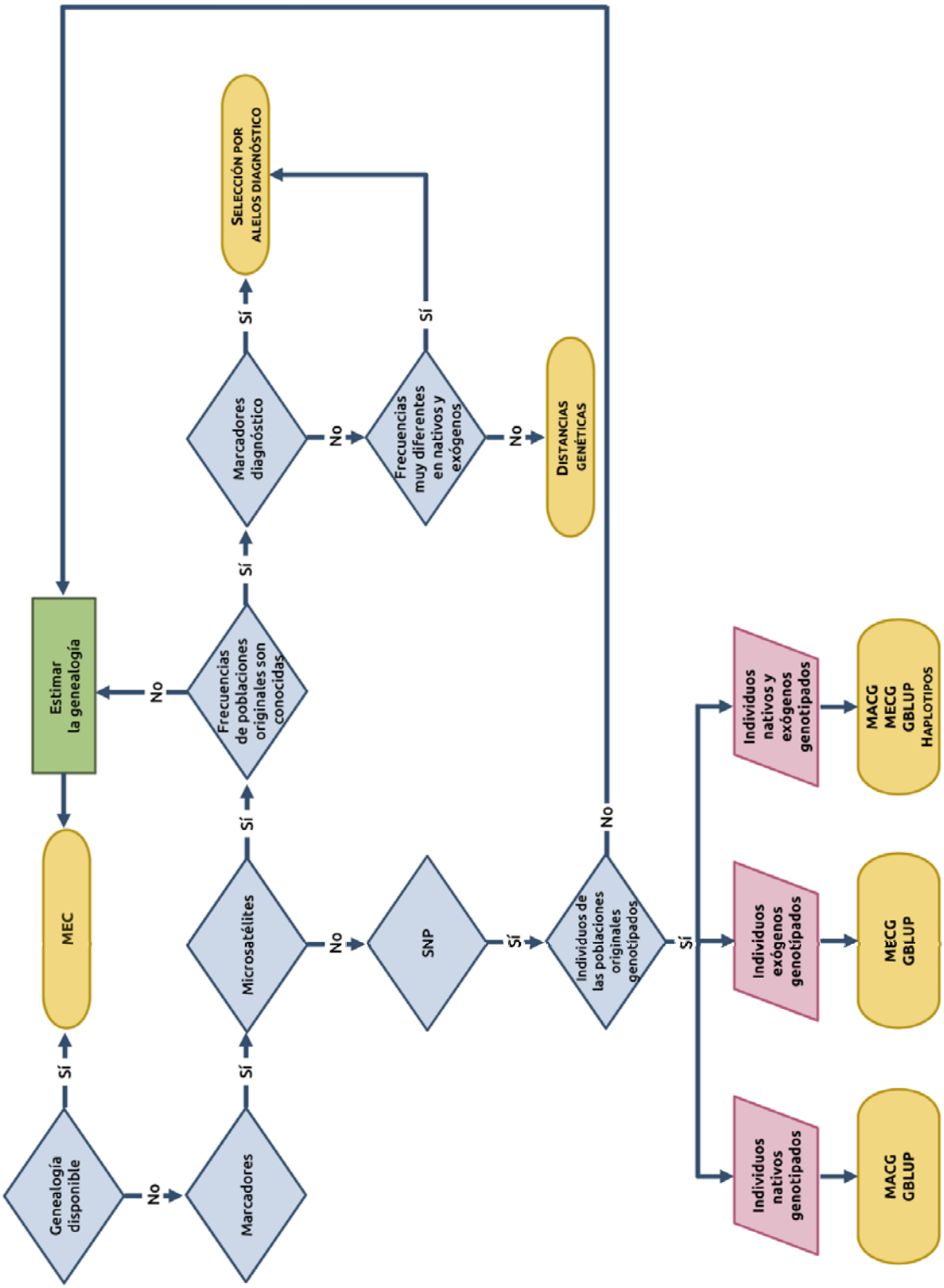

FIGURA D.1. Diagrama de flujo representando las opciones de actuación para revertir introgresión en función de la información disponible.

Cuando se trabaja con razas de animales domésticos en ocasiones hay que depender de índices morfológicos que miden el parecido con el patrón racial, como ocurre en FERNÁNDEZ et al. (2012). En este caso la información fenotípica de un carácter diagnóstico se utilizará como medida de identificación de aquellos individuos más puros.

El incremento en consanguinidad fruto del proceso de depuración es una consecuencia negativa de todos los métodos estudiados. Dicho incremento ha de tenerse en cuenta al diseñar el plan de manejo y monitorizarse a lo largo de todo el proceso. Poblaciones con consanguinidad elevada sufren, a raíz de la disminución en diversidad y aumento en la homocigosidad de deletéreos, una disminución en su eficacia biológica a todos los niveles: número de descendientes, supervivencia en los individuos juveniles, longevidad, intervalo entre partos, calidad y cantidad de esperma, aptitud maternal, tiempo de desarrollo, etc. (FRANKHAM et al. 2002).

El mantenimiento de la diversidad y la reducción de ΔF han de ser objetivos paralelos en el proceso de desintrogresión. Esto puede llevarse a cabo bien mediante un número mínimo de individuos que contribuyan cada generación, o bien imponiendo una restricción explícita en la tasa de consanguinidad cada generación. Adicionalmente puede optimizarse el esquema de apareamientos para evitar los apareamientos consanguíneos.

No obstante, hay que considerar que la consanguinidad es una consecuencia inevitable en la mayoría de los casos. Esto implica que en cada situación habrá que alcanzar una situación de compromiso. Por tanto, antes de empezar el proceso de desintrogresión habrá que evaluar qué valores de consanguinidad estamos dispuestos a aceptar en el caso correspondiente. Por ejemplo, en poblaciones grandes el impacto de la desintrogresión sobre la consanguinidad puede ser relativamente bajo, sin embargo, en el caso de poblaciones pequeñas y en peligro, la selección de un número reducido de individuos a fin de recuperar una pequeña proporción de genoma nativo puede ser potencialmente dañina e inaceptable.

Esto es lo que ocurre con el caballo de Przewalski (*Equus ferus przewalskii*). Tras casi extinguirse después de la Segunda Guerra Mundial, la población actual se recuperó notablemente, aunque su estado actual sigue siendo crítico (BOUMAN & BOUMAN 1994). En 2005 existían 1860 ejemplares, 300 de ellos viviendo en libertad

en Mongolia y en varias reservas en China (http://www.zoopraha.cz/en/about-zoo/history/przewalski-s-horses). La información genética exógena de un único caballo se ha estado extendiendo por la población durante 100 años, por lo que prácticamente todos los individuos tienen introgresión. La selección de un número reducido de individuos para recuperar el genoma Przewalski original causaría un incremento de la consanguinidad, en una población ya de por sí altamente emparentada, que perjudicaría demasiado a la especie convirtiéndose en una opción inviable.

Una vez finalizado el proceso de desintrogresión, la monitorización de la diversidad genética de la población pasará a ser el objetivo prioritario. Puesto que el incremento de la consanguinidad es una consecuencia directa de la desintrogresión, un control posterior será de fundamental importancia para poder mantener la máxima diversidad posible. Para ello sería recomendable continuar manejando la población con un método que minimice la tasa de consanguinidad, como el de contribuciones óptimas (CABALLERO & TORO 2000; SONESSON & MEUWISSEN 2001).

Hemos visto que la recuperación del genoma puro de una población tras su mezcla con individuos exógenos es posible y que la eficacia del proceso depende de las poblaciones involucradas así como de la información disponible. No obstante, la desintrogresión no es un proceso sin coste, sino que implica una pérdida de diversidad genética y un incremento en consanguinidad. Además, la recuperación del genoma completo no está siempre asegurada. Por tanto, los resultados de este trabajo también sugieren la importancia de la prevención de la mezcla de poblaciones que interese mantener en pureza.

En el caso de conservación de especies ganaderas, antes de llevar a cabo un cruce con una raza más productiva habrá que evaluar cuidadosamente los intereses, puesto que podría ser un proceso irreversible, y la pérdida de las adaptaciones locales puede producirse rápidamente (TABERLET et al. 2008). En el caso de poblaciones salvajes, habría que eliminar en lo posible las especies invasivas, así como controlar el impacto de las actividades humanas sobre los paisajes, para minimizar los daños de este tipo. En cualquier caso, la

monitorización de las poblaciones mediante registros genealógicos y/o datos moleculares sería recomendable.

CONCLUSIONES

1. Un fenómeno de introgresión no deseada puede ocurrir en especies salvajes y domésticas, amenazando la pureza de la población y requiriendo un proceso de depuración para recuperar dicho genoma.

2. La cantidad de genoma exógeno introducido y el tiempo durante el cual se produjo la mezcla son factores limitantes para el éxito de la depuración: hay que actuar tan pronto como sea posible.

3. La información genealógica o molecular puede utilizarse para llevar a cabo la depuración. La información de genotipado masivo es la que permite una mejor depuración de las poblaciones.

4. La diferenciación entre las poblaciones nativa y exógena es crucial para poder recuperar el genoma original con herramientas moleculares. Cuanto más parecidas sean las poblaciones, más difícil es reconocer los individuos puros y por tanto el éxito disminuirá.

5. El incremento de la consanguinidad es una consecuencia inevitable del proceso. Cuanta más cantidad de información exógena se elimina, más altos son los niveles de consanguinidad. La desintrogresión ha de llevarse a cabo incluyendo un mínimo de individuos que contribuyan a la siguiente generación o una restricción explícita en la tasa de consanguinidad.

CONCLUSIONS

1. Undesired introgression can happen in wild and domestic species, risking the purity of a population, and requires a process of recovery of the original genome.
2. The amount of exogenous alleles introgressed and the time elapsed until management starts are crucial to the success of the de-introgression process: acting as soon as possible is essential.
3. Genealogical and molecular information can be used to restore the native background. Dense molecular markers allow for better results.
4. Genetic differences between native and exogenous populations are crucial to recover the native genome using molecular tools. When the populations are too similar, detecting the purest individuals is difficult, and the success probability decreases.
5. A by-product of the recovery process is an increase of inbreeding. The more exogenous information is removed the larger this increase is. De-introgression must be carried out by imposing a minimal number of individuals contributing to the next generations, or an explicit restriction in the increase of inbreeding.

BIBLIOGRAFÍA

ALLENDORF, F.W., LEARY, R.F., SPRUELL, P. & WENBURG, J.K., 2001. The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution* **16:** 613-622.

ALLENDORF, F.W. & LUIKART, G., 2007. *Conservation and the Genetics of Populations*. Blackwell, Malden (MA).

BARBANERA, F., FORCINA, G., GUERRINI, M. & DINI, F., 2011. Molecular phylogeny and diversity of the Corsican red-legged partridge: hybridization and management issues. *Journal of Zoology* **285:** 56-65.

BARILANI, M., DEREGNAUCOURT, S., GALLEGO, S., GALLI, L., MUCCI, N., PIOMBO, R., PUIGCERVER, M., RIMONDI, S., RODRÍGUEZ-TEIJEIRO, J.D., SPANÒ, S. et al., 2005. Detecting hybridization in wild (*Coturnix c. coturnix*) and domesticated (*Coturnix c. japonica*) quail populations. *Biological Conservation* **126:** 445-455.

BARILLET, F., 2007. Genetic improvement for dairy production in sheep and goats. *Small Ruminant Research* **70:** 60-75.

BAUMUNG, R. & SÖLKNER, J., 2003. Pedigree and marker information requirements to monitor genetic variability. *Genetics Selection Evolution* **35:** 369-383.

BAYER, C.S.O., SACKMAN, A.M., BEZOLD, K., CABE, P.R. & MARSH, D.M., 2012. Conservation genetics of an endemic mountaintop salamander with an extremely limited range. *Conservation Genetics* **13:** 443-454.

BEAUMONT, M., BARRATT, E.M., GOTTELLI, D., KITCHENER, A.C., DANIELS, M.J., PRITCHARD, J.K. & BRUFORD, M.W., 2001. Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* **10:** 319-336.

BENNEWITZ, J., SIMIANER, H. & MEUWISSEN, T.H.E., 2008. Investigations on merging breeds in genetic conservation schemes. *Journal of Dairy Science* **91:** 2512-2519.

BENSON, J.F., HOSTETLER, J.A., ONORATO, D.P., JOHNSON, W.E., ROELKE, M.E., O'BRIEN, S.J., JANSEN, D. & OLI, M.K., 2011. Intentional genetic introgression influences survival of adults and subadults in a small, inbred felid population. *Journal of Animal Ecology* **80:** 958-967.

BLOTT, S.C., WILLIAMS, J.L. & HALEY, C.S., 1999. Discriminating among cattle breeds using genetic markers. *Heredity* **82:** 613-619.

BOLORMAA, S., HAYES, B.J., HAWKEN, R.J., ZHANG, Y., REVERTER, A. & GODDARD, M.E., 2011. Detection of chromosome segments of zebu and taurine origin and their effect on beef production and growth. *Journal of Animal Science* **89:** 2050-2060.

BOUMAN, I. & BOUMAN, J., 1994. The history of the Przewalski's horse, pp. 5-38 en *Przewalski's horse: The history and biology of an endangered species*, editado por L. BOYD & K. A. HOUPT. State University of New York Press, New York.

BOYER, M.C., MUHLFELD, C.C. & ALLENDORF, F.W., 2008. Rainbow trout (*Oncorhynchus mykiss*) invasion and the spread of hybridization with native westslope cutthroat trout (*Oncorhynchus clarkii lewisi*). *Canadian Journal of Fisheries and Aquatic Sciences* **65:** 658-669.

BROWNING, B.L. & BROWNING, S.R., 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics* **84:** 210-223.

BUTLER, K., FIELD, C., HERBINGER, C.M. & SMITH, B.R., 2004. Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data. *Molecular Ecology* **13:** 1589-1600.

CABALLERO, A. & TORO, M.A., 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research* **75:** 331-343.

CABALLERO, A. & TORO, M.A., 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics* **3:** 289-299.

CABRIA, M.T., MICHAUX, J.R., GÓMEZ-MOLINER, B.J., SKUMATOV, D., MARAN, T., FOURNIER, P., LÓPEZ DE LUZURIAGA, J. & ZARDOYA, R., 2011. Bayesian analysis of hybridization and introgression between the endangered european mink (*Mustela lutreola*) and the polecat (*Mustela putorius*). *Molecular Ecology* **20:** 1176-1190.

CAVALLI-SFORZA, L.L. & EDWARDS, A.W.F., 1967. Phylogenetic Analysis Models and Estimation Procedures. *American Journal of Human Genetics* **19:** 233-257.

CRISPO, E., MOORE, J.S., LEE-YAW, J.A., GRAY, S.M. & HALLER, B.C., 2011. Broken barriers: Human-induced changes to gene flow and introgression in animals. *Bioessays* **33:** 508-518.

DAETWYLER, H.D., WIGGANS, G.R., HAYES, B.J., WOOLLIAMS, J.A. & GODDARD, M.E., 2011. Imputation of Missing Genotypes From Sparse to High Density Using Long-Range Phasing. *Genetics* **189:** 317-U1028.

DALVIT, C., DE MARCHI, M. & CASSANDRO, M., 2007. Genetic traceability of livestock products: A review. *Meat Science* **77:** 437-449.

DANIELS, M.J. & CORBETT, L., 2003. Redefining introgressed protected mammals: when is a wildcat a wild cat and a dingo a wild dog? *Wildlife Research* **30:** 213-218.

DANTZIG, G.B., 1963. *Linear programming and extensions*. Princeton University Press, Princeton.

DAVEY, J.W., HOHENLOHE, P.A., ETTER, P.D., BOONE, J.Q., CATCHEN, J.M. & BLAXTER, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12:** 499-510.

DE CARA, M.A.R., FERNÁNDEZ, J., TORO, M.A. & VILLANUEVA, B., 2011. Using genome-wide information to minimise the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics* **128:** 456-464.

DRISCOLL, C., YAMAGUCHI, N., O'BRIEN, S.J. & MACDONALD, D.W., 2011. A suite of genetic markers useful in assessing wildcat (*Felis silvestris* ssp.)—domestic cat (*Felis silvestris catus*) admixture. *Journal of Heredity* **102:** S87-S90.

EDMANDS, S., 2007. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology* **16:** 463-475.

ELSHIRE, R.J., GLAUBITZ, J.C., SUN, Q., POLAND, J.A., KAWAMOTO, K., BUCKLER, E.S. & MITCHELL, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6:** e19379.

ENGELSMA, K.A., VEERKAMP, R.F., CALUS, M.P.L. & WINDIG, J.J., 2011. Consequences for diversity when prioritizing animals for conservation with pedigree or genomic information. *Journal of Animal Breeding and Genetics* **128:** 473-481.

FAO, 1998. *Secondary guidelines for development of national farm animal genetic resources management plans: Measurement of Domestic Animal Diversity (MoDAD)*. FAO, Rome.

FERNÁNDEZ, J. & TORO, M.A., 1999. The use of mathematical programming to control inbreeding in selection schemes. *Journal of Animal Breeding and Genetics* **116:** 447-466.

FERNÁNDEZ, J., CLEMENTE, I., AMADOR, C., MEMBRILLO, A., AZOR, P. & MOLINA, A., 2012. Use of different sources of information for the recovery and genetic management of endangered populations: example with the extreme case of Iberian pig Dorado strain. *Livestock Science*. Online. DOI: 10.1016/j.livsci.2012.07.019.

FRANKHAM, R., BALLOU, J.D. & BRISCOE, D.A., 2002. *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.

FREESE, C.H., AUNE, K.E., BOYD, D.P., DERR, J.N., FORREST, S.C., GATES, C.C., GOYAN, P.J.P., GRASSEL, S.M., HALBERT, N.D., KUNKEL, K. et al., 2007. Second chance for the plains bison. *Biological Conservation* **136:** 175-184.

GANDINI, G. & OLDENBROEK, K., 2007. Strategies for moving from conservation to utilisation, pp. 29-54 en *Utilisation and conservation of farm animal genetic resources*, editado por K. OLDENBROEK. Wageningen Academic Publishers, Wageningen, the Nederlands.

GATTEPAILLE, L.M. & JAKOBSSON, M., 2012. Combining Markers into Haplotypes Can Improve Population Structure Inference. *Genetics* **190:** 159-174.

GILMOUR, A.R., GOGEL, B.J., GULLIS, S.J. & WELHAM, S.J., 2002. *ASReml User Guide Release 1.0*. VSN International Ltd., Hemel Hepstead, UK.

GOODMAN, S.J., BARTON, N.H., SWANSON, G., ABERNETHY, K. & PEMBERTON, J.M., 1999. Introgression through rare hybridization: A genetic study of a hybrid zone between red and sika deer (genus *Cervus*) in Argyll, Scotland. *Genetics* **152:** 355-371.

GOSEY, J., 1991. Crossbreeding systems and the theory behind composite breeds, pp. 33 en *Proceedings, The Range Beef Cow Symposium XII*, Fort Collins, Colorado.

GOTTELLI, D., SILLEROZUBIRI, C., APPLEBAUM, G.D., ROY, M.S., GIRMAN, D.J., GARCIAMORENO, J., OSTRANDER, E.A. & WAYNE, R.K., 1994. Molecular genetics of the most endangered canid: the Ethiopian wolf *Canis simensis*. *Molecular Ecology* **3:** 301-312.

GROENEVELD, L.F., LENSTRA, J.A., EDING, H., TORO, M.A., SCHERF, B., PILLING, D., NEGRINI, R., FINLAY, E.K., JIANLIN, H., GROENEVELD, E. et al., 2010. Genetic diversity in farm animals – a review. *Animal Genetics* **41:** 6-31.

HALBERT, N.D. & DERR, J.N., 2007. A comprehensive evaluation of cattle introgression into US federal bison herds. *Journal of Heredity* **98:** 1-12.

HALEY, C., 2009. Human and livestock genetics: parallel evolution and horizontal exchange. *Journal of Animal Breeding and Genetics* **126:** 413-414.

HALL, S.J.G. & BRADLEY, D.G., 1995. Conserving livestock breed biodiversity. *Trends in Ecology & Evolution* **10:** 267-270.

HAYES, B.J., VISSCHER, P.M. & GODDARD, M.E., 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research* **91:** 47-60.

HEDRICK, P.W., 1995. Gene flow and genetic restoration: The Florida Panther as a case study. *Conservation Biology* **9:** 996-1007.

HINDAR, K., FLEMING, I.A., MCGINNITY, P. & DISERUD, A., 2006. Genetic and ecological effects of salmon farming on wild salmon: modelling from experimental results. *Ices Journal of Marine Science* **63:** 1234-1247.

HOHENLOHE, P.A., AMISH, S.J., CATCHEN, J.M., ALLENDORF, F.W. & LUIKART, G., 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources* **11:** 117-122.

INGVARSSON, P.K., 2001. Restoration of genetic variation lost - The genetic rescue hypothesis. *Trends in Ecology & Evolution* **16:** 62-63.

KIDD, A.G., BOWMAN, J., LESBARRERES, D. & SCHULTE-HOSTEDDE, A.I., 2009. Hybridization between escaped domestic and wild American mink (*Neovison vison*). *Molecular Ecology* **18:** 1175-1186.

KIRKPATRICK, S., GELATT, C.D. & VECCHI, M.P., 1983. Optimization by Simulated Annealing. *Science* **220:** 671-680.

KULLBACK, S., 1997. *Information theory and statistics*. Courier Dover Publications, Dover, New York.

LACY, R.C., ALAKS, G. & WALSH, A., 1996. Hierarchical analysis of inbreeding depression in *Peromyscus polionotus*. *Evolution* **50:** 2187-2200.

LACY, R.C., 1997. Errata. *Evolution* **51:** 1025-1025.

LARGIADÈR, C.R., 2006. Hybridization and introgression between native and alien species en *Biological invasions*, editado por W. NENTWIG. Springer, Berlin.

LENSTRA, J.A., GROENEVELD, L.F., EDING, H., KANTANEN, J., WILLIAMS, J.L., TABERLET, P., NICOLAZZI, E.L., SÖLKNER, J., SIMIANER, H., CIANI, E. et al., 2012. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Animal Genetics* **43:** 483–502.

LUIKART, G. & ENGLAND, P.R., 1999. Statistical analysis of microsatellite DNA data. *Trends in Ecology & Evolution* **14:** 253-256.

MACHUGH, D.E., SHRIVER, M.D., LOFTUS, R.T., CUNNINGHAM, P. & BRADLEY, D.G., 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and Zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146:** 1071-1086.

MADSEN, T., SHINE, R., OLSSON, M. & WITTZELL, H., 1999. Restoration of an inbred adder population. *Nature* **402:** 34-35.

MAPA 2003. Estudio y caracterización del sector equino en España. (Ministerio de Agricultura, Pesca y Alimentación)

MAUDET, C., LUIKART, G. & TABERLET, P., 2002. Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. *Journal of Animal Science* **80:** 942-950.

MEUWISSEN, T.H.E., HAYES, B.J. & GODDARD, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157:** 1819-1829.

MILLER, C.R., ADAMS, J.R. & WAITS, L.P., 2003. Pedigree-based assignment tests for reversing coyote (*Canis latrans*) introgression into the wild red wolf (*Canis rufus*) population. *Molecular Ecology* **12:** 3287-3301.

MORAIS, J., OOM, M.M., MALTA-VACAS, J. & LUÍS, C., 2005. Genetic structure of an endangered Portuguese semiferal pony breed, the Garrano. *Biochemical Genetics* **43:** 347-364.

MORIN, P.A., LUIKART, G., WAYNE, R.K. & GRP, S.N.P.W., 2004. SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19:** 208-216.

NEGRO, J.J., TORRES, M.J. & GODOY, J.A., 2001. RAPD analysis for detection and eradication of hybrid partridges (*Alectoris rufa* × *A.graeca*) in Spain. *Biological Conservation* **98:** 19-24.

NEI, M., 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* **70:** 3321-3323.

NEI, M., 1987. *Molecular evolutionary genetics*. Columbia. University Press, New York.

NORÉN, K., DALÉN, L., KVALØY, K. & ANGERBJÖRN, A., 2005. Detection of farm fox and hybrid genotypes among wild arctic foxes in Scandinavia. *Conservation Genetics* **6:** 885-894.

OLIVEIRA, R., GODINHO, R., RANDI, E., FERRAND, N. & ALVES, P.C., 2008. Molecular analysis of hybridisation between wild and domestic cats (*Felis silvestris*) in Portugal: implications for conservation. *Conservation Genetics* **9:** 1-11.

PADILLA, J.A., SANSINFORIANO, E., PAREJO, J.C., RABASCO, A. & MARTÍNEZ-TRANCÓN, M., 2009. Inference of admixture in the endangered Blanca Cacereña bovine breed by microsatellite analyses. *Livestock Science* **122:** 314-322

PARKER, H.G., KIM, L.V., SUTTER, N.B., CARLSON, S., LORENTZEN, T.D., MALEK, T.B., JOHNSON, G.S., DEFRANCE, H.B., OSTRANDER, E.A. & KRUGLYAK, L., 2004. Genetic structure of the purebred domestic dog. *Science* **304:** 1160-1164.

PERRY, W.L., FEDER, J.L., DWYER, G. & LODGE, D.M., 2001. Hybrid zone dynamics and species replacement between *Orconectes* crayfishes in a northern Wisconsin lake. *Evolution* **55:** 1153-1166.

R DEVELOPMENT CORE TEAM R Foundation for Statistical Computing 2010. R: A Language and Environment for Statistical Computing. Vienna, Austria

RANDI, E., 2008. Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology* **17:** 285-293.

REID, N., 2011. European hare (*Lepus europaeus*) invasion ecology: implication for the conservation of the endemic Irish hare (*Lepus timidus hibernicus*). *Biological Invasions* **13:** 559-569.

RHYMER, J.M., WILLIAMS, M.J. & BRAUN, M.J., 1994. Mitochondrial analysis of gene flow between New Zealand mallards (*Anas platyrhynchos*) and grey ducks (*A. superciliosa*). *Auk* **111:** 970-978.

RHYMER, J.M. & SIMBERLOFF, D., 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* **27:** 83-109.

ROY, M.S., GEFFEN, E., SMITH, D., OSTRANDER, E.A. & WAYNE, R.K., 1994. Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution* **11:** 553-570.

SCHAEFFER, L.R., BURNSIDE, E.B., GLOVER, P. & FATEHI, J., 2011. Crossbreeding results in Canadian dairy cattle for production, reproduction and conformation. *The Open Agriculture Journal* **5:** 63-72.

SCHERF, B.D. (Ed.), 2000. *WORLD WATCH LIST for domestic animal diversity*. FAO, Rome.

SHEPHERD, R.K., MEUWISSEN, T.H.E. & WOOLLIAMS, J.A., 2010. Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. *BMC Bioinformatics* **11:** 529.

SIMIANER, H., 2005. Decision making in livestock conservation. *Ecological Economics* **53:** 559-572.

SLATE, J., GRATTEN, J., BERALDI, D., STAPLEY, J., HALE, M. & PEMBERTON, J.M., 2009. Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136:** 97-107.

SONESSON, A.K. & MEUWISSEN, T.H.E., 2001. Minimization of rate of inbreeding for small populations with overlapping generations. *Genetical Research* **77:** 285-292.

SUŠNIK, S., BERREBI, P., DOVC, P., HANSEN, M.M. & SNOJ, A., 2004. Genetic introgression between wild and stocked salmonids and the prospects for using molecular markers in population rehabilitation: the case of the Adriatic grayling (*Thymallus thymallus* L. 1785). *Heredity* **93:** 273-282.

SUTHERLAND, W.J., ARMSTRONG-BROWN, S., ARMSWORTH, P.R., BRERETON, T., BRICKLAND, J., CAMPBELL, C.D., CHAMBERLAIN, D.E., COOKE, A.I., DULVY, N.K., DUSIC, N.R. et al., 2006. The identification of 100 ecological questions of high policy relevance in the UK. *Journal of Applied Ecology* **43:** 617-627.

TABERLET, P., VALENTINI, A., REZAEI, H.R., NADERI, S., POMPANON, F., NEGRINI, R. & AJMONE-MARSAN, P., 2008. Are cattle, sheep, and goats endangered species? *Molecular Ecology* **17:** 275-284.

TALLMON, D.A., LUIKART, G. & WAPLES, R.S., 2004. The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution* **19:** 489-496.

TORO, M.A., FERNÁNDEZ, J. & CABALLERO, A., 2009. Molecular characterization of breeds and its use in conservation. *Livestock Science* **120:** 174-195.

UGARTE, E., RUIZ, R., GABIÑA, D. & BELTRÁN DE HEREDIA, I., 2001. Impact of high-yielding foreign breeds on the Spanish dairy sheep industry. *Livestock Production Science* **71:** 3-10.

VAN TASSELL, C.P., SMITH, T.P.L., MATUKUMALLI, L.K., TAYLOR, J.F., SCHNABEL, R.D., LAWLEY, C.T., HAUDENSCHILD, C.D., MOORE, S.S., WARREN, W.C. & SONSTEGARD, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* **5:** 247-252.

VANRADEN, P.M., OLSON, K.M., WIGGANS, G.R., COLE, J.B. & TOOKER, M.E., 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science* **94:** 5673-5682.

VERNESI, C., CRESTANELLO, B., PECCHIOLI, E., TARTARI, D., CARAMELLI, D., HAUFFE, H. & BERTORELLE, G., 2003. The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): A microsatellite analysis. *Molecular Ecology* **12:** 585-595.

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. *Genetics Selection Evolution* **34:** 275-305.

VILÀ, C., WALKER, C., SUNDQVIST, A.K., FLAGSTAD, O., ANDERSONE, Z., CASULLI, A., KOJOLA, I., VALDMANN, H., HALVERSON, J. & ELLEGREN, H., 2003. Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf–dog hybrids. *Heredity* **90:** 17-24.

VITOUSEK, P.M., MOONEY, H.A., LUBCHENCO, J. & MELILLO, J.M., 1997. Human domination of Earth's ecosystems. *Science* **277:** 494-499.

VOGEL, L.S. & JOHNSON, S.G., 2008. Estimation of hybridization and introgression frequency in toads (Genus : *Bufo*) using DNA sequence variation at mitochondrial and nuclear loci. *Journal of Herpetology* **42:** 61-75.

VOLF, J., 1994. The Studbook, pp. 61-73 en *Przewalski's horse: The history and biology of an endangered species*, editado por L. BOYD & K. A. HOUPT. State University of New York Press, New York.

VORBURGER, C. & REYER, H.U., 2003. A genetic mechanism of species replacement in European waterfrogs? *Conservation Genetics* **4:** 141-155.

WINDIG, J.J. & ENGELSMA, K.A., 2010. Perspectives of genomics for genetic conservation of livestock. *Conservation Genetics* **11:** 635-641.

YANG, J., BENYAMIN, B., MCEVOY, B.P., GORDON, S., HENDERS, A.K., NYHOLT, D.R., MADDEN, P.A., HEATH, A.C., MARTIN, N.G., MONTGOMERY, G.W. et al., 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42:** 565-U131.