UNIVERSITAT POLITÈCNICA DE VALÈNCIA

MANAGEMENT OF GENETIC DIVERSITY IN CONSERVATION PROGRAMS USING GENOMIC COANCESTRY

This thesis has been submitted in fulfilment of the requirements for the degree of Doctor at the Universitat Politècnica de València.

By

Elisabet Morales González

Supervisors:

Beatriz Villanueva Gaviña

Jesús Fernández Martín

Valencia, June 2023

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

GESTIÓN DE LA DIVERSIDAD GENÉTICA EN PROGRAMAS DE CONSERVACIÓN UTILIZANDO PARENTESCO GENÓMICO

Esta tesis ha sido escrita y presentada como uno de los requisitos para optar al grado de Doctor por la Universitat Politècnica de València.

Por

Elisabet Morales González

Directores:

Beatriz Villanueva Gaviña

Jesús Fernández Martín

Valencia, junio 2023

Esta Tesis Doctoral se realizó en el Instituto de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC) de Madrid. El trabajo expuesto en el capítulo 2 se realizó en parte en el Instituto Roslin, de la Universidad de Edimburgo durante una estancia predoctoral. Los trabajos expuestos en la tesis han sido financiados con una beca predoctoral FPI (BES-2017-081070) y a través de proyectos del Plan Estatal de I+D+i del Ministerio de Ciencia e Innovación (proyectos CGL2016-75904-C2-2-P y PID2020-114426GB-C22) y de la Unión Europea ('European Union's Seventh Framework Program, KBBE.2013.1.2-10, under Grant Agreement No. 613611, FISHBOOST project' y 'European Commission Horizon 2020, Framework Programme through Grant Agreement No. 727315, MedAID project').

"*A journey of a thousand miles begins with one step*"

Lao-Tse

ACKNOWLEDGMENTS

This thesis has been possible thanks to many people who have lent me their knowledge, experience, help and support.

First, I would like to thank my thesis supervisors, Beatriz Villanueva and Jesús Fernández, for trusting me and giving me the opportunity to carry out this thesis with them. I am not able to express in these few lines everything you have done for me these years and how grateful I am to you. Thank you for your constant attention, your patience, motivation and for everything you have taught me during the development of this thesis. Your guidance helped me during the research and writing of this doctoral thesis. Thank you both for giving me so many opportunities to go to conferences and creating a great work environment.

Beatriz, thank you for teaching me to be more orderly and methodical. Also because of your perfectionism, because thanks to you, in spite of my absent-mindedness, today I can distinguish a good work and my first thought from a not so good one is "Beatriz would not have let me send/present it like that ".

Jesús, thank you for being my guide in learning programming and for being there in most of the congresses that I have attended. When I saw you nodding while I presented the work, I felt a little less nervous.

I would like to thank Agustín Blasco for his fundamental role as a tutor and for being my link with the Universidad Politécnica de Valencia.

I would also like to thank Miguel Ángel Toro, for all the good ideas that he has contributed to this thesis and for being there for all the questions we have asked him. Thank you for your willingness to always help.

To Ricardo Pong-Wong, thank you for allowing me to do a research stay at the Roslin Institute, even though it did not go as expected due to COVID. I also want to thank him for his constructive feedback on much of this research and for the software he initially passed on to us.

I would also like to thank all the members of the Animal Breeding Department

of INIA-CSIC who gave me the opportunity to join their team and use their research facilities. In particular, I would like to thank María Saura for her patience in my early days in the world of science, for teaching me so much during my final master's thesis, for help me with the first article of this thesis, and for still being there up to this moment with her support and advice. And to Almudena Fernández thank you for her great help in my beginnings in the world of bioinformatics and her guidance in the first analyses that were part of my final master's thesis and this thesis.

I would like to thank my colleagues from the cafe Alejandro, Bea, Marta, Consuelo, Diego, Adrián López García, Mónica, María Muñoz, Óscar, Adrián López Catalina, Edgar, Ramón, Anita, Carlos, Raúl. Some of you started the journey with me, some of you have been at the end and others have been a constant all this time. Thank you all for your friendship, for the hours shared, for the coffee moments (so necessary), laughter and advice and for everything we experienced during the congresses. Special thanks to María Muñoz for the car trips that have given rise to so many conversations during this last year, thanks for your advice. It has been a great luck for me to have you as a neighbor.

I especially want to thank Silvia García-Ballesteros, who has been a constant support over the years. From the first day you showed up in the office across the corridor, I knew we would get along well. Thank you for being such a good friend and partner. For solving my doubts, no matter how silly they were and for listening to me, for encouraging me in my low moments and for your affection. This stage would not have been the same without you.

A special thanks to my parents, for being there in all the important moments of my life, for the unconditional support and love that you always give me, thank you for reminding me of what is really important in life. Because without you I would not be the person I am.

And finally, to Jordi, my couple, thank you for being by my side during these years, for taking care of me and encouraging me to keep going. For your support, affection and understanding. For always being willing to help with anything I need, even

when I don't ask for your help. In short, thank you very much for being part of my life.

AGRADECIMIENTOS

Esta tesis ha sido posible gracias a mucha gente que me ha prestado su conocimiento, su experiencia, su ayuda y su apoyo.

En primer lugar, me gustaría agradecer a mis directores de tesis, Beatriz Villanueva y Jesús Fernández, por confiar en mí y haberme dado la oportunidad de realizar esta tesis con ellos. No soy capaz de expresar en estas pocas líneas todo lo que habéis hecho por mí estos años y lo agradecida que estoy con vosotros. Gracias por vuestra constante atención, vuestra paciencia, orientación, motivación y por todo lo que me habéis enseñado durante el desarrollo de esta tesis. Vuestra guía me ayudó durante la investigación y redacción de esta tesis de doctorado. Gracias a los dos por darme tantas oportunidades para ir a congresos y crear un buenísimo ambiente de trabajo.

Beatriz, gracias por enseñarme a ser más ordenada y metódica. También por tu perfeccionismo, porque gracias a ti, a pesar de mi despiste, hoy puedo distinguir un buen trabajo y mi primer pensamiento de uno no tan bueno es “Beatriz no me hubiera dejado mandar/presentar eso así”.

Jesús, gracias por ser mi guía en el aprendizaje de la programación y por estar ahí en la mayoría de los congresos a los que he asistido. Cuando te veía asentir mientras presentaba el trabajo me sentía un poquito menos nerviosa.

Quiero agradecer a Agustín Blasco, por su fundamental papel como tutor y por ser mi nexo con la Universidad Politécnica de Valencia.

También quiero agradecer a Miguel Ángel Toro, por todas las buenas ideas que ha aportado a esta tesis y por estar ahí para todas las consultas que le hemos hecho. Gracias por tu buena disposición siempre a ayudar.

A Ricardo Pong-Wong, gracias por permitirme hacer una estancia de investigación en el Instituto Roslin, aunque no saliera como esperábamos debido al COVID. También quiero darle las gracias por sus comentarios constructivos sobre gran parte de esta investigación y por el software que en un inicio nos pasó.

También me gustaría agradecer a todos los miembros del Departamento de

Mejora Genética Animal de INIA-CSIC que me dieron la oportunidad de unirme a su equipo y utilizar sus instalaciones de investigación. En particular, agradezco a María Saura por su paciencia en mis inicios en el mundo de la ciencia, por enseñarme tanto durante mi trabajo final de máster, con el primer artículo de esta tesis y por seguir ahí hasta este momento con su apoyo y sus consejos. Y a Almudena Fernández por su gran ayuda en mis inicios en el mundo de la bioinformática y su guía en los primeros análisis que formaron parte de mi trabajo final de máster y de esta tesis.

Agradezco a mis compañeros del café Alejandro, Bea, Marta, Consuelo, Diego, Adrián López García, Mónica, María Muñoz, Óscar, Adrián López Catalina, Edgar, Ramón, Anita, Carlos, Raúl. Algunos comenzasteis el camino conmigo, algunos habéis estado al final y otros habéis sido una constante todo este tiempo. Gracias a todos por vuestra amistad, por las horas compartidas, por los momentos de café (tan necesarios), risas y consejos y por todo lo vivido durante los congresos. Especialmente agradecer a María Muñoz por los viajes en coche que han dado pie a tantas conversaciones en este último año, gracias por tus consejos. Ha sido una gran suerte para mi tenerte de vecina.

Muy especialmente quiero agradecer a Silvia García-Ballesteros que ha sido un apoyo constante durante estos años. Desde el primer día que apareciste en el despacho de enfrente supe que nos llevaríamos bien. Gracias por ser tan buena amiga y compañera. Por solventar mis dudas por tontas que fueran y por escucharme, por darme ánimos en mis momentos bajos y por tu cariño. Esta etapa no habría sido igual sin ti.

Un especial agradecimiento a mis padres, por estar ahí en todos los momentos importantes de mi vida, por el apoyo incondicional y el amor que siempre me brindáis, gracias por recordarme lo que es realmente importante en la vida. Porque sin vosotros no sería la persona que soy.

Y por último, a Jordi, mi pareja, gracias por estar a mi lado durante estos años, por cuidar de mí y animarme a seguir adelante. Por tu apoyo, afecto y comprensión. Por estar siempre dispuesto a ayudar en cualquier cosa que necesite, hasta cuando no te pido ayuda. En definitiva, muchas gracias por formar parte de mi vida.

INDEX

ABSTRACT

A main objective in conservation programs is to maintain genetic diversity, and the most efficient management strategy to achieve it is to apply the Optimal Contributions method. This method optimizes the contributions of breeding candidates by minimizing the global coancestry. This leads to the highest levels of genetic diversity, when measured as expected heterozygosity, and to an effective control of the increase of inbreeding. The fundamental parameter of the method is the coancestry matrix which, traditionally, has been obtained from pedigree data. The current availability of genome-wide information allows us to estimate coancestries with higher precision. However, many different genomic coancestry measures have been proposed and it is unknown which measure is more efficient to minimize the loss of genetic diversity. Thus, the general aim of this thesis was to investigate the efficiency of different genomic coancestry matrices in the management of conserved populations when the Optimal Contributions method is applied to maximize genetic diversity. These coancestry matrices were evaluated with real and simulated data and for undivided and subdivided populations. **Chapter 1** presents a comparison of the efficiency in retaining genetic diversity (measured as expected heterozygosity) of six different genomic coancestry matrices using real data from a farm turbot population. The matrices compared were those based on: i) the proportion of shared alleles (θ_{SIM}); ii) deviations of the observed number of alleles shared by two individuals from the expected number ($\theta_{L\&H}$); iii) the realized relationship matrix obtained by VanRaden's method 1 (θ_{VR1}); iv) the realized relationship matrix obtained by VanRaden's method 2 (θ_{VR2}); v) the realized relationship matrix obtained by Yang´s method (θ_{YAN}); and vi) identical by descent segments (θ_{SEG}). They were computed using thousands of SNP genotypes obtained through 2b-RAD technology. Optimizations in **Chapter 1** were performed for a single generation as only genotype data from two generations (parents and offspring) were available. There were large differences in the magnitude of the different coancestry coefficients. Moreover, comparisons between coefficients greatly varied (especially for self-coancestry), being the lowest correlations between θ_{SIM}, $\theta_{L\&H}$ or θ_{SEG} and θ_{VR2} or θ_{YAN}. Results revealed that management with matrices based on the proportion of shared alleles or on segments (θ_{SIM}, $\theta_{L\&H}$ and θ_{SEG}) retained higher variability than those based on realized genomic relationship matrices

(θ_{VR1}, θ_{VR2} and θ_{YAN}). The higher the diversity achieved the lower was the number of individuals selected to contribute to the next generation. As expected, maximizing heterozygosity pushed alleles toward intermediate frequencies. However, it has been pointed out that moving allele frequencies away from initial frequencies may be undesirable as particular adaptations to the environment can be lost. In **Chapter 2**, stochastic simulations were used to investigate the efficiency of $\theta_{L\&H}$ and θ_{VR2} in the management of an undivided population across 50 generations and both matrices were compared not only in terms of the genetic diversity maintained but also in terms of the associated changes in allele frequencies across generations. The results indicate that the use of $\theta_{L\&H}$ resulted in a higher genetic diversity but also in a higher change of allele frequencies than the use of θ_{VR2}. The differences between strategies were reduced when only SNPs with a minimum allele frequency (MAF) above a particular threshold (MAF > 0.05 and MAF > 0.25) were used to compute $\theta_{L\&H}$ and θ_{VR2} and when the Optimal Contributions method was applied in populations of smaller sizes (N = 20 *vs* N = 100). In **Chapter 3**, the evaluation of $\theta_{L\&H}$ and θ_{VR2} was extended to subdivided populations, also via computer simulations. When populations are subdivided into different breeding groups, it is possible to give different weights to the within- and between-subpopulation components of genetic diversity. When a higher weight is given to the within-subpopulation component, the levels of inbreeding within subpopulations can be restricted. The use of $\theta_{L\&H}$ was the best option for managing subdivided populations as it maintains more global diversity, leads to less inbreeding within subpopulations and to changes in frequencies similar to those observed when using θ_{VR2} when a large weight is given to the within-subpopulation term.

RESUMEN

Un objetivo fundamental en los programas de conservación es mantener la diversidad genética y la estrategia de gestión más eficiente para lograrlo es aplicar el método de Contribuciones Óptimas. Este método optimiza las contribuciones de los candidatos a reproductores minimizando el parentesco global, lo que conduce a los niveles más altos de diversidad genética, medida como heterocigosis esperada, y a un control efectivo del aumento de consanguinidad. El parámetro fundamental de este método es la matriz de parentesco. Esta matriz se ha obtenido tradicionalmente a partir del pedigrí, pero la disponibilidad actual de genotipos para un gran número de polimorfismos de un solo nucleótido (SNP) nos permite estimarla con una mayor precisión. Sin embargo, se han propuesto muchas medidas de parentesco genómico y se desconoce qué medida es la más apropiada para minimizar la pérdida de diversidad genética. Por lo tanto, el objetivo general de esta tesis fue investigar la eficiencia de diferentes matrices genómicas de parentesco en la gestión de poblaciones en programas de conservación, cuando se aplica el método de Contribuciones Óptimas. Las distintas matrices de parentesco genómico fueron evaluadas con datos reales y con datos simulados, tanto para poblaciones no divididas como para poblaciones subdivididas. En el **Capítulo 1** se presenta una comparación de la eficiencia en la retención de la diversidad genética (medida como heterocigosis esperada) de seis matrices genómicas, utilizando datos reales de una población cultivada de rodaballo. Las matrices comparadas fueron aquellas basadas en: i) la proporción de alelos compartidos por dos individuos (θ_{SIM}); ii) las desviaciones del número observado de alelos compartidos por dos individuos respecto del número esperado ($\theta_{L\&H}$); iii) la matriz de relaciones genómicas obtenida a través el método 1 de VanRaden (θ_{VR1}); iv) la matriz de relaciones genómicas obtenida a través el método 2 de VanRaden (θ_{VR2}); v) la matriz de relaciones genómicas obtenida a través el método de Yang (θ_{YAN}); y vi) segmentos idénticos por descendencia (θ_{SEG}). Estas matrices se obtuvieron utilizando miles de genotipos de SNP obtenidos a través de la tecnología 2b-RAD. Las optimizaciones en el **Capítulo 1** se realizaron para una sola generación ya que solo estaban disponibles datos de genotipado para dos generaciones (padres e hijos). Las diferencias en la magnitud de los diferentes coeficientes de parentesco fueron grandes y las correlaciones entre ellos variaron

ampliamente (especialmente para el auto-parentesco). Las correlaciones más bajas fueron aquellas entre θ_{SIM}, $\theta_{L\&H}$ o θ_{SEG} y θ_{VR2} o θ_{YAN}. Los resultados mostraron que la gestión que utiliza matrices basadas en la proporción de alelos compartidos o en segmentos (θ_{SIM}, $\theta_{L\&H}$ y θ_{SEG}) retuvieron una mayor diversidad que aquella que utiliza matrices de relaciones genómicas (θ_{VR1}, θ_{VR2} y θ_{YAN}). Cuanto mayor fue la diversidad genética alcanzada, menor fue el número de individuos seleccionados para contribuir a la siguiente generación. Como era de esperar, la maximización de la heterocigosis llevó los alelos hacia frecuencias intermedias. Sin embargo, se ha señalado que alejar las frecuencias alélicas de las frecuencias iniciales puede ser indeseable, ya que se pueden perder adaptaciones particulares al medio. En el **Capítulo 2**, se utilizaron simulaciones estocásticas para investigar la eficiencia de $\theta_{L\&H}$ y θ_{VR2} en el manejo de poblaciones no divididas a lo largo de 50 generaciones y ambas matrices se compararon no solo en términos de la diversidad genética sino también en términos de los cambios asociados en las frecuencias alélicas. Los resultados indicaron que el uso de $\theta_{L\&H}$ resultó en una mayor diversidad genética pero también en un mayor cambio de frecuencias alélicas que el uso de θ_{VR2}. Las diferencias entre estrategias fueron menores cuando sólo se usaron SNP con una frecuencia del alelo menos común (MAF) por encima de un umbral particular (MAF > 0.05 y MAF > 0.25) para calcular $\theta_{L\&H}$ y θ_{VR2} y cuando se aplicó el método de Contribuciones Óptimas en poblaciones de tamaños más pequeños (se pasó de $N = 100$ a $N = 20$). En el **Capítulo 3**, la evaluación de $\theta_{L\&H}$ y θ_{VR2} se extendió a poblaciones subdivididas, también a través de simulaciones por ordenador. En poblaciones subdivididas, la diversidad genética se distribuye en dos componentes: dentro y entre subpoblaciones. Cuando se otorga un mayor peso al componente dentro de subpoblaciones, es posible restringir los niveles de consanguinidad dentro de subpoblaciones. Bajo este escenario, la utilización de $\theta_{L\&H}$ resultó ser la mejor opción para gestionar este tipo de poblaciones, ya que mantiene una mayor diversidad global, condujo a una menor consanguinidad dentro de subpoblaciones y a cambios en las frecuencias similares a los observados cuando se utilizó θ_{VR2}.

RESUM

Un objectiu fonamental en els programes de conservació és mantenir la diversitat genètica i l'estratègia de gestió més eficient per aconseguir-ho és aplicar el mètode de contribucions òptimes. Aquest mètode optimitza les contribucions dels reproductors candidats minimitzant el parentiu global, la qual cosa condueix als nivells més alts de diversitat genètica, mesurada com a heterocigosi esperada, i a un control efectiu de l'augment de consanguinitat. El paràmetre fonamental d'aquest mètode és la matriu de parentiu. Aquesta matriu s'ha obtingut tradicionalment a partir del pedigrí, però la disponibilitat actual de genotips per a un gran nombre de polimorfismes d'un sol nucleòtid (SNP) ens permet estimar-la amb més precisió. No obstant això, s'han proposat moltes mesures de parentiu genòmic i es desconeix quina mesura és la més apropiada per minimitzar la pèrdua de diversitat genètica. Per tant, l'objectiu general d'aquesta tesi va ser investigar l'eficiència de diferents matrius genòmiques de parentiu en la gestió de poblacions en programes de conservació, quan s'aplica el mètode de Contribucions Òptimes. Les diferents matrius de parentiu genòmic van ser avaluades amb dades reals i amb dades simulades, tant per a poblacions no dividides com per a poblacions subdividides. Al **Capítol 1** es va presentar una comparació de l'eficiència en la retenció de la diversitat genètica (mesurada com a heterocigosi esperada) de sis matrius genòmiques, utilitzant dades reals d'una població cultivada de rèvola. Les matrius comparades van ser aquelles basades en: i) la proporció d'al·lels compartits per dos individus (θ_{SIM}); ii) les desviacions del nombre observat d'al·lels compartits per dos individus del número esperat ($\theta_{L\&H}$); iii) la matriu de relacions genòmiques obtinguda a través del mètode 1 de VanRaden (θ_{VR1}); iv) la matriu de relacions genòmiques obtinguda a través del mètode 2 de VanRaden (θ_{VR2}); v) la matriu de relacions genòmiques obtinguda a través del mètode de Yang (θ_{YAN}); i vi) segments idèntics per descendència (θ_{SEG}). Aquestes matrius es van obtenir utilitzant milers de genotips de SNP obtinguts a través de la tecnologia 2b-RAD. Les optimitzacions al **Capítol 1** es van realitzar per a una sola generació ja que només estaven disponibles dades de genotipat per a dues generacions (pares i fills). Les diferències en la magnitud dels diferents coeficients de parentiu van ser grans i les correlacions entre ells van variar àmpliament (especialment per a l'autoparentiu). Les correlacions més baixes van ser aquelles entre θ_{SIM}, $\theta_{L\&H}$ o θ_{SEG}

i θ_{VR2} o θ_{YAN}. Els resultats van mostrar que la gestió que utilitza matrius basades en la proporció d'al·lels compartits o en segments (θ_{SIM}, $\theta_{L\&H}$ i θ_{SEG}) van retindre una major diversitat que aquella que utilitza matrius de relacions genòmiques (θ_{VR1}, θ_{VR2} i θ_{YAN}). Com més gran va ser la diversitat genètica aconseguida, menor va ser el nombre d'individus seleccionats per contribuir a la següent generació. Com era d'esperar, la maximització de l'heterocigosi va portar els al·lels cap a freqüències intermèdies. No obstant això, s'ha assenyalat que allunyar les freqüències al·lèliques de les freqüències inicials pot ser indesitjable, ja que es poden perdre adaptacions particulars al medi. Al **Capítol 2**, es van utilitzar simulacions estocàstiques per investigar l'eficiència de $\theta_{L\&H}$ i θ_{VR2} en el maneig de poblacions no dividides al llarg de 50 generacions i totes dues matrius es van comparar no sols en termes de la diversitat genètica sinó també en termes dels canvis associats en les freqüències al·lèliques. Els resultats van indicar que l'ús de $\theta_{L\&H}$ va resultar en una major diversitat genètica però també en un major canvi de freqüències al·lèliques que l'ús de θ_{VR2}. Les diferències entre estratègies van ser menors quan només es van fer servir SNP amb una freqüència de l'al·lel menys comú (MAF) per sobre d'un llindar particular (MAF > 0.05 i MAF > 0.25) per calcular $\theta_{L\&H}$ i θ_{VR2} i quan es va aplicar el mètode de Contribucions Òptimes en poblacions de mides més petites (es va passar de N = 100 a N = 20). Al **Capítol 3**, l'avaluació de $\theta_{L\&H}$ i θ_{VR2} es va estendre a poblacions subdividides, també a través de simulacions per ordinador. En poblacions subdividides, la diversitat genètica es compon de dos components: dins i entre subpoblacions. Quan s'atorga un major pes al component dins de subpoblacions, és possible restringir els nivells de consanguinitat dins de subpoblacions. Sota aquest escenari, la utilització de $\theta_{L\&H}$ va resultar ser la millor opció per gestionar aquest tipus de poblacions, ja que manté una major diversitat global, va conduir a una menor consanguinitat dins de subpoblacions i a canvis en les freqüències similars als observats quan es va utilitzar θ_{VR2}.

GENERAL INTRODUCTION

Genetic diversity

The maintenance of genetic diversity in animal populations is fundamental for their correct development and to avoid their extinction. For wild species, the International Union for Conservation of Nature and Natural Resources (IUCN) recognizes the need to preserve genetic diversity as it is the foundation for biodiversity and is necessary for long-term survival, adaptation, and resilience of populations, species, and entire ecosystems. In 1964, the IUCN established the Red List of Threatened Species in order to inform on the global extinction risk status of animal, fungus and plant species, and catalyze actions for biodiversity conservation. In this list, threatened animal species fall into the categories of critically endangered, endangered, and vulnerable (IUCN, 1994). Currently, there are more than 147,500 species on the IUCN Red List, with more than 41,000 species threatened with extinction (IUCN, 2022).

Extinction population risk is not limited to wild species but also applies to farm species. The Food and Agriculture Organization of the United Nations (FAO), which is the institution that monitors the status of livestock genetic diversity worldwide, has also warned about the threats facing populations of these species. In fact, despite an increasing number of actions aimed at preserving biodiversity, the proportion of local breeds at risk of extinction is increasing exponentially (FAO, 2015; FAO, 2019). Indiscriminate crossbreeding and replacement of well adapted local breeds by exotic high-output breeds, are the main causes of genetic erosion that causes local breeds to become extinct or seriously endangered (FAO, 2015; Taberlet *et al*., 2008; Biscarini *et al*., 2015). Important losses in genetic diversity also occur within breeds. For some high-yielding breeds, within-breed diversity has been rapidly reduced due to high selection intensities and the use of few very popular sires (e.g., FAO, 2015).

Genetic diversity is the set of differences in the DNA sequence between species, between populations within species, and between individuals within populations (Woolliams & Oldenbroek, 2017). It is important to minimize its loss through population management given that high genetic diversity levels would increase the likelihood that populations will be able to respond effectively to challenges such as the emergence of new diseases or climatic changes, and to ensure their long-term survival (Frankham *et*

al., 2010). The larger the genetic diversity, the higher the probability of the population containing individuals with alleles contributing to the adaptation to specific conditions and, thus, the higher the probability that the population survives. Genetic diversity also provides the raw material for breeding programs aimed at improving productivity in farm populations and is essential for obtaining genetic progress through selection for economically important traits. Also, in livestock populations, the genetic diversity among breeds, strains or lines can be used to exploit complementarity and heterosis through crossbreeding programs.

When managing populations, in addition to minimizing the loss of genetic diversity, it is also necessary to control the rate at which inbreeding increases (ΔF), particularly to avoid its negative consequences in the short term. Inbreeding increases homozygosity at the expense of heterozygosity and this increase in homozygosity in turn increases the incidence of homozygous recessive defects which lead to the decrease of the population mean for many quantitative traits, particularly those related to fitness, a phenomenon known as inbreeding depression (Falconer & Mackay, 1996; Caballero, 2020). Inbreeding depression can have important consequences in the short term through the reductions in viability and fertility that can increase the extinction risk.

If V_A is the existing additive genetic variance in a particular generation, then the loss in the next generation is $V_A \Delta f$, where Δf is the rate at which coancestry increases (Falconer & Mackay, 1996). Therefore, minimizing the rate at which genetic diversity is lost can be achieved by minimizing the rate at which the average coancestry increases in the population. Under random mating, minimizing Δf is equivalent to minimizing ΔF (Caballero & Toro, 2000; Villanueva *et al*., 2010). Under non-random mating, the rate at which genetic diversity is lost is better measured by Δf. This is because the inbreeding coefficient for a particular individual is the coancestry coefficient between its parents but at the population level, the relationship between the average coancestry in a particular generation and the average inbreeding in the next generation depends on mating decisions (e.g., whether or not matings between relatives are avoided). Nevertheless, inbreeding depression depends on the levels of inbreeding and not on coancestry and thus, both coancestry and inbreeding are central concepts in population management. Therefore, in

genetic conservation programs, the main objectives are to maintain the largest possible amount of genetic diversity and to avoid inbreeding depression, particularly in fitness-related traits.

Both the loss of genetic diversity and the increase in inbreeding are associated to genetic drift that is the change in allele frequencies in a finite population from generation to generation due to random sampling. Genetic drift is greater in small populations, but it can be also important in large populations as its magnitude actually depends on the effective population size (N_e) (Frankham, 2005). The effective population size, another very important population parameter, is the size of a hypothetical idealized population (a randomly mated population with equal numbers of males and females, contributing uniform numbers of progeny, and not subject to other forces that change genetic diversity, such as mutation, migration and selection) that would result in the same ΔF or genetic drift. The effective size can be estimated from the rate of coancestry or from the rate of inbreeding as $N_e = 1/2\Delta f$ or $1/2\Delta F$. Both estimates are equivalent with random mating (Caballero & Toro, 2000) or with non-random mating if the level of non-randomness is constant across generations (Villanueva *et al*., 2010). In general, populations under conservation programs are small and, consequently, have a small N_e. Populations under genetic improvement programs can have also small N_e due to a reduced number of individuals contributing to the next generation.

In order to characterize, manage and monitor populations, specific measures of genetic diversity are required. One of the most commonly used measures is the expected heterozygosity (H_e), also called gene diversity (Nei, 1973) which is the heterozygosity that would be present in a population at Hardy-Weinberg equilibrium with the same allele frequencies as the population of interest. For a single locus, $H_e = 1 - \sum_{i=1}^{n} p_i^2$, where p_i is the frequency of allele i and n is the number of alleles. Single-locus H_e can be then averaged over all loci. Note that the maximum H_e will occur when all the alleles of a given locus are at the same frequency. Importantly, high levels of H_e also imply high levels of additive genetic variance and, thus, high genetic responses from natural or artificial selection (Falconer & Mackay, 1996). Moreover, maximizing H_e is equivalent to maximizing the effective number of alleles, that is, the number of equally frequent

alleles that would lead to the same H_e as in the studied population (Kimura & Crow, 1964). Most studies on conservation genetics focus on H_e for the management and monitoring of genetic diversity (e.g., de Cara *et al*., 2011, 2013; Gómez-Romano *et al*., 2013; Eynard *et al*., 2016; Kleinman-Ruiz *et al*., 2019). This measure has been also used to monitor the loss of genetic diversity in selection programs (Eynard *et al*., 2016) and to characterize the genetic diversity conserved in gene banks (Eynard *et al*., 2018). It should be noted that H_e is directly related to the average coancestry coefficient of the population (f) as $H_e = 1 - f$ (Frankham *et al*., 2010; Toro *et al*., 2009). Consequently, populations with low levels of coancestry will have high levels of H_e.

Although H_e is the most used measure of genetic diversity, there are other measures including the i) observed heterozygosity (H_o) which is the proportion of heterozygous individuals in the population averaged across all loci in the genome; and ii) allelic diversity that is simply the number of different allelic variants segregating in the population. Allelic diversity is known to be more sensitive to bottlenecks than H_e and reflects better past fluctuations in N_e (Nei *et al*., 1975; Luikart *et al*., 1998). Allelic diversity is also essential for the long-term evolutionary potential of populations because the limit of selection response is determined by the initial number of alleles (assuming that mutation is negligible), regardless of the allele frequencies (James, 1970; Hill & Rasbash 1986; Caballero & García-Dorado 2013; Vilas *et al*., 2015). Finally, allelic diversity would be a better tool to monitoring the loss of rare variants (i.e. variants with alleles at very low frequencies) given that this loss has a limited impact on the average levels of H_e but may have a large effect on allelic diversity (Eynard *et al*., 2016).

Managing genetic diversity and inbreeding

One of the simplest management strategies to maintain genetic diversity and control inbreeding is to equalize parental contributions; i.e., all individuals in the population contribute exactly with the same number of offspring to the next generation (Gowe *et al*., 1959; Wang, 1997). This strategy leads to rates of inbreeding and coancestry that are about half as large as those obtained under panmixia; i.e., random

contributions and mating (Fernández & Caballero, 2001). When equalizing contributions, the lower magnitude of genetic drift (Fernández & Caballero, 2001) diminishes the probability of random loss of alleles, and the lower levels of inbreeding reduce the depression in fitness-related traits. However, the equalization of parental contributions also reduces the intensity of natural selection, given that differences in fecundity among parents are obviated except for complete mating failures (Wang, 1997; Sánchez *et al*., 2003). Therefore, as a side effect, there is a tendency to accumulate detrimental variants and to increase the genetic load of individuals.

A great amount of research was carried out in the 90s with the aim of developing selection and mating strategies to control inbreeding and avoid its negative consequences (i.e., inbreeding depression and reduced diversity) in artificial selection programs. Proposed selection strategies included to i) increase the number of selected individuals; ii) increase the number of selected individuals and allow them to contribute differentially; iii) restrict the number of individuals selected per family; and iv) reduce the emphasis given to family information in the selection criterion (e.g., Toro & Pérez-Enciso, 1990; Villanueva *et al.*, 1994). These strategies considered rates of gain and inbreeding separately and, although they were succesful in controlling the increase in inbreeding, they also led to losses in the response to selection. A selection strategy that simultaneously manage genetic gain and inbreeding in selection decisions was subsequently developed by Meuwissen (1997) and Grundy *et al*. (1998). This dynamic method places a direct constraint on ΔF while the contributions of selected candidates (i.e., number of offspring to be produced form each breeding candidate) is optimized for maximizing genetic gain. This methodology is known as the Optimal Contribution (OC) selection method. Selection decisions are optimized in order to manage ΔF without implying any loss in genetic gain. With the OC method, the numbers of individuals selected and their contributions (i.e., number of offspring) are not fixed but optimized each generation by taking into account not only the estimated breeding values of the candidates but also their genetic relationships (Meuwissen, 1997; Grundy *et al*., 1998; Woolliams *et al.*, 2015). The fundamental parameter for optimizing the contributions of all potential breeding candidates is the additive genetic relationship matrix, **A**, or

equivalently the coancestry matrix ($\boldsymbol{\theta}$) as $\boldsymbol{\theta} = \mathbf{A}/2$. Matrix $\boldsymbol{\theta}$ contains the coancestry coefficients between all pairs of candidates (Falconer & Mackay, 1996; Lynch & Walsh, 1998).

Although the OC method was developed in the context of genetic improvement programs, its application to conservation programs is straightforward (Fernández *et al*., 2003; Villanueva *et al.*, 2004). In this case, the objective is to minimize the rate of coancestry with the aim of maintaining the highest possible levels of genetic diversity and reducing the rise of inbreeding. It must be highlighted that, due to the relationship between f and H_e (i.e., $H_e = 1 - f$), the OC methodology applied to conservation programs is directed to the maximization of the genetic diversity measured as H_e.

Given that H_e reaches its maximum value at intermediate allele frequencies, in principle, maximizing H_e would have an extra positive effect in terms of conserving rare alleles. In fact, if rare alleles are pushed toward intermediate frequencies, their probability of being lost would be reduced (Fernández *et al*., 2004). However, it has also been pointed out that moving allele frequencies away from initial frequencies may be undesirable as particular adaptations to the environment can be lost and the frequency of deleterious mutations can increase (Lacy, 2000; Frankham, 2008; Saura *et al*., 2008) and affect the fitness of the population (Schoen *et al*., 1998; Fernández & Caballero, 2001; Theodorou & Couvet, 2003; Rodríguez-Ramilo *et al*., 2006). Thus, when applying OC method it is worth to investigate how allele frequencies have changed in the population.

Management strategies described so far refer to undivided populations. However, for several reasons, both farm and wild populations can be subdivided into different breeding groups that are more or less disconnected. These reasons can be logistic (e.g., resource and space limitations to keep the population in one single location) or biological (e.g., different subpopulations may be characterized by local adaptations) (Fernández *et al*., 2008). Although the subdivision of populations has some advantages, it can also lead to problems as subpopulations often have low N_e which leads to rapid increases in inbreeding and losses of genetic diversity, increasing the risk of their extinction (Falconer & Mackay, 1996). Many studies on subdivided populations have investigated the distribution of the total genetic diversity into within and between

subpopulations terms (e.g., Wright, 1931; Eding & Meuwissen, 2001; Comps *et al*., 2001; Foulley & Ollivier, 2006; Caballero & Rodríguez-Ramilo, 2010; Whitlock, 2011). However, studies investigating the management of subdivided populations are scarce. Due to the problems that isolation can cause for subpopulations, it is important to allow (or even force) a certain degree of migration between them. A simple procedure to achieve this is the one-migrant-per-generation rule (OMPG) (see Wang, 2004), based on the island model derived by Wright (1931). It allows for an average of one migrant per generation and subpopulation and was the standard approach to manage subdivided populations in the past. A limitation of the OMPG strategy is that it does not account for the genetic structure of the population (i.e., the particular relationship between subpopulations and the inbreeding within subpopulations). Therefore, the subpopulations involved in the exchange of individuals are chosen at random which could be suboptimal for the control of inbreeding. Moreover, the average number of migrants is always one. To solve these drawbacks, Fernández *et al*. (2008) developed an extension of the OC method to optimally manage subdivided populations. Their method determines the optimal contribution of each individual to maximize the global genetic diversity but also implicitly optimizes the migration flow between subpopulations, determining the optimal migration rate and the specific subpopulations involved in the exchange of individuals. Additionally, with subdivided populations there is a possibility of assigning different relative weights to the between- and within-subpopulation components of coancestry. Increasing the relative weight given to within-subpopulation component allows the control of inbreeding within subpopulations. All in all, the consequences of the management of subdivided populations through the OC method must be assessed in terms of the level of the global genetic diversity maintained, the distribution of this diversity between- and within-subpopulations, the inbreeding within subpopulations, the particular migration pattern, and the change in allele frequencies.

Managing genetic diversity using genomic information

Traditionally, the coancestry matrix (the central element in the OC method) has been computed from pedigree records (Meuwissen, 1997; Grundy *et al*., 1998; Fernández

et al., 2003). Pedigree-based coancestry coefficients provide expectations of the proportion of alleles that two particular animals have in common; for example, it implicitly assumes that half of the alleles of a pair of full-sibs are identical by descent from their parents. However, these expectations can differ from the exact proportions not only for full-sibs but for any other type of relationship, with the exception of parent-offspring and ignoring sex chromosomes (Christensen *et al.*, 1996). Molecular markers can be used to estimate these proportions with a high degree of precision and, therefore, to reflect the true proportion of the genome in common. Also, pedigree recording can be very difficult, if not impossible, in wild animal populations (Garant & Kruuk, 2005; Keller *et al*., 2011) while molecular information can be obtained for all types of populations and from practically any sample of animal origin. Finally, in populations of livestock species, where obtaining pedigree records is a common practice, pedigrees often contain errors. For instance, the pedigree error rate in dairy cattle was 10% in the UK (Visscher *et al*., 2002) and 7 to 9% in Ireland (McClure *et al*., 2018). In a conservation context, Oliehoek & Bijma (2009) showed that when the parental assignment error rate is above 35%, the OC method preserves less genetic diversity than simply equalizing the contributions of the candidates.

The efficiency of using molecular coancestries in OC for maintaining genetic diversity was first investigated by Fernández *et al*. (2005). They simulated a coancestry matrix computed from a reduced number of microsatellite molecular markers and concluded that the exclusive use of molecular information in OC was of very limited value when the aim was to maintain genetic diversity. However, when the matrix is computed from dense SNP panels, several studies have showed that genomic coancestry is more effective than pedigree-based coancestry for maintaining genetic diversity, measured as H_e (de Cara *et al*., 2011; Gómez-Romano *et al*., 2013) or as ΔF (Eynard *et al*., 2016). These previous studies used specific genomic measures. For instance, the coancestry measure used by de Cara *et al*. (2011) and Gómez-Romano *et al*. (2013) was simply the proportion of alleles shared by two individuals. Eynard *et al*. (2016) used a similar measure and also one based on the realized genomic relationship matrix proposed by Yang *et al*. (2010). However, many other measures of genomic coancestry (and

inbreeding) have been proposed (Villanueva *et al*., 2021; Caballero *et al*., 2022) and their efficiency when used in OC for maintaining diversity and controlling inbreeding need to be evaluated.

The different measures of genomic coancestry proposed so far can be grouped into different categories. The simplest way of measuring coancestry based on SNP data is to compute SNP-by-SNP similarity (Nejati-Javaremi *et al.*, 1997). This measure, used by de Cara *et al*. (2011) and Gómez-Romano *et al*. (2013), does not distinguish between identity-by-state (IBS) and identity-by-descent (IBD). The equivalent measure of inbreeding is the proportion of homozygous SNPs in an individual. Note that IBS values are typically higher than IBD values given that alleles can be IBS for two reasons: i) they are IBD (copies of the same allele of the base population); or ii) they are not IBD, but coming from two alleles that were equal in the base population (e.g., Toro *et al*., 2014). A second category of genomic coancestries measures try to put IBS in an IBD scale. In their formulation, allele frequencies in a base population taken as a reference to compute coancestry are used to correct for the similarities originally present (Toro *et al*., 2002, 2014). In most cases these allele frequencies are not available and current frequencies are used to compute coancestry, but this can lead to biased estimates. Within this category, one of the most commonly used measure is that based on the deviations of the observed number of alleles shared by two individuals from the expected numbers under Hardy-Weinberg equilibrium. This coancestry coefficient is also computed on a SNP-by-SNP basis and was first proposed by Li & Horvitz (1953) for inbreeding and subsequently adapted for coancestry by Toro *et al.* (2002). It is worth noting that this coefficient correlates perfectly with the similarity (IBS) coancestry (and inbreeding) coefficient although they are at different scales (Villanueva *et al*., 2021). A third category corresponds to coancestry measures obtained from different realized genomic relationship matrices (VanRaden, 2008; Yang *et al*., 2010) that have been widely used in genome-wide evaluations and genome-wide association studies. These matrices have been also widely used to obtain genomic inbreeding coefficients (e.g., Keller *et al*., 2011; Bjelland *et al*., 2013; Pryce *et al*., 2014; Zhang *et al*., 2015; Mastrangelo *et al*., 2016; Solé *et al*., 2017; Caballero *et al*., 2020; Villanueva *et al*., 2021; Caballero *et al*., 2022).

The coancestry measures included in this third category are also computed on a SNP-by-SNP basis and their formulation includes base population allele frequencies. Finally, another genomic coancestry measure proposed is that based on IBD segments which are defined as long continuous segments of DNA that are identical in two individuals (Gusev *et al*., 2009; de Cara *et al*., 2013; Thompson, 2013; Chiang *et al*., 2016; Gómez-Romano *et al*., 2016b; Saada *et al*., 2020; Smith *et al*., 2022). Although what it is strictly observed is similarity (IBS) between individuals, it is expected that long enough segments shared by two individuals are very likely to be IBD. The equivalent measure for genomic inbreeding is that obtained from runs of homozygosity or ROH (McQuillan *et al*., 2008).

Although the main objective in conservations programs is to maintain the maximum possible genetic diversity, it has been also claimed that preserving the original allele frequencies may be desirable (Lacy, 2000; Frankham, 2008; Saura *et al*., 2008). Therefore, it is also important to evaluate the changes in allele frequencies through the management period. Using different genomic coefficients of coancestry in the OC methodology may have a different impact on the diversity maintained and the change in frequencies. In fact, Gómez-Romano *et al*. (2016a) suggested that while OC using a genomic coancestry matrix based on allele sharing tends to move allele frequencies towards intermediate values (0.5 in the case of biallelic SNPs), OC using realized genomic relationship matrices would lead to solutions where allele frequencies would tend to be unchanged. Recently, Meuwissen *et al*. (2020) investigated the use of different coancestry genomic matrices in OC in the context of a breeding program where the objective is to maximize genetic gain while restricting at the same time the increase in inbreeding (and the loss of genetic diversity). They concluded that the suggestion of Gómez-Romano *et al*. (2016a) seems to be correct since the use of the matrix based on allele sharing led to higher genetic drift than the use of realized genomic relationship matrices. However, more research is needed to find out if, in the context of a conservation program, different genomic coancestry matrices used in OC have the same effects. In addition, it would be interesting to know the effect of using these genomic coancestry matrices in OC not only on genetic diversity measured as H_e but also on allelic diversity.

Another area that requires further research is the optimal management of subdivided populations. Studies investigating the efficiency of the OC method applied to this type of populations (Fernández *et al*., 2008; Caballero *et al*., 2010; Ávila *et al*., 2011) have been carried out using pedigree-based coancestry coefficients. There is thus, a need to investigate the consequences of using different genomic coancestry measures in terms of the global genetic diversity maintained, the distribution of this diversity between- and within-subpopulations and the change in allele frequencies.

SNP arrays in farm and wild species

High density SNP arrays allow large numbers of individuals to be rapidly and cost-effectively genotyped for large numbers of genetic markers. Several studies have showed that SNP arrays played an essential role in the conservation of genetic diversity (e.g., Engelsma *et al*., 2012; de Cara *et al*., 2011; Gómez-Romano *et al*., 2013; Eynard *et al*., 2016). There are currently dense panels of SNPs available for most farm terrestrial species, including cattle (777K; Rincon *et al*., 2011), pigs (660K; van Son *et al*., 2017), sheep (600K; Kijas *et al*., 2016), horses (670K; Schaefer *et al*., 2017), chickens (600K SNP; Kranis *et al*., 2013) and alpaca (76K; Calderon *et al*., 2021). Also, although still far from that in terrestrial animals, the development of genomic tools in aquaculture species has been very significant. For example, there are commercial SNP arrays developed for common carp (250K*;* Xu *et al*., 2014), salmon (200K; Yáñez *et al*., 2016; Barría *et al*., 2019) and rainbow trout (665K; Bernard *et al*., 2022). Combined arrays for several species have been also developed to save costs. For instance, an array of 60K SNPs has been recently developed for European seabass and gilthead seabream (Peñaloza *et al*., 2021). For species of commercial interest for which there are no arrays developed, genotyping by sequencing (GBS) techniques, including restriction site- associated DNA sequencing (RAD-seq) (Baird *et al*., 2008) and derivatives, can been applied to obtain SNP data at the population level without the need for a reference genome (Houston *et al*., 2020).

For non-farm species, progress in the development of SNP arrays has been much slower than in farm species given that the direct economic return from the maintenance of non-farm populations of these species is less obvious and resources are very limited (Norman *et al*., 2019). However, there are a handful of medium-high density SNP arrays designed for several species including arrays of ~90K SNPs for the Antarctic fur seal (Humble *et al*., 2020), ~50K SNPs for zebra finch (Lee *et al*., 2021), ~10K SNPs for the house sparrow (Hagen *et al*., 2013; Lundregan *et al*., 2018), ~500K for the great tit (van Bers *et al*., 2012; Kim *et al*., 2018), 9K for the polar bear (Malenfant *et al*., 2015) and 50K for the bald eagle (Judkins *et al*., 2020).

Notwithstanding, with the advent of next generation sequencing, new avenues have opened to include genomics in studies on wild populations of non-model species. For instance, research using different numbers of SNPs has been carried out for Iberian linx (~1500 SNPs; Kleinman-Ruiz *et al*., 2017), Tasmanian devil (~200 SNPs; Hogg *et al*., 2019), invasive comb jelly (~100 SNPs; Pujolar *et al*., 2022), or brown bear (~100 SNPs; Norman & Spong, 2015). Also, although most SNP arrays have been developed for domestic species, they can be used to obtain subsets of SNPs for wild relatives. For example, SNP arrays developed for bovine and ovine species have been used for genome analyses in reindeer (Kharzinova *et al*., 2015) and loci selected from the canine SNP array has been used for wildlife monitoring in grey wolf (Kraus *et al*., 2015). Bovine arrays have been also used for genetic analyses in antelope (Ogden, 2012) and addax (Ivy *et al*., 2016) species and a chicken array has been used for North American prairie grouse species (Minias *et al*., 2019).

All this molecular information will be very useful to manage the maintenance of genetic diversity in both undivided and subdivided populations under conservation programs. For this reason, the study of the outcomes of OC management when using different matrices of genomic coancestry is so valuable.

References

Ávila, V., Fernández, J., Quesada, H. & Caballero, A. (2011). An experimental evaluation

with *Drosophila melanogaster* of a novel dynamic system for the management of subdivided populations in conservation programs. *Heredity*, 106, 765–774. https://doi.org/10.1038/hdy.2010.117.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A. *et al*. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS ONE*, 3, e3376. https://doi.org/10.1371/journal.pone.0003376.

Barría, A., Christensen, K.A., Yoshida, G., Jedlicki, A., Leong, J.S., Rondeau, E.B. *et al*. (2019). Whole genome linkage disequilibrium and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding population using a high-density SNP array. *Frontiers in Genetics*, 10, 498. https://doi.org/10.3389/fgene.2019.00498.

Bernard, M., Dehaullon, A., Gao, G., Paul, K., Lagarde, H., Charles, M. *et al*. (2022). Development of a high-density 665 K SNP array for rainbow trout genome-wide genotyping. *Frontiers in Genetics*, 13, 941340. https://doi.org/10.3389/fgene.2022.941340.

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J. & Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Frontiers in Genetics*, 6, 33. https://doi.org/10.3389/fgene.2015.00033.

Bjelland, D.W., Weigel, K., Vukasinovic, N. & Nkrumah, J.D. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science*, 96, 4697–4706. https://doi.org/10.3168/jds.2012-6435.

Caballero, A. (2020). Quantitative genetics. Cambridge University Press.

Caballero, A., Fernández, A., Villanueva, B. & Toro, M. A. (2022). A comparison of marker-based estimators of inbreeding and inbreeding depression. *Genetics Selection Evolution*, 54, 82. https://doi.org/10.1186/s12711-022-00772-0.

Caballero, A. & García-Dorado, A. (2013). Allelic diversity and its implications for the rate of adaptation. *Genetics*, 195, 1373–1384.

https://doi.org/10.1534/genetics.113.158410.

Caballero, A. & Rodríguez-Ramilo, S. T. (2010). A new method for the partition of allelic diversity within and between subpopulations. *Conservation Genetics*, 11, 2219–2229. https://doi.org/10.1007/s10592-010-0107-7.

Caballero, A., Rodríguez-Ramilo, S. T., Avila, V. & Fernández, J. (2010). Management of genetic diversity of subdivided populations in conservation programmes. *Conservation Genetics*, 11, 409–419. https://doi.org/10.1007/s10592-009-0020-0.

Caballero, A. & Toro, M. (2000). Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetics Research*, 75, 331–343. https://doi.org/10.1017/S0016672399004449.

Caballero, A., Villanueva, B. & Druet, T. (2020). On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. *Evolutionary Applications*, 14, 416–428. https://doi.org/10.1111/eva.13126.

Calderon, M., More, M. J., Gutierrez, G. A. & Ponce de León, F. A. (2021). Development of a 76k alpaca (*Vicugna pacos*) single nucleotide polymorphisms (SNPs) microarray. *Genes*, 12, 291. https://doi.org/10.3390/genes12020291.

Chiang, C. W., Ralph, P. & Novembre, J. (2016). Conflation of short identity-by-descent segments bias their inferred length distribution. *G3: Genes, Genomes, Genetics*, 6, 1287–1296. https://doi.org/10.1534/g3.116.027581.

Comps, B., Gömöry, D., Letouzey, J., Thiébaut, B. & Petit, R. J. (2001). Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, 157, 389–397. https://doi.org/10.1093/genetics/157.1.389.

Christensen, K., Fredholm, M., Winterø, A. K., Jørgensen, J. N. & Andersen, S. (1996). Joint effect of 21 marker loci and effect of realized inbreeding on growth in pigs. *Animal Science*, 62, 541–546. https://doi.org/10.1017/S1357729800015083.

de Cara, M.A.R., Fernández, J., Toro, M.A. & Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics*, 128, 456–464. https://doi.org/10.1111/j.1439-0388.2011.00971.x.

de Cara, M.A.R., Villanueva, B., Toro, M.A. & Fernández, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology*, 22, 6091–6099. https://doi.org/10.1111/mec.12560.

Eding, H. & Meuwissen, T. H. E. (2001). Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics*, 118, 141–159. https://doi.org/10.1046/j.1439-0388.2001.00290.x.

Engelsma, K.A., Veerkamp, R.F., Calus, M.P.L., Bijma, P. & Windig, J.J. (2012). Pedigree- and marker-based methods in the estimation of genetic diversity in small groups of Holstein cattle. *Journal of Animal Breeding and Genetics*, 129, 195–205. https://doi.org/10.1111/j.1439-0388.2012.00987.x.

Eynard, S. E., Windig, J. J., Hiemstra, S. J. & Calus, M. P. (2016). Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genetics Selection Evolution*, 48, 33. https://doi.org/10.1186/s12711-016-0210-4.

Eynard, S. E., Windig, J. J., Hulsegge, I., Hiemstra, S. J. & Calus, M. P. (2018). The impact of using old germplasm on genetic merit and diversity—A cattle breed case study. *Journal of Animal Breeding and Genetics*, 135, 311–322. https://doi.org/10.1111/jbg.12333.

Falconer, D.S. & Mackay, F.C. (1996). Introduction to Quantitative Genetics. 4th ed. Longman Group Ltd, Harlow, Essex, England.

FAO. (2015). The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture. FAO Commission on Genetic Resources for Food and Agriculture, Rome, Italy.

FAO. (2019). The State of the World's Biodiversity for Food and Agriculture. FAO

Commission on genetic resources for food and agriculture assessments, Rome, Italy.

Fernández, J. & Caballero, A. (2001). Accumulation of deleterious mutations and equalization of parental contributions in the conservation of genetic resources. *Heredity*, 86, 480–488. https://doi.org/10.1046/j.1365-2540.2001.00851.x.

Fernández, J., Toro, M.A. & Caballero, A. (2003). Fixed contributions designs *vs*. minimization of global coancestry to control inbreeding in small populations. *Genetics*, 165, 885–894. https://doi.org/10.1093/genetics/165.2.885.

Fernández, J., Toro, M. A. & Caballero, A. (2004). Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conservation Biology*, 18, 1358–1367. https://doi.org/10.1111/j.1523-1739.2004.00341.x.

Fernández, J., Toro, M. A. & Caballero, A. (2008). Management of subdivided populations in conservation programs: development of a novel dynamic system. *Genetics*, 179, 683–692. https://doi.org/10.1534/genetics.107.083816.

Fernández, J., Villanueva, B., Pong-Wong, R. & Toro, M. A. (2005). Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genetics*, 170, 1313–1321. https://doi.org/10.1534/genetics.104.037325.

Foulley, J. L. & Ollivier, L. (2006). Estimating allelic richness and its diversity. *Livestock Science*, 101, 150–158. https://doi.org/10.1016/j.livprodsci.2005.10.021.

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, 126, 131–140. https://doi.org/10.1016/j.biocon.2005.05.002.

Frankham, R. (2008). Genetic adaptation to captivity in species conservation programs. *Molecular Ecology*, 17, 325–333. https://doi.org/10.1111/j.1365-294X.2007.03399.x.

Frankham, R., Ballou, J. D. & Briscoe D. A. (2010). *Introduction to conservation genetics*. 2nd ed. Cambridge: Cambridge University Press, United Kingdom.

Garant, D. & Kruuk, L. E. (2005). How to use molecular marker data to measure evolutionary parameters in wild populations. *Molecular Ecology*, 14, 1843–1859. https://doi.org/10.1111/j.1365-294X.2005.02561.x.

Gómez-Romano, F., Villanueva, B., de Cara, M.A.R & Fernández, J. (2013). Maintaining genetic diversity using molecular coancestry: The effect of marker density and effective population size. *Genetics Selection Evolution*, 45, 38. https://doi.org/10.1186/1297-9686-45-38.

Gómez-Romano, F.; Villanueva, B.; Fernández, J.; Woolliams, J.A.; & Pong-Wong, R. (2016a). The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution*, 48, 2. https://doi.org/10.1186/s12711-015-0172-y.

Gómez-Romano, F., Villanueva, B., Sölkner, J., de Cara, M. Á. R., Mészáros, G., Pérez O'Brien, A. M. & Fernández, J. (2016b). The use of coancestry based on shared segments for maintaining genetic diversity. *Journal of Animal Breeding and Genetics*, 133, 357–365. https://doi.org/10.1111/jbg.12213.

Gowe, R. S., Robertson, A. & Latter, B. D. H. (1959). Environment and poultry breeding problems: 5. The design of poultry control strains. *Poultry Science*, 38, 462–471. https://doi.org/10.3382/ps.0380462.

Grundy, B., Villanueva, B. & Wooliams, J.A. (1998). Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetics Research*, 72, 159–168. https://doi.org/10.1017/S0016672398003474.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L. *et al*. (2009). Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, 19, 318–326. https://doi.org/10.1101/gr.081398.108.

Hagen, I. J., Billing, A. M., Rønning, B., Pedersen, S. A., Pärn, H., Slate, J. & Jensen, H. (2013). The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10 K SNP-chip for the house sparrow (*Passer domesticus*). *Molecular Ecology Resources*, 13, 429–439.

https://doi.org/10.1111/1755-0998.12088.

Hill, W. G. & Rasbash, J. (1986). Models of long term artificial selection in finite population. *Genetics Research*, 48, 41–50. https://doi.org/10.1017/S0016672300024642.

Hogg, C. J., Wright, B., Morris, K. M., Lee, A. V., Ivy, J. A., Grueber, C. E. & Belov, K. (2019). Founder relationships and conservation management: empirical kinships reveal the effect on breeding programmes when founders are assumed to be unrelated. *Animal Conservation*, 22, 348–361. https://doi.org/10.1111/acv.12463.

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L. *et al*. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, 21, 389–409. https://doi.org/10.1038/s41576-020-0227-y.

Humble, E., Paijmans, A. J., Forcada, J. & Hoffman, J. I. (2020). An 85K SNP array uncovers inbreeding and cryptic relatedness in an Antarctic fur seal breeding colony. *G3: Genes, Genomes, Genetics*, 10, 2787–2799. https://doi.org/10.1534/g3.120.401268.

IUCN. (1994). IUCN Red List Categories. IUCN Species Survival Commission. Prepared by the *Standards and Petitions Committee*. IUCN, Gland, Switzerland and Cambridge, U.K. Downloadable from https://portals.iucn.org/library/sites/library/files/documents/RL-1994-001.pdf.

IUCN (2022). Guidelines for Using the IUCN Red List Categories and Criteria. Version 15.1. Prepared by the *Standards and Petitions Committee*. IUCN, Gland, Switzerland and Cambridge, U.K. Downloadable from https://www.iucnredlist.org/documents/RedListGuidelines.pdf.

Ivy, J. A., Putnam, A. S., Navarro, A. Y., Gurr, J. & Ryder, O. A. (2016). Applying SNP-derived molecular coancestry estimates to captive breeding programs. *Journal of Heredity*, 107, 403–412. https://doi.org/10.1093/jhered/esw029.

James, J. W. (1970). The founder effect and response to artificial selection. *Genetics Research*, 16, 241-250.

Judkins, M. E., Couger, B. M., Warren, W. C. & Van Den Bussche, R. A. (2020). A 50K SNP array reveals genetic structure for bald eagles (*Haliaeetus leucocephalus*). *Conservation Genetics*, 21, 65–76. https://doi.org/10.1007/s10592-019-01216-x.

Keller, M.C., Visscher, P.M. & Goddard, M.E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics*, 189, 237–249. https://doi.org/10.1534/genetics.111.130922.

Kijas, J. W., Hadfield, T., Naval Sanchez, M. & Cockett, N. (2016). Genome-wide association reveals the locus responsible for four-horned ruminant. *Animal Genetics*, 47, 258–262. https://doi.org/10.1111/age.12409.

Kim, J. M., Santure, A. W., Barton, H. J., Quinn, J. L., Cole, E. F., Great Tit HapMap Consortium *et al*. (2018). A high-density SNP chip for genotyping great tit (*Parus major*) populations and its application to studying the genetic architecture of exploration behaviour. *Molecular Ecology Resources*, 18, 877–891. https://doi.org/10.1111/1755-0998.12778.

Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite population. *Genetics*, 49, 725. https://doi.org/10.1093/genetics/49.4.725.

Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F. *et al*. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, 18, 556. https://doi.org/10.1186/s12864-017-3946-5.

Kleinman-Ruiz, D., Soriano, L., Casas-Marce, M., Szychta, C., Sánchez, I., Fernández, J. & Godoy, J. A. (2019). Genetic evaluation of the Iberian lynx ex situ conservation programme. *Heredity*, 123, 647–661. https://doi.org/10.1038/s41437-019-0217-z.

Kharzinova, V. R., Sermyagin, A. A., Gladyr, E. A., Okhlopkov, I. M., Brem, G. & Zinovieva, N. A. (2015). A study of applicability of SNP chips developed for

bovine and ovine species to whole-genome analysis of reindeer Rangifer tarandus. *Journal of Heredity*, 106, 758-761. https://doi.org/10.1093/jhered/esv081.

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S. *et al*. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics*, 14, 59. https://doi.org/10.1186/1471-2164-14-59.

Kraus, R. H., Vonholdt, B., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R. *et al*. (2015). A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources*, 15, 295–305. https://doi.org/10.1111/1755-0998.12307.

Lacy, R. C. (2000). Should we select genetic alleles in our conservation breeding programs? *Zoo Biology*, 19, 279–282. https://doi.org/10.1002/1098-2361(2000)19:4<279::AID-ZOO5>3.0.CO;2-V.

Lee, K. D., Millar, C. D., Brekke, P., Whibley, A., Ewen, J. G., Hingston, M. *et al*. (2021). The design and application of a 50 K SNP chip for a threatened Aotearoa New Zealand passerine, the hihi. *Molecular Ecology Resources*, 22, 415–429. https://doi.org/10.1111/1755-0998.13480.

Li, C.C. & Horvitz, D.G. (1953). Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics*, 5, 107–117.

Luikart, G., Allendorf, F. W., Cornuet, J. M. & Sherwin, W. B. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, 89, 238–247. https://doi.org/10.1093/jhered/89.3.238.

Lundregan, S. L., Hagen, I. J., Gohli, J., Niskanen, A. K., Kemppainen, P., Ringsby, T. H. *et al*. (2018). Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Molecular Ecology*, 27, 3498–3514. https://doi.org/10.1111/mec.14811.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Vol. 1.

Sunderland, MA: Sinauer.

Malenfant, R. M., Coltman, D. W. & Davis, C. S. (2015). Design of a 9K illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. *Molecular Ecology Resources*, 15, 587–600. https://doi.org/10.1111/1755-0998.12327.

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M. T. & Portolano, B. (2016). Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal*, 10, 746–754. https://doi.org/10.1017/S1751731115002943.

McClure, M. C., McCarthy, J., Flynn, P., McClure, J. C., Dair, E., O'Connell, D. K. & Kearney, J. F. (2018). SNP data quality control in a national beef and dairy cattle system and highly accurate SNP based parentage verification and identification. *Frontiers in Genetics*, 9, 84. https://doi.org/10.3389/fgene.2018.00084.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L. *et al*. (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, 83, 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007.

Meuwissen, T.H.E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75, 934–940. https://doi.org/10.2527/1997.754934x.

Meuwissen, T. H. E., Sonesson, A. K., Gebregiwergis, G. & Woolliams, J. A. (2020). Management of genetic diversity in the era of genomics. *Frontiers in Genetics*, 11, 880. https://doi.org/10.3389/fgene.2020.00880.

Minias, P., Dunn, P. O., Whittingham, L. A., Johnson, J. A. & Oyler-McCance, S. J. (2019). Evaluation of a Chicken 600K SNP genotyping array in non-model species of grouse. *Scientific Reports*, 9, 6407. https://doi.org/10.1038/s41598-019-42885-5.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the*

National Academy of Sciences U.S.A., 70, 3321–3323. https://doi.org/10.1073/pnas.70.12.3321.

Nei, M., Maruyama, T. & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 29, 1–10. https://doi.org/10.2307/2407137.

Nejati-Javaremi, A., Smith, C. & Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science*, 75, 1738–1745. https://doi.org/10.2527/1997.7571738x.

Norman, A.J & Spong G. (2015). Single nucleotide polymorphism-based dispersal estimates using noninvasive sampling. *Ecology and Evolution*. 5, 3056–3065. https://doi.org/10.1002/ece3.1588.

Norman, A. J., Putnam, A. S. & Ivy, J. A. (2019). Use of molecular data in zoo and aquarium collection management: benefits, challenges, and best practices. *Zoo Biology*, 38, 106–118. https://doi.org/10.1002/zoo.21451.

Ogden, R., Baird, J., Senn, H. & McEwing, R. (2012). The use of cross-species genome-wide arrays to discover SNP markers for conservation genetics: a case study from Arabian and scimitar-horned oryx. *Conservation Genetics Resources*, 4, 471–473. https://doi.org/10.1007/s12686-011-9577-2.

Oliehoek, P. A. & Bijma, P. (2009). Effects of pedigree errors on the efficiency of conservation decisions. *Genetics Selection Evolution*, 41, 9. https://doi.org/10.1186/1297-9686-41-9.

Peñaloza, C., Manousaki, T., Franch, R., Tsakogiannis, A., Sonesson, A. K., Aslam, M. L. *et al*. (2021). Development and testing of a combined species SNP array for the European seabass (*Dicentrarchus labrax*) and gilthead seabream (*Sparus aurata*). *Genomics*, 113, 2096–2107. https://doi.org/10.1016/j.ygeno.2021.04.038.

Pryce, J.E., Haile-Mariam, M., Goddard, M.E. & Hayes, B.J. (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution*. 46, 71.

https://doi.org/10.1186/s12711-014-0071-7.

Pujolar, J. M., Limborg, M. T., Ehrlich, M. & Jaspers, C. (2022). High throughput SNP chip as cost effective new monitoring tool for assessing invasion dynamics in the comb jelly *Mnemiopsis leidyi*. *Frontiers in Marine Science*, 9, 1019001. https://doi.org/10.3389/fmars.2022.1019001.

Rincon, G., Weber, K. L., Van Eenennaam, A. L., Golden, B. L. & Medrano, J. F. (2011). Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys. *Journal of Dairy Science*, 94, 6116–6121. https://doi.org/10.3168/jds.2011-4764.

Rodríguez-Ramilo, S. T., Moran, P. & Caballero, A. (2006). Relaxation of selection with equalization of parental contributions in conservation programs: an experimental test with *Drosophila melanogaster*. *Genetics*, 172, 1043–1054. https://doi.org/10.1534/genetics.105.051003.

Saada, J.N., Kalantzis, G., Shyr, D., Cooper, F., Robinson, M., Gusev, A. & Palamara, P. F. (2020). Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. *Nature Communications*, 11, 6130. https://doi.org/10.1038/s41467-020-19588-x.

Sánchez, L., Bijma, P. & Woolliams, J. A. (2003). Minimizing inbreeding by managing genetic contributions across generations. *Genetics*, 164, 1589–1595. https://doi.org/10.1093/genetics/164.4.1589.

Saura, M., Pérez-Figueroa, A., Fernández, J., Toro, M. A. & Caballero, A. (2008). Preserving population allele frequencies in ex situ conservation programs. *Conservation Biology*, 22, 1277–1287. https://doi.org/10.1111/j.1523-1739.2008.00992.x.

Schaefer, R. J., Schubert, M., Bailey, E., Bannasch, D. L., Barrey, E., Bar-Gal, G. K. *et al*. (2017). Developing a 670k genotyping array to tag~ 2M SNPs across 24 horse breeds. *BMC Genomics*, 18, 565. https://doi.org/10.1186/s12864-017-3943-8.

Schoen, D. J., David, J. L. & Bataillon, T.M. (1998). Deleterious mutation accumulation

and the regeneration of genetic resources. *Proceedings of the National Academy of Sciences U.S.A.*, 95, 394–399.

Smith, J., Qiao, Y. & Williams, A. L. (2022). Evaluating the utility of identity-by-descent segment numbers for relatedness inference via information theory and classification. *G3: Genes, Genomes, Genetics*, 12, jkac072. https://doi.org/10.1093/g3journal/jkac072.

Solé, M., Gori, A. S., Faux, P., Bertrand, A., Farnir, F., Gautier, M. & Druet, T. (2017). Age-based partitioning of individual genomic inbreeding levels in Belgian Blue cattle. *Genetics Selection Evolution*, 49, 92. https://doi.org/10.1186/s12711-017-0370-x.

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. https://doi.org/10.1111/j.1365-294X.2007.03475.x.

Theodorou, K. & Couvet, D. (2003). Familial versus mass selection in small populations. *Genetics Selection Evolution*, 35, 425. https://doi.org/10.1051/gse:2003032.

Thompson, E. A. (2013). Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*, 194, 301–326. https://doi.org/10.1534/genetics.112.148825.

Toro, M.A., Barragán, C., Óvilo, C., Rodrigáñez, J., Rodríguez, C. & Silió, L. (2002). Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetics*, 3, 309–320. https://doi.org/10.1023/A:1019921131171.

Toro, M.A., Fernández, J. & Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science*, 120, 174–195. https://doi.org/10.1016/j.livsci.2008.07.003.

Toro, M.A. & Pérez-Enciso, M. (1990). Optimization of selection response under restricted inbreeding. *Genetics Selection Evolution*, 22, 93. https://doi.org/10.1186/1297-9686-22-1-93.

Toro, M.A., Villanueva, B. & Fernández, J. (2014). Genomics applied to management strategies in conservation programmes. *Livestock Science*, 166, 48–53. https://doi.org/10.1016/j.livsci.2014.04.020.

van Bers, N. E. M., Santure, A. W., Van Oers, K., Cauwer, I. S., Dibbits, B. W., Mateman *et al*. (2012). The design and crosspopulation application of a genome-wide SNP chip for the great tit *Parus major*. *Molecular Ecology Resources*, 12, 753–770. https://doi.org/10.1111/j.1755-0998.2012.03141.x.

van Son, M., Enger, E. G., Grove, H., Ros-Freixedes, R., Kent, M. P., Lien, S. & Grindflek, E. (2017). Genome-wide association study confirm major QTL for backfat fatty acid composition on SSC14 in Duroc pigs. *BMC Genomics*, 18, 369. https://doi.org/10.1186/s12864-017-3752-0.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*. 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

Vilas, A., Pérez-Figueroa, A., Quesada, H. & Caballero, A. (2015). Allelic diversity for neutral markers retains a higher adaptive potential for quantitative traits than expected heterozygosity. *Molecular Ecology*, 24, 4419–4432. https://doi.org/10.1111/mec.13334.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E. *et al*. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics Selection Evolution*, 53, 42. https://doi.org/10.1186/s12711-021-00635-0.

Villanueva, B., Pong-Wong, R., Woolliams, J.A. & Avendaño, S. (2004). Managing genetic resources in selected and conserved populations. In: Simm, G., Villanueva, B., Sinclair, K.D., Townsend, S. (eds.). *Farm Animal Genetic Resources*. BSAS Nottingham University Press, Nottingham, United Kingdom. 113–132.

Villanueva, B., Sawalha, R.M., Roughsedge, T., Rius-Vilarrasa, E. & Woolliams, J.A. (2010). Development of a genetic indicator of biodiversity for farm animals.

Livestock Science, 129, 200–207. https://doi.org/10.1016/j.livsci.2010.01.025.

Villanueva, B., Woolliams, J.A. & Simm, G. (1994). Strategies for controlling rates of inbreeding in MOET nucleus schemes for beef cattle. *Genetics Selection Evolution*, 26, 517. https://doi.org/10.1186/1297-9686-26-6-517.

Visscher, P. M., Woolliams, J. A., Smith, D. & Williams, J. L. (2002). Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *Journal of Dairy Science*, 85, 2368–2375. https://doi.org/10.3168/jds.S0022-0302(02)74317-8.

Wang, J. (1997). More efficient breeding systems for controlling inbreeding and effective size in animal populations. *Heredity*, 79, 591–599. https://doi.org/10.1038/hdy.1997.204.

Wang, J. (2004). Application of the one-migrant-per-generation rule in conservation and management. *Conservation Biology*, 18, 332–343. https://doi.org/10.1111/j.1523-1739.2004.00440.x.

Whitlock, M. C. (2011). G'_{ST} and D do not replace F_{ST}. *Molecular Ecology*, 20, 1083–1091. https://doi.org/10.1111/j.1365-294X.2010.04996.x.

Woolliams, J.A., Berg, P., Dagnachew, B.S. & Meuwissen, T.H.E. (2015). Genetic contributions and their optimisation. *Journal of Animal Breeding and Genetics*, 132, 89–99. https://doi.org/10.1111/jbg.12148.

Woolliams JA & Oldenbroek JK. (2017). Genetic diversity issues in animal populations in the genomic era. In: *Management of Animal Genetic Resources*, Wageningen Academic Publishers, Wageningen.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159. https://doi.org/10.1093/genetics/16.2.97.

Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y. *et al*. (2014). Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics*, 15, 307. https://doi.org/10.1186/1471-2164-15-307.

Yang, J., Benyamin, B., Mcevoy, B.P., Gordon, S., Henders, A.K., Nyholt, D. R. *et al.* (2010). Common SNPs explain a large proportion of heritability for human height. *Nature Genetics*, 42, 565–569. https://doi.org/10.1038/ng.608.

Yáñez, J. M., Naswa, S., Lopez, M. E., Bassini, L., Correa, K., Gilbey, J. *et al*. (2016). Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular Ecology Resources*. 16, 1002–1011. https://10.1111/1755-0998.12503.

Zhang, Q., Calus, M. P., Guldbrandtsen, B., Lund, M. S. & Sahana, G. (2015). Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genetics*, 16, 88. https://doi.org/10.1186/s12863-015-0227-7.

OBJECTIVES

General Objective

The general objective of the thesis was to evaluate the efficiency of different genomic coancestry matrices in the management of populations subject to conservation programs when the Optimal Contribution method is applied to maximize genetic diversity.

Specific Objectives

Chapter 1

1.1. Compare statistics (means, standard deviations, and correlations) for six different measures of genomic coancestry proposed in the literature in a farm turbot population.

1.2. Assess the genetic diversity (measured as expected heterozygosity) maintained when the six different genomic coancestry matrices are used in the Optimal Contribution method to manage the turbot population. Only genotype data from two consecutive generations (i.e., parents and offspring) were available to achieve this Objective.

Chapter 2

2.1. Evaluate, through computer simulations, the genetic diversity maintained when two different genomic coancestry matrices are used in the Optimal Contribution method for managing undivided populations across 50 generations (i.e., diversity is assessed in the short and long term). One of the matrices favors solutions that tend to move allele frequencies towards 0.5 (i.e., to increase genetic diversity), while the second matrix favors solutions that tend to keep allele frequencies closer to those in the original population. Genetic diversity was measured as expected heterozygosity and as allelic diversity.

2.2. Evaluate the allele frequency trajectories resulting from the use of both coancestry matrices across the 50 generations of management.

Chapter 3

3.1. Evaluate, through computer simulations, the genetic diversity maintained and the allele frequency trajectories when the two genomic coancestry matrices are used in the Optimal Contribution method for managing subdivided populations across 10 generations.

3.2. Determine the distribution of genetic diversity within and between subpopulations and the migratory flow between subpopulations in subdivided populations managed using the Optimal Contribution method.

CHAPTER 1

Evaluating different genomic coancestry matrices for managing genetic variability in turbot

Elisabet Morales-González[1]*, María Saura[1], Almudena Fernández[1], Jesús Fernández[1], Ricardo Pong-Wong[2], Santiago Cabaleiro[3], Paulino Martínez[4], Anaís Martín-García and Beatriz Villanueva[1]

[1]Departamento de Mejora Genética Animal, INIA, Ctra. de La Coruña, km 7.5, 28040 Madrid, Spain.

[2]Genetics and Genomics, The Roslin Institute and R(D)SVS of the University of Edinburgh, Midlothian EH25 9RG, Roslin, UK.

[3]CETGA, Cluster de Acuicultura de Galicia, Punta do Couso, s/n. 15695, Aguiño-Ribeira, A Coruña, Spain.

[4]Departamento de Xenética, Universidade de Santiago de Compostela, Campus de Lugo, 27002 Lugo, Spain.

*Corresponding author

The content of this chapter has been published in *Aquaculture*:

https://doi.org/10.1016/j.aquaculture.2020.734985.

Abstract

In population management, the most efficient method to control the increase of inbreeding and the associated loss of genetic variability is the Optimal Contributions method. This method optimizes the contributions of breeding candidates by minimizing the weighted global coancestry. Traditionally, coancestry coefficients have been estimated from pedigree data but the current availability of genome-wide information allows us to estimate them with higher precision. In recent years, developments of genomic tools in aquaculture species have been very significant. For turbot, a species with an increasing aquaculture value, the whole genome has been recently assembled and genetic and physical maps have been refined. Although several measures of genomic coancestry have been proposed, their relative efficiency for maintaining genetic variability is unknown. The objectives of this study were to compare different measures of genomic coancestry for turbot, and to evaluate their efficiency for retaining genetic variability when using the Optimal Contributions method. We used genomic data obtained through 2b-RAD technology for a domesticated population to achieve the objectives. The different genome-wide coancestry matrices compared were based on: i) the proportion of shared alleles; ii) deviations of the observed number of alleles shared by two individuals from the expected number; iii) the realized relationship matrix obtained by VanRaden's method 1; iv) the realized relationship matrix obtained by VanRaden's method 2; v) the realized relationship matrix obtained by Yang´s method; and vi) identical by descent segments. The amount of genetic variability retained when using each coancestry matrix was measured as the expected heterozygosity in the next generation. Results revealed that coancestry coefficients showing high correlations between them gave similar results from the optimization. The genetic variability retained was about 5% higher when using the matrices based on the proportion of shared alleles, deviations of the observed number of alleles shared or segments than when using the three genomic relationship matrices. Matrices retaining more variability showed a higher ability to discriminate relationships between individuals. The higher the diversity achieved the lower was the number of fish selected to contribute to the next generation.

Keywords: expected heterozygosity, coancestry, loss of variability, optimal

contributions, RAD-Seq, *Scophthalmus maximus*

1. Introduction

Two key objectives in the genetic management of populations are to maintain genetic variability and to avoid inbreeding depression (i.e., the reduction in mean phenotypic performance with increasing levels of inbreeding). This is particularly important in aquaculture breeding as the high fecundity of fish facilitates obtaining thousands of offspring from very few parents, increasing the risk of high inbreeding rates and low variability. In this sense, emphasis should be given to the optimal way of creating base populations from which selection programs start given that the genetic variability of the traits originally included in the breeding objective and those that will be included in the future will condition the success of the programs (Fernández *et al*., 2014).

The maintenance of genetic variability and the avoidance of inbreeding depression can be achieved by applying the Optimal Contributions (OC) method that optimizes contributions of breeding candidates to the next generation by minimizing the weighted global coancestry (Meuwissen, 1997; Grundy *et al*., 1998; Fernández *et al*., 2003; Villanueva *et al*., 2004; Woolliams *et al*., 2015). In terms of practical application of OC, aquaculture species have an advantage over terrestrial species as the high reproductive capacity of fish avoids the need of including additional constraints usually required when optimal contributions are higher than those biologically possible.

The central element of the OC method is the coancestry matrix (θ) that contains the coancestry coefficients (f) between all pairs of breeding candidates. These coefficients have been usually computed from pedigree data but with the availability of genotypes for large numbers of single nucleotide polymorphisms (SNPs) in recent years, pedigree-based estimates are being replaced with more accurate genomic estimates (Speed & Balding, 2015; Wang, 2016; Goudet *et al*., 2018; Supple & Shapiro, 2018). In fact, using genomic f simply measured as the proportion of alleles shared by two individuals (Nejati-Javaremi *et al*., 1997) in OC has been shown to lead to higher genetic diversity maintained than using pedigree-based coefficients (de Cara *et al*., 2011; Gómez-

Romano *et al*., 2013). Other measures of genomic *f* have been developed including those based on i) the deviations of the observed number of alleles shared by two individuals from the expected numbers under Hardy-Weinberg equilibrium (Li & Horvitz, 1953); and ii) identical by descent (IBD) segments, defined as continuous segments of DNA that are identical in two individuals (Gusev *et al*., 2009; de Cara *et al*., 2013; Gómez-Romano *et al*., 2016). In addition, genomic *f* can be obtained from the realized relationship matrices proposed by VanRaden (2008) and Yang *et al*. (2010) which are widely used in genome-wide association studies and in genomic selection.

Although still far from that in terrestrial animals, the development of genomic tools in aquaculture species has been very significant. For turbot (*Scophthalmus maximus*), a species with an increasing aquaculture value, important investments in recent years have been done and powerful tools are now available (Maroso *et al*., 2018). The whole genome has been recently assembled and genetic and physical maps have been refined. However, as for most fish species, commercial SNP arrays are still lacking. Instead, genotyping-by-sequencing technologies are receiving increasing attention in these species. These technologies are simple and cost-effective and include restriction-site associated DNA sequencing (RAD-Seq) technologies such as the original RAD-Seq, 2b-RAD and ddRAD (Li & Wang, 2017; Robledo *et al*., 2017; Sato *et al*., 2019, Zenger *et al*., 2019).

The objectives of this study were to compare different measures of genomic coancestry for turbot, and to determine the efficiency of different genomic coancestry matrices in retaining genetic variability measured as the expected heterozygosity when used in OC. Thousands of SNP genotypes obtained through 2b-RAD technology were used to achieve the objectives.

2. Material and Methods

2.1. Animal samples and genotypes

The turbot genome is small (~524 Mb) compared to other vertebrates. The latest

version of the genetic map (Maroso *et al*., 2018) includes 22 linkage groups (LGs) in accordance with the turbot karyotype constitution. Data available for this study came from an experiment developed at CETGA (Aquaculture Cluster of Galicia, Spain) within the framework of the European project FISHBOOST (http://www.fishboost.eu/). Genome-wide SNP data were available for 1,152 individuals (591 males and 561 females) with known sex belonging to 36 full-sib families (including 12 paternal and 11 maternal half-sib families) and for their parents (23 sires and 23 dams). Parents were sampled from the turbot CETGA population that represents a population of Atlantic origin (Maroso *et al*., 2018).

Genotypes were obtained by genotyping-by-sequencing using a 2b-RAD-sequencing approach. This method uses restriction enzymes that cut DNA at both sides of the recognition site at a fixed distance, producing short DNA fragments of identical size (33–36 bp). These fragments are subsequently sequenced on next-generation platforms (Wang *et al*., 2012; Robledo *et al*., 2017). In comparison with other RAD-based techniques, an advantage of 2b-RAD is that it facilitates the sampling and sequencing of identical sites across individuals. Details of the approach taken are given in Maroso *et al*. (2018). Briefly, after mapping to the turbot reference genome (Figueras *et al*., 2016) and applying quality filters, an initial set of 25,511 SNPs was obtained. Of these, only those present in 80% of the parents and with a minimum coverage of 10x were retained. This set of SNPs was used as a reference to obtain the SNPs in the offspring. Markers showing Mendelian errors (offspring genotype being inconsistent with Mendelian transmission, given the parental genotypes), unmapped SNPs and those with MAF < 0.015 in the parental population as well as those with extreme departures of Hardy-Weinberg equilibrium ($p < 0.001$) were removed. Also, for tags containing multiple polymorphisms only one SNP was retained. After quality control a total of 18,125 SNPs were retained. Then, the software BEAGLE 4.1 (Browning & Browning, 2011) was used to infer missing genotypes, and only those SNPs with high reliability (call rate > 90%) after imputation were kept. Imputation led to an increase of about 13% in the number of genotypes available across individuals. After quality control and imputation, a total of 18,097 SNPs were available for analysis.

2.2. Genomic coancestry coefficients

Six different genome-wide coancestry coefficients were considered and they are described below. They were all computed for the offspring generation (i.e., the breeding candidates).

1. f_{SIM}: SNP-by-SNP similarity between two individuals; i.e., the proportion of alleles shared by two individuals. Specifically, the coancestry coefficient between individuals *i* and *j* ($f_{SIM(i,j)}$) was computed as

$$f_{SIM\ (i,j)} = \frac{\sum_{k=1}^{S}\sum_{l=1}^{2}\sum_{m=1}^{2} I_{kl(i)m(j)}}{4S}$$

where *S* is the number of SNPs for which individuals *i* and *j* had genotype and $I_{kl(i)m(j)}$ is the identity of allele *l* of individual *i* with allele *m* of individual *j* for SNP *k* that takes the value of 1 if both alleles are identical and 0 otherwise (Nejati-Javaremi *et al*., 1997).

2. $f_{L\&H}$: Coancestry coefficient measured as the excess in the observed number of alleles shared by two individuals relative to the expected homozygosity under Hardy-Weinberg equilibrium. It was computed as

$$f_{L\&H(i,j)} = \frac{S f_{SIM(i,j)} - \sum_{k=1}^{S}\sum_{l=1}^{2} p_{kl}^2}{S - \sum_{k=1}^{S}\sum_{l=1}^{2} p_{kl}^2}$$

where p_{kl} is the allelic frequency of allele *l* of SNP *k* (Li & Horvitz, 1953; Toro *et al*., 2002).

3. f_{VR1}: Coancestry coefficient computed from the realized relationship matrix obtained by VanRaden's method 1 (VanRaden, 2008). The coancestry coefficient between individuals *i* and *j* was computed as

$$f_{VR1(i,j)} = \frac{\sum_{k=1}^{S}(x_{ki} - 2p_k)(x_{kj} - 2p_k)}{4\sum_{k=1}^{S} p_k(1-p_k)}$$

where x_{ki} is the genotype of individual *i* for SNP *k* that was coded as 0, 1 or 2 for genotypes AA, AB and BB, respectively and p_k is the frequency of the allele of SNP *k* whose homozygote genotype is coded as 2.

4. f_{VR2}: Coancestry coefficient computed from the realized relationship matrix obtained by VanRaden's method 2 (VanRaden, 2008). The coancestry coefficient between individuals *i* and *j* was computed as

$$f_{VR2(i,j)} = \frac{1}{2S}\sum_{k=1}^{S}\frac{(x_{ki} - 2p_k)(x_{kj} - 2p_k)}{2p_k(1-p_k)}$$

5. f_{YAN}: Coancestry coefficient computed from the realized relationship matrix of Yang *et al*. (2010). In this case, off-diagonal elements are computed as in VanRaden's second method, while diagonal elements are computed by considering that self-relationships are expected to be equal to 1 plus inbreeding:

$$f_{YAN(i,i)} = \frac{1}{2} + \frac{1}{2S}\sum_{k=1}^{S}\frac{x_{ki}^2 - (1+2p_k)x_{ki} + 2p_k^2}{2p_k(1-p_k)}$$

Note that coefficients f_{VR1}, f_{VR2} and f_{YAN} are based on the fact that the coancestry coefficient between individuals *i* and *j* equals $g_{(i,j)}/2$, where $g_{(i,j)}$ is the realized relationship between those individuals.

6. f_{SEG}: Coancestry coefficients based on IBD segments (de Cara *et al*., 2013). In particular, the coancestry between individuals *i* and *j* was computed as

$$f_{SEG(i,j)} = \frac{\sum_{k=1}^{S}\sum_{a_i=1}^{2}\sum_{b_j=1}^{2}(L_{seg_k}(a_i, b_j))}{4l}$$

where $L_{seg_k}(a_i, b_j)$ is the length of the shared IBD segment *k* measured over homologue *a* of individual *i* and homologue *b* of individual *j*, and *l* is the length of the genome covered by SNPs (i.e., the actual length minus the summed length of gaps longer than the maximum distance allowed between two consecutive SNPs in a segment). Estimation of f_{SEG} requires thus that phases of SNP genotypes are known, and they were obtained using the software BEAGLE 4.1 (Browning & Browning, 2011). The criteria used to define a segment were the following: i) the minimum length was set to 0.4 Mb; ii) the minimum density was set to 1 SNP every 50 kb; iii) the maximum distance allowed between two consecutive SNPs in a segment was 0.1 Mb; and iv) a maximum of 1 missing genotype was permitted in a segment. These criteria were based on the distribution of the distance between

consecutive SNPs and the density of SNPs observed across the genome (see later in the Results section). Note that runs of homozygosity (ROH), that refer to identity of DNA segments within a particular individual, can be obtained from self-coancestries.

Frequencies to be used in coefficients $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN} should be those in the base population. In our case, the oldest generation with frequencies available was the parental generation. Thus, the frequencies used were those of the parents of the breeding candidates.

2.3. Optimization of contributions

In order to evaluate the amount of genetic variability retained when using different coancestry matrices in the management of populations, the OC method (Fernández *et al*., 2003; Villanueva *et al*., 2004; Woolliams *et al*., 2015) was applied. The problem to be solved is concerned with the allocation of contributions of the candidates to produce the next generation so as to minimize the weighted global coancestry, and it can be formulated as

Minimize $\mathbf{c}^T\boldsymbol{\theta}\,\mathbf{c}$

subject to the following constraints:

$\mathbf{Q}^T\mathbf{c} \leq ½\,\mathbf{1}$

$c_i \geq 0$ for i = 1,…, *n* fish candidates

where **c** is the (*n* x 1) vector of solutions (i.e., contributions or proportions of offspring generated by each candidate), $\boldsymbol{\theta}$ is the coancestry matrix, **Q** is a (*n* x 2) known incidence matrix indicating the sex of the candidates with 0's and 1's, and **1** is a (2 x 1) vector of ones. The first inequality ensures that half of the contributions come from males and half come from females. The optimization problem was solved using Lagrangian multipliers (Meuwissen, 1997; Woolliams *et al*., 2015). Note that with this approach, **c** can contain negative values for some candidates. The contribution of candidates with $c_i < 0$ was then set to 0 and the optimization was repeated until all elements of **c** were non-negative.

Candidates considered in the optimization were the 1,152 offspring genotyped. The different coancestry matrices ($\boldsymbol{\theta}_{SIM}$, $\boldsymbol{\theta}_{L\&H}$, $\boldsymbol{\theta}_{VR1}$, $\boldsymbol{\theta}_{VR2}$, $\boldsymbol{\theta}_{YAN}$ and $\boldsymbol{\theta}_{SEG}$) were used in the optimization and their efficiency was compared in terms of the genetic variability retained. The amount of genetic variability retained was measured as the heterozygosity expected in the next generation (H_e) that equals to $1 - f_w$, where f_w is the average coancestry of the selected candidates (including self-coancestries) weighted by contributions. The average coancestry f_w was computed as $\mathbf{c_x}'\boldsymbol{\theta}_{SIM}\mathbf{c_x}$, where $\mathbf{c_x}$ is the vector of optimal contributions resulting of the optimization using matrix $\boldsymbol{\theta}_x$ (and $\boldsymbol{\theta}_x$ is $\boldsymbol{\theta}_{SIM}$, $\boldsymbol{\theta}_{L\&H}$, $\boldsymbol{\theta}_{VR1}$, $\boldsymbol{\theta}_{VR2}$, $\boldsymbol{\theta}_{YAN}$ or $\boldsymbol{\theta}_{SEG}$). Note that f_w is the expected inbreeding in the next generation under random mating. For comparison, the expected heterozygosity was also computed for the scenario where contributions were equalized (i.e., all candidates contribute and all do it equally).

3. Results

Detailed information of the SNPs located on the 22 LGs are given in Table 1. In general, the number of SNPs per LG increased with increasing LG length. It averaged 824 SNPs per LG and ranged from 576 (LG22) to 1,131 (LG02). The average SNP density for the whole genome was 34.90 SNPs/Mb and ranged from 28.42 (LG05) to 42.70 (LG17) SNPs/Mb. The average LG length was 24 Mb and the average distance between adjacent SNPs was 0.029 Mb and ranged from 0.025 (LG23) to 0.035 (LG07) Mb (Table 1). About 96% of adjacent SNPs were at distances ≤ 0.1 Mb (Figure 1). The distribution of MAF (Figure 2) indicated that almost 70% of the SNPs (12,466 SNPs) had a MAF ≥ 0.05 and about 8% (1,338 SNPs) had a MAF < 0.01.

Figure 3 shows the distribution of the length of ROH obtained from self-coancestries. The length of the ROH ranged from 0.40 to 5.55 Mb and a high proportion of short segments was observed. As described above, the minimum length chosen for defining a segment was 0.4 Mb. It is expected that ROH of this length come from common ancestors born 50 generations ago, when assuming a recombination rate of 2.5 cM/Mb (Bouza *et al*., 2007) and the fact that ROH length (in cM) equals $100/2g$ on

average, where *g* is the number of generations since the common ancestor (e.g., Howrigan *et al*., 2011). The restrictions on the minimum density required and the maximum distance allowed between two consecutive SNPs when defining a segment were imposed to avoid that sparsely covered genomic regions artificially inflate estimates of *f*. Note that the minimum average density per LG (Table 1) was 28.4 SNP/Mb (i.e., 0.0284 SNP/kb or 1 SNP every 35.1 kb) and that more than 95% of SNP pairs were at distances shorter than 0.1 Mb (Figure 1). The minimum number of SNPs in a segment was 9.

Estimates of the different measures of genomic *f* are given in Table 2. Estimates of f_{SIM} were much higher than estimates for the other coefficients, in particular those for $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN} that were close to 0. The estimate of f_{SEG} was also low but higher than that of $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN}. This is particularly true for global coancestry and for coancestry between individuals (off-diagonal elements). For sef-coancestries, the differences between the different measures were much smaller.

The correlations between the different coancestry measures estimated using all SNPs are shown in Figure 4. Correlations including self-coancestries and coancestries between individuals ranged from 0.72 to 1.00. Coefficients f_{SIM}, $f_{L\&H}$ and f_{SEG} showed pairwise correlations between them of nearly 1.00. Coefficients f_{VR1}, f_{VR2} and f_{YAN} showed also very high correlations between them ($\geq$ 0.96) and correlations between these coefficients ($f_{VR1,}$ f_{VR2} and f_{YAN}) and f_{SIM}, $f_{L\&H}$ and f_{SEG} were relatively lower but still high (> 0.70). However, corresponding correlations for self-coancestries (or inbreeding) evidenced more extreme results. Those between f_{VR1} or f_{VR2} with f_{SIM}, $f_{L\&H}$ or f_{SEG} were very low and even negative. In particular, self-coancestry coefficients from $\boldsymbol{\theta_{VR2}}$ were those presenting negative correlations (up to –0.40) with self-coancestry coefficients from $\boldsymbol{\theta_{SIM}}$, $\boldsymbol{\theta_{L\&H}}$ or $\boldsymbol{\theta_{SEG}}$. However, self-coancestry pairwise correlations between f_{YAN} and f_{SIM}, $f_{L\&H}$ and f_{SEG} were relatively high (> 0.70). The correlation of self-coancestry coefficients from $\boldsymbol{\theta_{VR2}}$ with those from $\boldsymbol{\theta_{YAN}}$ although positive, was relatively low (0.26).

Table 1. Number of SNPs, length (in Mb), SNP density (in SNP/Mb) and average distance between SNPs (in Mb) for each linkage group (LG).

LG[a]	No. SNPs	Length	Density	Distance
01	1,027	26.87	38.22	0.026
02	1,129	31.89	35.40	0.028
03	782	21.32	36.68	0.027
04	901	29.50	30.54	0.033
05	702	24.81	28.30	0.035
06	849	25.19	33.70	0.030
07	697	24.31	28.67	0.035
08	1,058	30.93	34.21	0.029
09	928	25.79	35.98	0.028
10	919	25.10	36.61	0.027
11	801	27.22	29.43	0.034
12	793	25.24	31.42	0.032
13	720	19.99	36.02	0.028
14	775	21.45	36.13	0.027
15	882	24.13	36.55	0.027
16	807	23.80	33.91	0.029
17	677	15.88	42.63	0.023
19	753	21.78	34.57	0.029
20	775	22.75	34.07	0.029
21	759	21.35	35.55	0.028
22	576	14.91	38.63	0.026
23	787	19.89	39.57	0.025
Total	18,097	524.10	–	–

[a]We follow the LG nomenclature of Maroso *et al.* (2018) where LG18 has been merged with LG8 (i.e., LG8 + LG18 is now LG8) when compared to the previous version of the map given by Figueras *et al.* (2016).

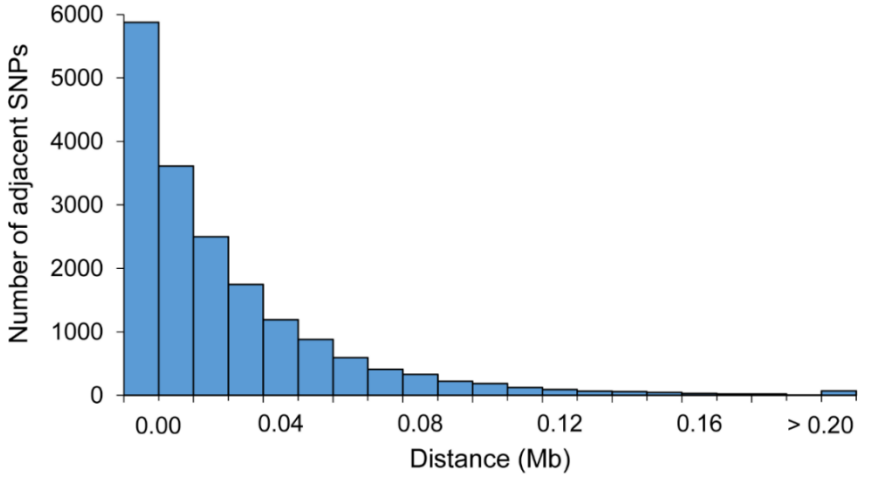

Figure 1. Distribution of the distance between adjacent SNPs (in Mb) in the genome.

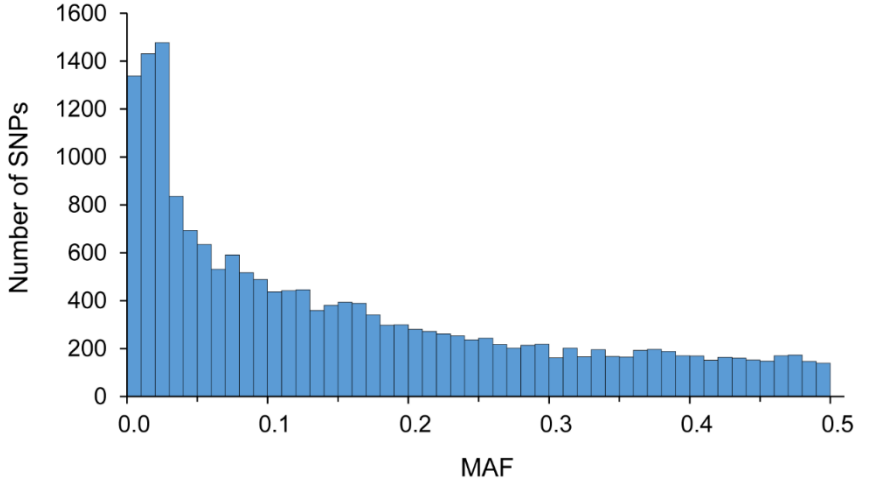

Figure 2. Distribution of the minimum allele frequency (MAF).

The correlations between the different coancestry measures estimated using all SNPs are shown in Figure 4. Correlations including self-coancestries and coancestries between individuals ranged from 0.72 to 1.00. Coefficients f_{SIM}, $f_{L\&H}$ and f_{SEG} showed pairwise correlations between them of nearly 1.00. Coefficients f_{VR1}, f_{VR2} and f_{YAN} showed also very high correlations between them (≥ 0.96) and correlations between these coefficients ($f_{VR1,}$ f_{VR2} and f_{YAN}) and f_{SIM}, $f_{L\&H}$ and f_{SEG} were relatively lower but still high

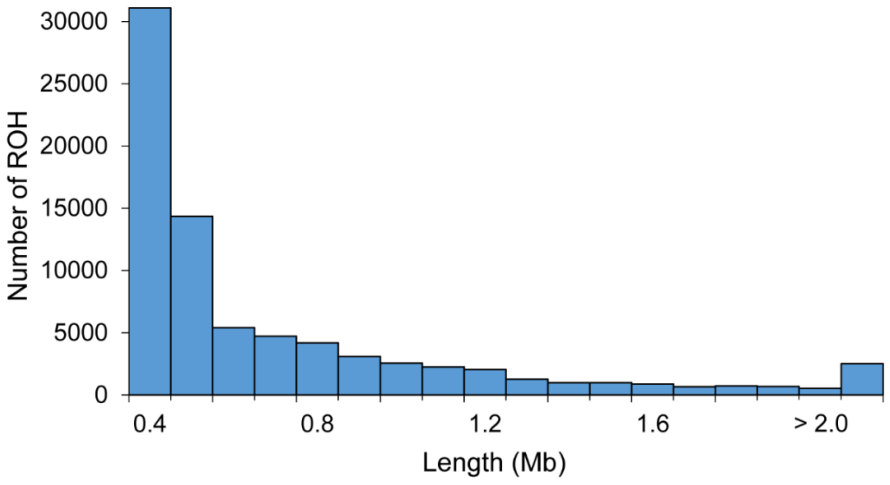

Figure 3. Distribution of the length of runs of homozygosity (ROH), in Mb.

(> 0.70). However, corresponding correlations for self-coancestries (or inbreeding) evidenced more extreme results. Those between f_{VR1} or f_{VR2} with f_{SIM}, $f_{L\&H}$ or f_{SEG} were very low and even negative. In particular, self-coancestry coefficients from $\boldsymbol{\theta_{VR2}}$ were those presenting negative correlations (up to –0.40) with self-coancestry coefficients from $\boldsymbol{\theta_{SIM}}$, $\boldsymbol{\theta_{L\&H}}$ or $\boldsymbol{\theta_{SEG}}$. However, self-coancestry pairwise correlations between f_{YAN} and f_{SIM}, $f_{L\&H}$ and f_{SEG} were relatively high (> 0.70). The correlation of self-coancestry coefficients from $\boldsymbol{\theta_{VR2}}$ with those from $\boldsymbol{\theta_{YAN}}$ although positive, was relatively low (0.26).

When the different measures of coancestry were used in OC, it was observed that the genetic variability retained, in terms of H_e, was about 5% higher when using $\boldsymbol{\theta_{SIM}}$, $\boldsymbol{\theta_{L\&H}}$ or $\boldsymbol{\theta_{SEG}}$ than when using $\boldsymbol{\theta_{VR1}}$, $\boldsymbol{\theta_{VR2}}$ or $\boldsymbol{\theta_{YAN}}$ (Table 3). In fact, the variability retained when using $\boldsymbol{\theta_{VR1}}$, $\boldsymbol{\theta_{VR2}}$ or $\boldsymbol{\theta_{YAN}}$ in OC was only slightly higher than that retained when equalizing contributions (H_e = 0.222). We also observed differences in the number of selected individuals when using the different matrices in the optimization. The higher heterozygosity retained with $\boldsymbol{\theta_{SIM}}$, $\boldsymbol{\theta_{L\&H}}$ and $\boldsymbol{\theta_{SEG}}$ was accompanied by an important decrease in the number of fish selected to contribute and by an increase in the variance of contributions, particularly with $\boldsymbol{\theta_{SIM}}$ and $\boldsymbol{\theta_{L\&H}}$. When using $\boldsymbol{\theta_{SIM}}$, $\boldsymbol{\theta_{L\&H}}$ or $\boldsymbol{\theta_{SEG}}$ only 8–13% of the candidates were selected in comparison with 49–56% when using $\boldsymbol{\theta_{VR1}}$, $\boldsymbol{\theta_{VR2}}$ or $\boldsymbol{\theta_{YAN}}$.

Table 2. Descriptive statistics of the different coefficients of coancestry.

	Mean	Standard deviation	Minimum	Maximum
Global coancestry				
f_{SIM}	0.777	0.018	0.740	0.927
$f_{L\&H}$	0.004	0.081	–0.157	0.673
f_{VR1}	0.002	0.062	–0.120	0.715
f_{VR2}	0.002	0.064	–0.089	0.974
f_{YAN}	0.002	0.063	–0.089	0.689
f_{SEG}	0.091	0.044	0.015	0.482
Self-coancestries (diagonal elements) (N = 1,152)				
f_{SIM}	0.887	0.012	0.861	0.927
$f_{L\&H}$	0.497	0.053	0.378	0.673
f_{VR1}	0.497	0.060	0.364	0.715
f_{VR2}	0.501	0.174	0.298	0.974
f_{YAN}	0.497	0.036	0.423	0.689
f_{SEG}	0.382	0.028	0.331	0.482
Coancestry between individuals (off-diagonal elements) (N = 662,976)				
f_{SIM}	0.777	0.018	0.741	0.888
$f_{L\&H}$	0.003	0.080	–0.157	0.498
f_{VR1}	0.002	0.061	–0.121	0.497
f_{VR2}	0.002	0.063	–0.089	0.637
f_{YAN}	0.002	0.063	–0.089	0.637
f_{SEG}	0.091	0.043	0.015	0.359

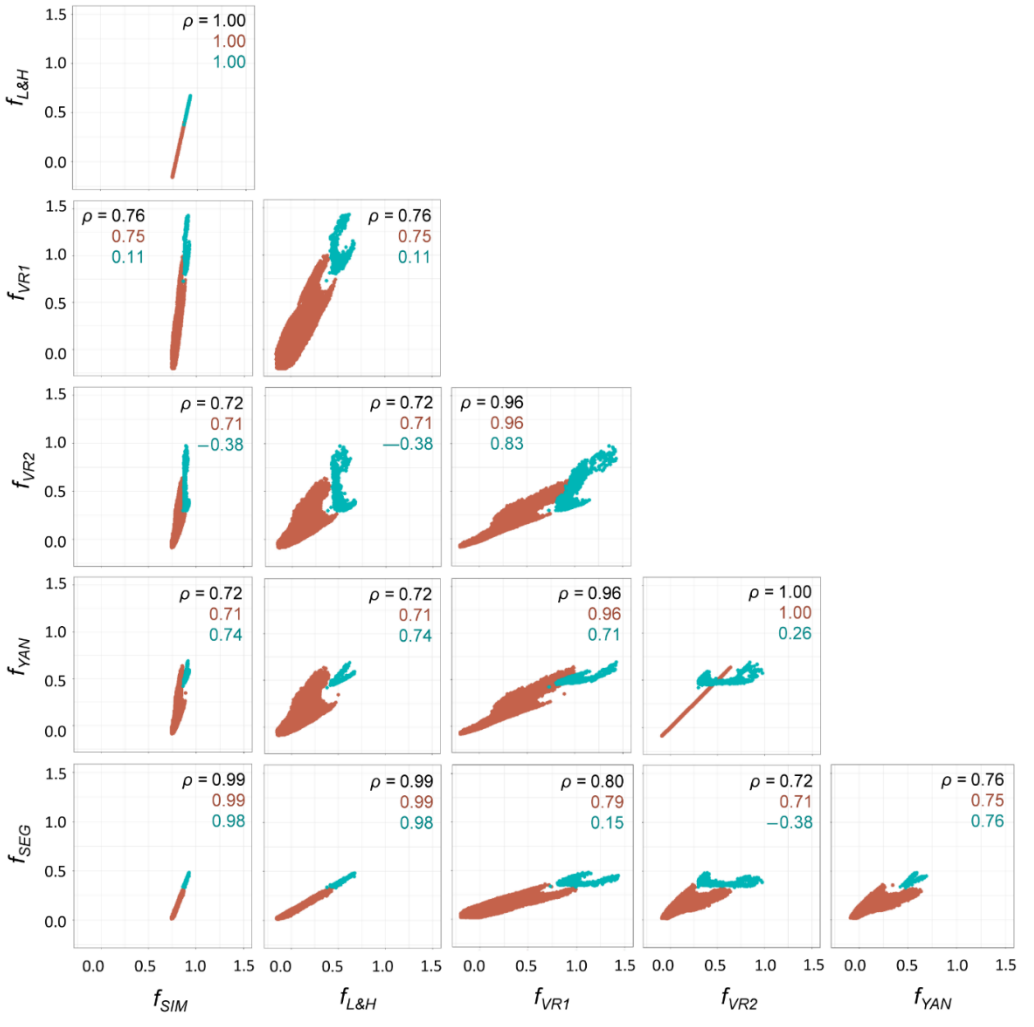

Figure 4. Scatter plots for coancestry coefficients f_{SIM}, $f_{L\&H}$, f_{VR1}, f_{VR2}, f_{YAN} and f_{SEG} against each other and corresponding correlation coefficients (ρ) when using all available SNPs. Self-coancestries are given in blue and coancestries between individuals are given in brown. Correlations in black include all data (self-coancestries and coancestries between individuals). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

In this study, we have made use of new genomic tools recently developed for turbot to compare different estimators of coancestry based on genomic information. Then, the different genomic coancestry matrices have been evaluated in terms of their

efficiency in retaining genetic variability measured as expected heterozygosity in the next generation, when implementing the OC method. The different matrices differed in the variability retained and particularly, in the number of fish needed to be kept to produce the next generation. Using matrices θ_{SIM}, $\theta_{L\&H}$ or θ_{SEG} in OC led to higher variability and lower number of fish selected to contribute than using θ_{VR1}, θ_{VR2} or θ_{YAN}.

The results obtained in this study are highly relevant in aquaculture breeding where the high fecundity typical from fish facilitates obtaining thousands of offspring from one single couple, increasing the risk of high inbreeding rates. This is particularly important when creating base populations from where breeding programs will start given that the genetic variability of traits potentially to be included in the breeding objective will condition the success of the programs. Fernández *et al*. (2014) clearly showed the benefits of using OC with genome-wide information when compared with strategies equalizing contributions. In their study, they used matrix θ_{SIM} but given the variety of genomic coancestry measures it was necessary to compare their relative efficiency for maintaining genetic variability when used in OC. Once the breeding program starts, the OC method should be applied as originally proposed; i.e., for obtaining the contributions of breeding candidates that maximize genetic gain obtained through selection while restricting at the same time the increase in coancestry and inbreeding (Meuwissen, 1997; Grundy *et al*., 1998; Woolliams *et al*., 2015). Although some studies have compared some genomic matrices in this context (e.g., Eynard *et al*., 2016), it is still unclear which is the most efficient matrix to obtain the highest genetic gain while restricting the rate of inbreeding.

Using SNP chip and whole-genome sequence cattle data, Eynard *et al*. (2016) compared the efficiency of using some of these genomic coancestry coefficients in OC for maintaining alleles segregating in the population. However, the most widely used measure of genetic variability is the expected heterozygosity (Nei, 1973), also called gene diversity, that represents the expected proportion of heterozygotes if the population were in Hardy-Weinberg equilibrium (Fernández & Bennewitz, 2017). The relevance of the expected heterozygosity is that it measures the ability of the population to respond to selection in the short term. Also, Fernández *et al*. (2004) showed that the strategies that

maximize expected heterozygosity through OC keep levels of allelic diversity as high as strategies that maximize allelic diversity itself.

The magnitude of the different coancestry measures compared varied greatly and that was mostly due to the time at which base populations were assumed for each coefficient. In particular, the magnitude of f_{SIM} was much higher than that of other coefficients. This was expected as f_{SIM} reflects, by definition, relationships coming from a common ancestor going back to a very distant base population in which all alleles were unique (Doekes *et al*., 2018). The coancestry coefficient based on the proportion of shared IBD segments (f_{SEG}) was lower than f_{SIM} but still higher than $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN}. The base population for f_{SEG} was assumed to be 50 generations ago (see above) whereas that for $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN} was the previous generation as these coefficients were computed using the allele frequencies in the parents. In the experiment that provided the data analyzed, all parents contributed equally and average values for $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN} were still close to 0 in the offspring.

The average ROH length obtained here for turbot was 0.77 Mb. This length is much smaller than the average length that is usually found in terrestrial species (Peripolli *et al*., 2016). To the best of our knowledge there is only one study describing ROH in fish, in particular in rainbow trout (D'Ambrosio *et al*., 2019) and it also reveals a larger average fragment size (~4 Mb). This can be explained by the differences in the size of the genome of the different species. While the genome size in rainbow trout is of the same order of magnitude than the size in terrestrial species (~2–3 Gb), the turbot genome length is about 3.5-fold smaller. Chromosome length ranges from 40 to 90 Mb in trout and from 15 to 30 Mb in turbot. Thus, long ROH segments are not expected to be found in the latter.

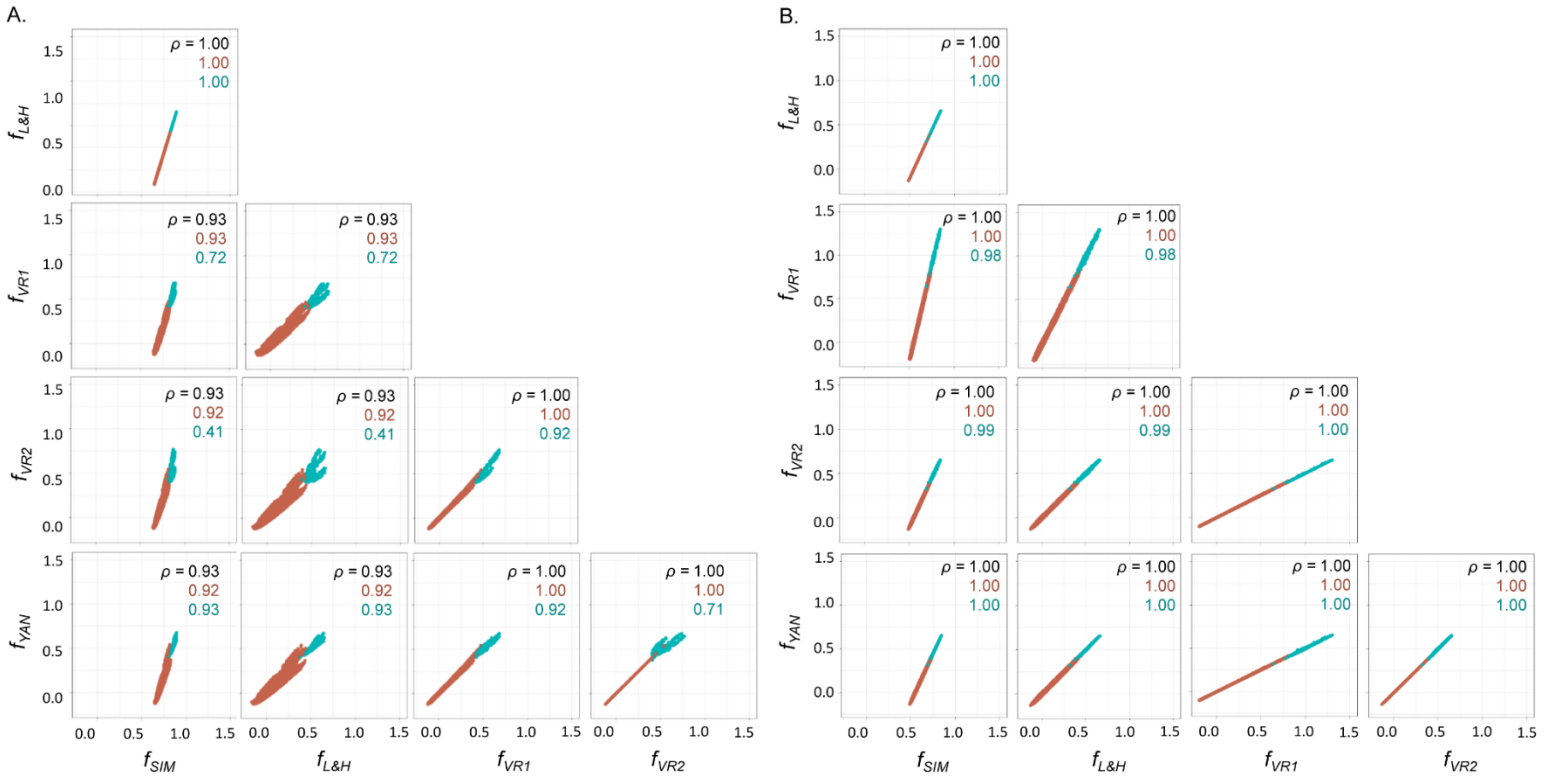

Figure 5. Scatter plots for coancestry coefficients f_{SIM}, $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN} against each other and corresponding correlation coefficients (ρ) when using only SNPs with MAF ≥ 0.05 (A) or MAF ≥ 0.25 (B). Self-coancestries are given in blue and coancestries between individuals are given in brown. Correlations in black include all data (self-coancestries and coancestries between individuals).

With RAD-sequencing technologies only a fraction of the genome is sequenced and genotyped. However, the SNP density in our study (~35 SNPs/Mb) was higher than that in D'Ambrosio *et al*. (2019) who used a 50K SNP chip for trout, and of the same order of magnitude than that found in studies on terrestrial species that used predesigned commercial SNP arrays (e.g., the Illumina PorcineSNP60 BeadChip). Thus, the technology used here represents an efficient and cost-effective genotyping option for measuring diversity in aquaculture species. Other useful applications of genotyping by sequencing in aquaculture species, for which genomic resources are typically more limited than in terrestrial species, includes performing genome-wide association studies and genomic selection for traits of interest in aquaculture (Robledo *et al*., 2017).

It is clear that, given their definitions, coefficients $f_{L\&H}$, f_{VR1}, f_{VR2} and f_{YAN} can be negative and higher that 1. In fact, $f_{L\&H}$ ranges from $-\infty$ to +1 and VanRaden´s coefficients range from −1 to $+\infty$. This enters in conflict with Malécot's definition of the coefficient of coancestry between two individuals that is defined as the probability that an allele drawn randomly from individual X is identical by descent to a gene drawn randomly from individual Y at an autosomal locus (Malécot, 1948) which must necessarily range from 0 to 1. Wang (2014) suggested to interpret the coancestry coefficient in terms of Wright's (1921) original correlation concept of relatedness rather than in terms of Malécot's probability of IBD. This allows negative values for coancestry coefficients. However, genomic coefficients are still out of the legitimate range [−1, 1] and thus, theoretical work is needed for an appropriate interpretation of these coefficients.

VanRaden's coefficients give higher weight to rare alleles and this may be the reason for the low positive and the negative correlations for the self-coancestries when these coefficients are involved. When correlations were re-estimated using only those SNPs with MAF ≥ 0.05 or MAF ≥ 0.25, there was a clear increase in the absolute value of those correlations that were initially positive and a change in the sign of those that were initially negative (Figure 5). This effect was more evident for the correlations involving f_{VR2} as this measure of coancestry is the measure giving more weight to rare alleles. All correlations (excluding those between self-coancestries) were > 0.9 when using SNPs with MAF ≥ 0.05. When using SNPs with MAF ≥ 0.25, all correlations

between the different coefficients were practically 1 (including correlations between self-coancestries). Note that this re-estimation filtering SNPs for MAF was not performed for f_{SEG} because it would lead to breaking off the segments and to an underestimation of f_{SEG}.

High correlations between coancestry measures did translate in similar results from the optimization. The use of θ_{SIM}, $\theta_{L\&H}$ and θ_{SEG} in OC led to the same variability retained (H_e = 0.24). This was expected given that pairwise correlations between f_{SIM}, $f_{L\&H}$ and f_{SEG} were practically 1. Similarly, the very high pairwise correlations between f_{VR1}, f_{VR2} and f_{YAN} ($\geq$ 0.96) led to the same H_e (0.22) when θ_{VR1}, θ_{VR2} and θ_{YAN} were used in OC. The benefit of using θ_{SIM}, $\theta_{L\&H}$ and θ_{SEG} over θ_{VR1}, θ_{VR2} and θ_{YAN} does not only lay in the higher genetic variability retained but also in the important decrease in the number of fish selected to contribute to the next generation. This result could be explained by a differential ability of the coefficients evaluated to discriminate relationships and agrees with that of Eynard *et al*. (2016) who when comparing θ_{SIM} and θ_{YAN} in cattle, found that with the latter all candidates were selected whereas with the former only 33% of the candidates were selected. The need of maintaining more individuals when using θ_{VR1}, θ_{VR2} or θ_{YAN} would increase the maintenance cost because more space would be required although this may be of less importance in fish than in terrestrial species.

The levels of H_e were relatively low (between 0.22 and 0.24), although similar to those found in other domesticated fish populations. For instance, Kijas *et al*. (2016) found that H_e was clearly lower in a farmed population of Atlantic salmon (H_e = 0.20) than in wild populations of the same species (H_e = 0.31). Estimates of H_e obtained from microsatellites for farmed populations of turbot (Coughlan *et al*., 1998; Bouza *et al*., 2002; Exadactylos *et al*., 2007), Atlantic salmon (Skaala *et al*., 2004) and carp (Ren *et al*., 2018) were also clearly lower than corresponding estimates for wild populations. These low levels of H_e are in accordance with the low estimates of effective population size (N_e) obtained for farmed fish populations. Saura *et al*. (2018) have recently given estimates of N_e for turbot (for the population studied here), gilthead seabream and carp of 28, 40 and 22 fish, respectively. Estimates lower than 50 (the critical value that is recommended to avoid inbreeding depression and retain fitness in the short-term) have

been also obtained for farmed gilthead seabream (Brown *et al*., 2005), coho salmon (Gallardo *et al*., 2004; Yáñez *et al*., 2014) and rainbow trout (Pante *et al*., 2001). These findings are in line with our results and highlight the necessity of broadening genetic diversity when base populations are built for starting breeding programs in aquaculture. With the increasing availability of genomic information in aquaculture species, base populations could be optimally designed using genomic estimates of relationships within and between candidate strains following the approach used here. Our results thus suggest that matrices θ_{SIM}, $\theta_{L\&H}$ and θ_{SEG} would be the most efficient to achieve the purpose of maximizing diversity in base populations.

5. Conclusions

The magnitude of the different estimates of coancestry measures in the population analyzed differed greatly. These differences can be explained by the differences in the time period where base populations were assumed for each coefficient. Correlations between the different coancestry and inbreeding (i.e., self-coancestries) coefficients were in general high, except for those involving self-coancestries computed using VanRaden and Yang´s realized relationship matrices. In particular, self-coancestries from θ_{VR2} showed low positive or even negative correlations with other coefficients. The genetic variability retained in the selected candidates in terms of expected heterozygosity was about 5% higher when using θ_{SIM}, $\theta_{L\&H}$ and θ_{SEG} than when using θ_{VR1}, θ_{VR2} or θ_{YAN}. The different matrices also led to different numbers of fish selected. The higher the diversity achieved the lower was the number of fish selected to contribute. Thus, matrices θ_{SIM}, $\theta_{L\&H}$ and θ_{SEG} show a higher ability to discriminate relationships between individuals.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Program (KBBE.2013.1.2-10) under grant agreement n°

613611, the Ministerio de Ciencia, Innovación y Universidades, Spain (grant CGL2016-75904-C2-2-P) and Fondos FEDER. R. Pong-Wong is funded by the Biotechnology and Biological Sciences Research Council through Institute Strategic Program Grant funding (BBS/E/D/30002275).

Ethics statement

This study was carried out in accordance with the recommendations of the ethical regulations and with the approvement of the Regional Government of Xunta de Galicia (registered under the code ES150730055401/16/PROD.VET.047ROD.01).

References

Bouza, C., Presa, P., Castro, J., Sánchez, L. & Martínez, P. (2002). Allozyme and microsatellite diversity in natural and domestic populations of turbot (*Scophthalmus maximus*) in comparison with other Pleuronectiformes. *Canadian Journal of Fisheries and Aquatic Sciences*, 59, 1460–1473. https://doi.org/10.1139/f02-114.

Bouza, C., Hermida, M., Pardo, B.G., Fernández, C., Fortes, G.G., Castro, J. *et al*. (2007). A microsatellite genetic map of the turbot (*Scophthalmus maximus*). *Genetics*, 177, 2457–2467. https://doi.org/10.1534/genetics.107.075416.

Brown, R.C., Woolliams, J.A. & McAndrew, B.J. (2005). Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture*, 247, 219–225. https://doi.org/10.1016/j.aquaculture.2005.02.002.

Browning, S. R. & Browning, B. L. (2011). Haplotype phasing: existing methods and new developments. *Nature Reviews Genetics*, 12, 703–714. https://doi.org/10.1038/nrg3054.

Coughlan, J.P., Imsland, A. K., Galvin, P.T., Fitzgerald, R.D., Naevdal, G. & Cross, T.F. (1998). Microsatellite DNA variation in wild populations and farmed strains of

turbot from Ireland and Norway: a preliminary study. *Journal of Fish Biology*, 52, 916–922. https://doi.org/10.1111/j.1095-8649.1998.tb00592.x.

D'ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C. *et al*. (2019). Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. *Genetics Selection Evolution*, 51, 26. https://doi.org/10.1186/s12711-019-0468-4.

de Cara, M.A.R., Fernández, J., Toro, M.A. & Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics*, 128, 456–464. https://doi.org/10.1111/j.1439-0388.2011.00971.x.

de Cara, M.A.R., Villanueva, B., Toro, M.A. & Fernández, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology*, 22, 6091–6099. https://doi.org/10.1111/mec.12560.

Doekes, H. P., Veerkamp, R. F., Bijma, P., Hiemstra, S. J. & Windig, J. (2018). Value of the Dutch Holstein Friesian germplasm collection to increase genetic variability and improve genetic merit. *Journal of Dairy Science*, 101, 10022–10033. https://doi.org/10.3168/jds.2018-15217.

Exadactylos, A., Rigby, M. J., Geffen, A. J. & Thorpe, J. P. (2007). Conservation aspects of natural populations and captive-bred stocks of turbot (*Scophthalmus maximus*) and Dover sole (*Solea solea*) using estimates of genetic diversity. *ICES Journal of Marine Science*, 64, 1173-1181. https://doi.org/10.1093/icesjms/fsm086.

Eynard, S. E., Windig, J. J., Hiemstra, S. J. & Calus, M. P. (2016). Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genetics Selection Evolution*, 48, 33. https://doi.org/10.1186/s12711-016-0210-4.

Fernández, J. & Bennewitz, J. (2017). Defining genetic diversity based on genomic tools. *Genomic Management of Animal Genetic Resources*, 1st ed.; Oldenbroek, JK, Ed, 49–76.

Fernández, J., Toro, M.A. & Caballero, A. (2003). Fixed contributions designs *vs*. minimization of global coancestry to control inbreeding in small populations. *Genetics*, 165, 885–894. https://doi.org/10.1093/genetics/165.2.885.

Fernández, J., Toro, M. A. & Caballero, A. (2004). Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conservation Biology*, 18, 1358–1367. https://doi.org/10.1111/j.1523-1739.2004.00341.x.

Fernández, J., Toro, M. A., Sonesson, A.K. & Villanueva, B. (2014). Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Frontiers in Genetics*, 5, 414. https://doi.org/10.3389/fgene.2014.00414.

Figueras, A., Robledo, D., Corvelo, A., Hermida, M., Pereiro, P., Rubiolo, J. A. *et al*. (2016). Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. *DNA Research*, 23, 181–192. https://doi.org/10.1093/dnares/dsw007.

Gallardo, J.A., García, X., Lhorente, J.P. & Neira, R. (2004). Inbreeding and inbreeding depression of female reproductive traits in two populations of coho salmon selected using BLUP predictors of breeding values. *Aquaculture*, 234, 111–122. https://doi.org/10.1016/j.aquaculture.2004.01.009.

Gómez-Romano, F., Villanueva, B., de Cara, M.A.R & Fernández, J. (2013). Maintaining genetic diversity using molecular coancestry: The effect of marker density and effective population size. *Genetics Selection Evolution*, 45, 38. https://doi.org/10.1186/1297-9686-45-38.

Gómez-Romano, F., Villanueva, B., Sölkner, J., de Cara, M. Á. R., Mészáros, G., Pérez O'Brien, A. M. & Fernández, J. (2016). The use of coancestry based on shared segments for maintaining genetic diversity. *Journal of Animal Breeding and Genetics*, 133, 357–365. https://doi.org/10.1111/jbg.12213.

Goudet, J., Kay, T. & Weir, B. S. (2018). How to estimate kinship. *Molecular Ecology*,

27, 4121–4135. https://doi.org/10.1111/mec.14833.

Grundy, B., Villanueva, B. & Wooliams, J.A. (1998). Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetics Research*, 72, 159–168. https://doi.org/10.1017/S0016672398003474.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L. *et al*. (2009). Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, 19, 318–326. https://doi.org/10.1101/gr.081398.108.

Howrigan, D. P., Simonson, M. A. & Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics*, 12, 460. https://doi: 10.1186/1471-2164-12-460.

Kijas, J. W., Hadfield, T., Naval Sanchez, M. & Cockett, N. (2016). Genome-wide association reveals the locus responsible for four-horned ruminant. *Animal Genetics*, 47, 258–262. https://doi.org/10.1111/age.12409.

Li, C.C. & Horvitz, D.G. (1953). Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics*, 5, 107–117.

Li, Y. H. & Wang, H. P. (2017). Advances of genotyping-by-sequencing in fisheries and aquaculture. *Reviews in Fish Biology and Fisheries*, 27, 535–559. https://doi.org/10.1007/s11160-017-9473-2.

Malécot, G., 1948. Les mathématiques de l'hérédité. Masson et Cie, Paris, France.

Maroso, F., Hermida, M., Millán, A., Blanco, A., Saura, M., Fernández, A. *et al*. (2018). Highly dense linkage maps from 31 full-sibling families of turbot (*Scophthalmus maximus*) provide insights into recombination patterns and chromosome rearrangements throughout a newly refined genome assembly. *DNA Research*, 25, 439–450. https://doi.org/10.1093/dnares/dsy015.

Meuwissen, T.H.E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75, 934–940. https://doi.org/10.2527/1997.754934x.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences U.S.A.*, 70, 3321–3323. https://doi.org/10.1073/pnas.70.12.3321.

Nejati-Javaremi, A., Smith, C. & Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science*, 75, 1738–1745. https://doi.org/10.2527/1997.7571738x.

Pante, M.J.R., Gjerde, B. & McMillan, I. (2001). Effect of inbreeding on body weight at harvest in rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*, 192, 201–211. https://doi.org/10.1016/S0044-8486(00)00467-1.

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R. & Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics*, 48, 255–271. https://doi.org/10.1111/age.12526.

Ren, W., Hu, L., Guo, L., Zhang, J., Tang, L., Zhang, E. *et al*. (2018). Preservation of the genetic diversity of a local common carp in the agricultural heritage rice–fish system. *Proceedings of the National Academy of Sciences U.S.A.*, 115, E546–E554. https://doi.org/10.1073/pnas.1709582115.

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P. & Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, 10, 670–682. https://doi.org/10.1111/raq.12193.

Sato, M., Hosoya, S., Yoshikawa, S., Ohki, S., Kobayashi, Y., Itou, T. & Kikuchi, K. (2019). A highly flexible and repeatable genotyping method for aquaculture studies based on target amplicon sequencing using next-generation sequencing technology. *Scientific Reports*, 9, 6904. https://doi.org/10.1038/s41598-019-43336-x.

Saura, M., Caballero, A., Santiago, E., Morales, E., Fernández, A. *et al*. (2018). The importance of ensuring genetic variability when establishing selection programmes in aquaculture. *XIX Reunión Nacional de Mejora Animal*, León, Spain, June 14–15.

Skaala, Ø., Høyheim, B., Glover, K. & Dahle, G. (2004). Microsatellite analysis in domesticated and wild Atlantic salmon (*Salmo salar L.*): allelic diversity and identification of individuals. *Aquaculture*, 240, 131–143. https://doi.org/10.1016/j.aquaculture.2004.07.009.

Speed, D. & Balding, D. J. (2015). Relatedness in the post-genomic era: is it still useful? *Nature Reviews Genetics*, 16, 33–44. https://doi.org/10.1038/nrg3821.

Supple, M. A. & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, 19, 131. https://doi.org/10.1186/s13059-018-1520-3.

Toro, M.A., Barragán, C., Óvilo, C., Rodrigáñez, J., Rodríguez, C. & Silió, L. (2002). Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetic*, 3, 309–320. https://doi.org/10.1023/A:1019921131171.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

Villanueva, B., Pong-Wong, R., Woolliams, J.A. & Avendaño, S. (2004). Managing genetic resources in selected and conserved populations. In: Simm, G., Villanueva, B., Sinclair, K.D., Townsend, S. (eds.). *Farm Animal Genetic Resources*. BSAS Nottingham University Press, Nottingham, United Kingdom. 113–132.

Wang, J. (2014). Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology*, 27, 518–530. https://doi.org/10.1111/jeb.12315.

Wang, J. (2016). Pedigrees or markers: which are better in estimating relatedness and inbreeding coefficient? *Theoretical Population Biology*, 107, 4–13. https://doi.org/10.1016/j.tpb.2015.08.006.

Wang, S., Meyer, E., McKay, J. K. & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, 9, 808–810. https://doi.org/10.1038/nmeth.2023.

Woolliams, J.A., Berg, P., Dagnachew, B.S. & Meuwissen, T.H.E. (2015). Genetic contributions and their optimisation. *Journal of Animal Breeding and Genetics*, 132, 89–99. https://doi.org/10.1111/jbg.12148.

Wright, S. (1921). Systems of mating. *Genetics*, 6, 111−178.

Yang, J., Benyamin, B., Mcevoy, B.P., Gordon, S., Henders, A.K., Nyholt, D. R. *et al.* (2010). Common SNPs explain a large proportion of heritability for human height. *Nature Genetics*, 42, 565−569. https://doi.org/10.1038/ng.608.

Yáñez, J. M., Naswa, S., Lopez, M. E., Bassini, L., Correa, K., Gilbey, J. *et al*. (2016). Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular Ecology Resources*, 16, 1002−1011. https://10.1111/1755-0998.12503.

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., Jerry, D. R. & Raadsma, H. W. (2019). Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Frontiers in Genetics*, 9, 693. https://doi.org/10.3389/fgene.2018.00693.

CHAPTER 2

Changes in allele frequencies when different genomic coancestry matrices are used for maintaining genetic diversity

Elisabet Morales-González [1*], Jesús Fernández [1], Ricardo Pong-Wong [2], Miguel A. Toro [3] and Beatriz Villanueva [1]

[1]Departamento de Mejora Genética Animal, INIA, Ctra. de La Coruña, km 7.5, 28040 Madrid, Spain.

[2]Genetics and Genomics, The Roslin Institute and R(D)SVS of the University of Edinburgh, Midlothian EH25 9RG, Roslin, UK.

[3]Departamento de Producción Agraria, ETSI Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, 28040 Madrid, Spain

*Corresponding author.

The content of this chapter has been published in *Genes*:

https://doi.org/10.3390/genes12050673.

Abstract

A main objective in conservation programs is to maintain genetic variability. This can be achieved using the Optimal Contributions (OC) method that optimize the contributions of candidates to the next generation by minimizing the global coancestry. However, it has been argued that maintaining allele frequencies is also important. Different genomic coancestry matrices can be used on OC and the choice of the matrix will have an impact not only on the genetic variability maintained, but also on the change in allele frequencies. The objective of this study was to evaluate, through stochastic simulations, the genetic variability maintained and the trajectory of allele frequencies when using two different genomic coancestry matrices in OC to minimize the loss of diversity: i) the matrix based on deviations of the observed number of alleles shared between two individuals from the expected numbers under Hardy-Weinberg equilibrium ($\boldsymbol{\theta_{LH}}$); and ii) the matrix based on VanRaden´s genomic relationship matrix ($\boldsymbol{\theta_{VR}}$). The results indicate that the use of $\boldsymbol{\theta_{LH}}$ resulted in a higher genetic variability than the use of $\boldsymbol{\theta_{VR}}$. However, the use of θ_{VR} maintained allele frequencies closer to those in the base population than the use of $\boldsymbol{\theta_{LH}}$.

Keywords: genetic diversity; allele frequencies; genomic coancestry matrix; optimal contributions.

1. Introduction

Genetic diversity is a prerequisite for populations to be able to face future environmental changes and to ensure long-term survival (Frankham *et al*., 2010). Thus, a common objective in genetic conservation programs is to minimize the loss of genetic variability. This can be achieved using the Optimal Contributions (OC) method that optimize the contributions of candidates to the next generation by minimizing the global coancestry (Meuwissen, 1997; Grundy *et al*., 1998; Fernández *et al*., 2003) It has been demonstrated that OC maximizes genetic diversity measured as expected heterozygosity (Caballero & Toro, 2000) which is proportional to the additive genetic variance of quantitative traits (Falconer & Mackay, 1996).

A different objective in genetic conservation programs can be to maintain allele frequencies to preserve the uniqueness of a particular population, since current frequencies are the result not only of genetic drift, but also of previous selection processes (Lacy, 2000; Frankham, 2008; Saura *et al*., 2008). Selection and drift can lead to a given allele responsible for a desirable trait at a high frequency. Moreover, trying to move the frequency to intermediate values to increase genetic variability would remove the uniqueness of the population. Thus, changes in the genetic composition of populations may be undesirable, particularly when dealing with ex situ conservation programs where the final aim is the reintroduction to the wild (Saura *et al*., 2008).

When the OC method is applied using pedigree information to compute coancestries, both objectives (maximum heterozygosity and maintenance of allele frequencies) are achieved (Saura *et al*., 2008) but this is not the case when coancestries are computed from molecular marker data. Previous studies have showed that using a coancestry matrix (θ) computed from large numbers of single nucleotide polymorphisms (SNPs) in OC is more efficient for maintaining diversity than using the pedigree-based coancestry matrix (de Cara *et al*., 2011, 2013; Gómez-Romano *et al*., 2013). However, given that the highest expected heterozygosity is obtained at intermediate allele frequencies, a consequence of applying OC using a θ based on SNP genotypes is that the genetic composition of the population is modified (Saura *et al*., 2008; de Cara *et al*., 2011, 2013a; Fernández *et al*., 2004; de Cara *et al*., 2013b).

Different genomic coancestry matrices have been proposed for being used in OC (de Cara *et al*., 2011, 2013a, Eynard *et al*., 2016; **Chapter 1**; Meuwissen *et al*., 2020). They include the matrix that describes deviations of the observed numbers of alleles shared by two individuals from the expected numbers under Hardy-Weinberg equilibrium (Li & Horvitz, 1953), and those obtained from genomic relationship matrices currently used in genomic predictions (**Chapter 1**; Meuwissen *et al*., 2020). In **Chapter 1** we showed that the expected heterozygosity retained through OC was higher when using the matrix proposed by Li & Horvitz (1953) than when using different genomic relationship matrices (i.e., the VaRaden's matrices based on method 1 and 2 (VanRaden, 2008) and the Yang's matrix (Yang *et al*., 2010)). However, as mentioned above, the

genomic θ used in OC will have an impact not only on the diversity maintained, but also on the trajectory of the change in allele frequencies. Gómez-Romano *et al*. (2016) suggested that while OC using a genomic coancestry matrix that simply measures the proportion of alleles shared by two individuals (Nejati-Javaremi *et al*., 1997) and that correlates perfectly with Li & Horvitz's matrix, favors solutions that tend to move allele frequencies towards 0.5, OC using VanRaden's matrices would lead to solutions that tend to keep allele frequencies closer to those in the original population (i.e., allele frequencies would tend to be unchanged). This has been recently confirmed by Meuwissen *et al*. (2020) in the context of OC aimed at maximizing genetic gain through selection while restricting the increase in inbreeding (i.e., restricting the loss of genetic diversity).

In general, populations under conservation programs are small and genetic drift leads to a loss of diversity and changes in allele frequencies. The magnitude of these drift effects depends on the effective population size (N_e) which can be estimated from genomic coancestry. However, Toro *et al*. (2020) have recently questioned the meaning of N_e when genomic matrices are used in OC. In particular, when optimal management is carried out using marker information, genetic diversity can increase in the initial generations implying negative estimates of N_e. Also, in the long term, N_e does not attain an asymptotic value, but it shows an unpredictable behavior. Their findings were based on OC using Nejati-Javaremi´s matrix (Nejati-Javaremi *et al.*, 1997) and it is unclear if they hold when other genomic coancestry matrices are used.

The objective of this study was to evaluate, through computer simulations, the genetic variability maintained and the trajectory of allele frequencies when different genomic coancestry matrices are used in OC. Estimates of N_e obtained from the change in heterozygosity computed from different genomic matrices were also compared.

2. Materials and Methods

Scenarios simulated involved the management of populations through the OC method using two different genomic coancestry matrices, for 50 discrete generations. Management started from a base population with family structure. The same base

population was used for the 100 replicates run and it was created in two steps. Firstly, a population at mutation-drift equilibrium was generated. Secondly, the population was expanded in order to have enough individuals for sampling the 100 replicates (see below).

2.1. Generation of the base population

A population at mutation-drift equilibrium was generated by simulating 10,000 discrete generations of random mating for a population of 100 individuals (50 males and 50 females). Sires and dams were sampled with replacement and were mated at random. Each mating produced 2 offspring (1 of each sex). Thus, N_e was equal to 100. The genome was composed of 20 chromosomes of 1 Morgan each. Two types of biallelic loci (SNP and unobserved loci) were simulated and they differed simply in their subsequent use. SNP loci were used for management after the base population was created whereas unobserved loci were used for measuring diversity and changes of allele frequencies, and for estimating N_e across generations. A total of 500,000 SNPs and 500,000 unobserved loci were simulated per chromosome. At the initial generation, all loci were fixed. The mutation rate per locus and generation was $\mu = 2.5\times10^{-6}$ for all loci. The number of new mutations per generation was sampled from a Poisson distribution with mean $2N_e n_c \mu n_l$, where n_c is the number of chromosomes (i.e., 20) and n_l is the total number of loci per chromosome (i.e., 1,000,000). Mutations were then randomly distributed across individuals, chromosomes and loci, switching allele 1 to allele 2 and vice versa. When generating the gametes, the number of crossovers per chromosome was drawn from a Poisson distribution with mean equal to 1. Crossovers were randomly distributed without interference. At the end of the process the expected heterozygosity measured at both types of loci had stabilized (mutation-drift equilibrium). After this, the population was expanded during 4 generations with the aim of having enough individuals to sample 100 different replicates. During these 4 generations, each individual was randomly allocated to 8 different mates and each mating produced 1 offspring. In this way, the number of individuals in the population was multiplied by 4 each generation. After these 4 generations, the population was composed by 25,600 individuals and constituted the base population ($t = 0$). There were a total of 56,017 SNPs and 55,840 unobserved loci still

segregating in $t = 0$. The expected heterozygosity (H_e) computed with all loci (SNPs and unobserved loci) still segregating was 0.1811 and the linkage disequilibrium (measured as r^2, the squared correlation between pairs of loci) between consecutive loci was 0.131.

2.2 Management strategies

Management was performed on populations of two different sizes ($N = 20$ and $N = 100$ individuals, half of each sex) using the OC method across 50 generations. Population size was kept constant across generations. The founder individuals for each replicate were randomly sampled from the base population. Note that, given that the set of individuals sampled in $t = 0$ differs across replicates, the number of segregating loci can also differ. In most scenarios (see below), all loci segregating in $t = 0$ were used for managing the population, for measuring diversity and changes of allele frequencies, and for estimating N_e.

The problem to be solved in the OC method is related to the allocation of contributions, i.e., the number of offspring each candidate should produce the next generation. The pursued strategy is to minimize the global coancestry weighted by those contributions; i.e. minimize $\mathbf{c}^T\boldsymbol{\theta}\,\mathbf{c}$, where $\mathbf{c}$ is a N x 1 vector of proportions of offspring left by each candidate (i.e., the vector of solutions) and $\boldsymbol{\theta}$ is the coancestry matrix. A restriction was imposed in the optimization such as the sum of the contributions of males and females is the same and equal to ½; i.e. $\mathbf{Q}^T\mathbf{c} = ½\ \mathbf{1}$, where $\mathbf{Q}$ is a (N x 2) known incidence matrix indicating the sex of the candidates with 0's and 1's, and $\mathbf{1}$ is a (2 x 1) vector of ones. The optimization problem was solved using Lagrangian multipliers (Meuwissen, 1997; Woolliams *et al*., 2015). Note that with this approach, $\mathbf{c}$ can contain negative values for some candidates. The contribution of candidates with $c_i < 0$ was then set to 0 and the optimization was repeated with the remaining candidates until all elements of $\mathbf{c}$ were non-negative. Finally, the contribution of individual i (c_i), which is a proportion, was converted to a number of offspring by multiplying c_i by $2N$ and rounding to the nearest integer but ensuring that the number of offspring of each sex equals to $N/2$. Each parent was randomly allocated to different mates (among the selected individuals) to produce its offspring.

Two management strategies were investigated and they differed in the genomic coancestry matrix used in the optimization of contributions. Under strategy S_{O_LH}, the coancestry matrix used was matrix $\boldsymbol{\theta_{LH}}$ which describes the excess in the observed number of alleles shared by two individuals relative to the expected number under Hardy-Weinberg equilibrium (Li & Horvitz, 1953; Toro *et al.*, 2002). Specifically, the coancestry coefficient between individuals *i* and *j* was computed as

$$f_{LH(i,j)} = \frac{\sum_{k=1}^{S} f_{OBS(i,j)k} - S + 2\sum_{k=1}^{S} p_k(1-p_k)}{2\sum_{k=1}^{S} p_k(1-p_k)},$$

where $f_{OBS(i,j)}$ is the proportion of alleles shared by individuals *i* and *j*, *S* is the number of SNPs and p_k is the frequency of the reference allele (allele *B*) of SNP *k* in *t* = 0. Under strategy S_{O_VR}, the coancestry matrix used was matrix θ_{VR} which is based on the genomic relationship matrix obtained from VanRaden's method 2 (VanRaden, 2008). Specifically, the coancestry coefficient between individuals *i* and *j* was computed as

$$f_{VR(i,j)} = \frac{1}{2S}\sum_{k=1}^{S} \frac{(x_{ki} - 2p_k)(x_{kj} - 2p_k)}{2p_k(1-p_k)},$$

where x_{ki} is the genotype of individual *i* for SNP *k*, coded as 0, 1 or 2 for genotypes *AA*, *AB* and *BB*, respectively, and p_k is as defined for f_{LH}.

In most scenarios, both coancestry matrices were computed every generation using all SNPs that were segregating in *t* = 0. However, we analyzed two additional scenarios where two different minor allele frequency (MAF) thresholds were imposed for the SNPs to be used to compute the coancestry matrices: i) using only SNPs with MAF > 0.05; and ii) using only SNPs with MAF > 0.25. The first threshold (MAF > 0.05) was considered because it is commonly applied when analyzing real data to reduce the number of potential genotyping errors. The second threshold (MAF > 0.25) was considered to explore the influence of rare alleles on the performance of the coancestry matrices investigated. It is known that with VanRaden's method rare alleles contribute more to the coancestry coefficient than common alleles (Gómez-Romano *et al.*, 2016; Forni *et al.*, 2011). It is, thus, interesting to determine how the differences between management strategies S_{O_LH} and S_{O_VR} vary in the different MAF scenarios.

Management in these additional scenarios was performed for 30 generations.

Furthermore, as a benchmark, we simulated a strategy (strategy S_E) where the contributions of all candidates were equalized (i.e., all individuals contributed with two offspring to the next generation). This is the simplest management strategy that has been proposed to maintain genetic diversity by increasing N_e. It should be noticed that when dealing with populations in which the relationships between individuals are homogeneous (all equally related) this strategy leads to a N_e close to $2N$.

2.3. Parameters evaluated

Management strategies were compared in terms of the genetic variability retained and the trajectory of the allele frequencies across generations for the SNPs and for the unobserved loci. Also, strategies were compared in terms of the number of individuals selected to produce the next generation (N_S) and the number of loci still segregating in a given generation, both for SNPs and for unobserved loci. The amount of genetic variability retained was measured as the expected heterozygosity (H_e) computed as $1 - \sum_{k=1}^{L} \sum_{l=1}^{2} p_{kl}^2$, where L is the number of loci (SNPs or unobserved loci) and p_{lk} is the frequency of allele l of locus k.

In order to evaluate the 'distance' between frequencies in a given generation t and frequencies in $t = 0$, we used the Kullback–Leibler (KL) divergence criterion, which measures how different is a particular distribution from a reference distribution (Kullback, 1997), which here is the distribution of allele frequencies in $t = 0$. The KL divergence between current frequencies and frequencies in $t = 0$ was computed as

$$KL = \sum_{k=1}^{L} \sum_{l=1}^{2} {p'}_{kl} log \frac{{p'}_{kl}}{p_{kl}},$$

where p_{kl} is the frequency of allele l of locus k in $t = 0$, and ${p'}_{kl}$ is the corresponding frequency in the current generation ($t > 0$). The summation over alleles included only alleles with ${p'}_{kl} > 0$.

Finally, N_e was estimated from the change in heterozygosity in SNP loci. Thus, N_e in generation t was computed as $N_e = 1/2\Delta H_e$, where ΔH_e equals $(H_{e(t-1)} - H_{e(t)})/H_{e(\tau-1)}$. All results presented are averages over the 100 replicates.

3. Results

3.1 Expected heterozygosity and Kullback–Leibler divergence for populations of size N = 100

For populations of size $N = 100$, strategy S_{O_LH} led to higher genetic variability (measured as H_e) than strategy S_{O_VR} (Table 1) and the difference between both strategies increased across generations. In particular, H_e was about 1%, 4% and 11% higher with S_{O_LH} than with S_{O_VR} in $t = 1$, 10 and 50, respectively. With S_{O_LH}, H_e even slightly increased in the initial generations while with S_{O_VR}, H_e decreased from the start. Also, H_e obtained with strategy S_{O_VR} was very similar to H_e obtained with strategy S_E. Table 1 also shows that S_{O_VR} maintained allele frequencies closer to those in the base population than S_{O_LH} given that the *KL* values for S_{O_LH} were ≥ 100% higher than for S_{O_VR}. The differences in *KL* between both strategies increased across generations. Also, at later generations S_{O_VR} was slightly more efficient in maintaining the initial frequencies than S_E, a strategy that is expected to maximize N_e and, thus, to minimize genetic drift.

Standard errors (computed across replicates) ranged from 4.91×10^{-5} to 9.54×10^{-5} for H_e and from 0.16×10^{-5} to 7.39×10^{-5} for *KL*.

The use of both matrices ($\boldsymbol{\theta_{LH}}$ and $\boldsymbol{\theta_{VR}}$) in OC also led to different numbers of individuals selected as parents of the next generation (N_S). In particular, with S_{O_LH} between 10% and 30% fewer individuals were selected than with S_{O_VR} (Table 1). In fact, with the latter almost all individuals were selected in all generations up to $t = 10$. The difference in N_S entailed a difference in the number of loci that remained segregating across generations that was much higher with S_{O_VR} than with S_{O_LH} (Table 1), particularly in the initial generations. As for H_e and for *KL*, strategies S_{O_VR} and S_E led to very similar values of N_S.

Table 1. Expected heterozygosity (H_e, in %) and Kullback Leibler divergence for unobserved loci (KL x 10^2), number of selected candidates (N_S), and number of SNPs (S) and unobserved loci (U) segregating across generations (t) when contributions are equalized (S_E) and when they are optimized using Li & Horvitz (S_{O_LH}) and VanRaden (S_{O_VR}) coancestry matrices computed with SNPs with MAF > 0.00 in a population of 100 individuals.

	S_E					S_{O_LH}*					S_{O_VR}*				
t	H_e	KL	N_S	S	U	H_e	KL	N_S	S	U	H_e	KL	N_S	S	U
1	19.17	0.06	100	51,035	50,894	+0.14	+0.14	–39	–2,239	–2,246	0.00	0.00	0	+8	+18
2	19.12	0.12	100	49,873	49,737	+0.21	+0.23	–36	–3,206	–3,229	0.00	0.00	0	–22	0
3	19.07	0.18	100	48,852	48,729	+0.28	+0.30	–35	–3,792	–3,847	0.00	0.00	0	–61	–52
4	19.03	0.24	100	47,946	47,828	+0.35	+0.37	–35	–4,182	–4,261	0.00	0.00	–1	–113	–101
5	18.98	0.30	100	47,108	47,003	+0.41	+0.43	–33	–4,384	–4,499	0.00	–0.01	–1	–162	–157
10	18.73	0.57	100	43,777	43,691	+0.68	+0.68	–30	–4,731	–4,975	0.00	–0.03	–2	–399	–401
15	18.51	0.82	100	41,311	41,217	+0.89	+0.86	–28	–4,523	–4,855	–0.01	–0.06	–5	–595	–587
20	18.27	1.06	100	39,313	39,229	+1.08	+0.99	–26	–4,152	–4,567	–0.01	–0.09	–6	–714	–720
30	17.82	1.50	100	36,231	36,140	+1.40	+1.16	–24	–3,329	–3,896	+0.01	–0.18	–9	–906	–899
40	17.38	1.90	100	33,854	33,759	+1.67	+1.24	–22	–2,517	–3,215	+0.03	–0.26	–11	–995	–970
50	16.95	2.28	100	31,940	31,848	+1.92	+1.27	–21	–1,786	–2,594	+0.05	–0.35	–12	–1,081	–1,036

*S_{O_LH} and S_{O_VR} values are those deviated from S_E.

Table 2. Average frequency of the minor allele in generation 0 (x 10^2) across generations (*t*) for SNPs and unobserved loci when contributions are equalized (S_E) and when they are optimized using Li & Horvitz (S_{O_LH}) and VanRaden (S_{O_VR}) coancestry matrices in a population of 100 individuals.

	SNPs			Unobserved loci		
t	S_E	S_{O_LH}	S_{O_VR}	S_E	S_{O_LH}	S_{O_VR}
0	13.45	13.45	13.45	13.39	13.39	13.39
1	13.44	13.68	13.45	13.39	13.60	13.40
2	13.44	13.81	13.45	13.39	13.72	13.40
3	13.44	13.94	13.45	13.38	13.82	13.39
4	13.44	14.06	13.44	13.38	13.93	13.39
5	13.44	14.17	13.44	13.38	14.02	13.39
10	13.44	14.67	13.41	13.38	14.44	13.36
15	13.45	15.08	13.37	13.39	14.77	13.33
20	13.44	15.42	13.32	13.39	15.05	13.29
30	13.44	15.96	13.23	13.39	15.46	13.23
40	13.45	16.36	13.12	13.39	15.75	13.15
50	13.45	16.67	13.01	13.40	15.98	13.07

Table 2 shows the evolution across generations of the average frequency of the minor allele in $t = 0$. This average frequency was practically constant with S_E and slightly decreased with S_{O_VR}. However, with S_{O_LH}, it increased from ~1% in $t = 1$ to 16–19% in $t = 50$. Thus, it is clear that S_{O_LH} leads average frequencies upward (ultimately towards 0.5) and S_{O_VR} tends to maintain them. As expected, these patterns were more evident for the SNPs than for the unobserved loci.

Figures 1 and 2 show the frequency (*f*) distribution also for minor alleles in $t = 0$ in this generation and after 10 and 30 generations of management, using different sets of SNPs to compute coancestries. When using all SNPs segregating in $t = 0$, the distributions for SNPs and unobserved loci were very similar (Figures 1a and 2a). However, when using only SNPs with MAF > 0.05 or MAF > 0.25, the distribution for SNPs was greatly

affected. When using SNPs with MAF > 0 or MAF > 0.05 (Figures 1a and 1b), a greater number of SNPs was fixed with S_{O_LH} than with S_{O_VR} across generations (see class f = 0.00). However, more loci (SNPs and unobserved loci) with low frequencies ($0.00 < f \leq 0.15$) were observed with S_{O_VR} than with S_{O_LH} and more loci with higher frequencies ($f > 0.4$) were observed with S_{O_LH} than with S_{O_VR}. Thus, although more alleles are fixed with S_{O_LH}, those that are kept segregating increase their frequency, while with S_{O_VR} the frequencies tend to be maintained. The highest difference between SNPs and unobserved loci was found when only SNPs with MAF > 0.25 were used to estimate the coancestry matrices (Figures 1c and 2c). These differences are due to the fact that no MAF filtering was done for the unobserved loci.

Figure 3 shows the trajectories of H_e and *KL* across generations for unobserved loci under strategies S_{O_LH} and S_{O_VR} using the three different sets of SNPs. The heterozygosity maintained with S_{O_LH} decreased as the MAF criterion chosen for the SNPs used to estimate coancestries become more restrictive given that the number of SNPs used decreased. In fact, the small increase in H_e observed in the initial generations when using all SNPs (MAF > 0.00) was not observed when using only the SNPs with MAF > 0.05 or MAF > 0.25. In parallel, the *KL* divergence with S_{O_LH} also decreased when increasing the severity of the restriction imposed on the SNPs used. However, with S_{O_VR} the changes observed in H_e and *KL* when using different set of SNPs were very small.

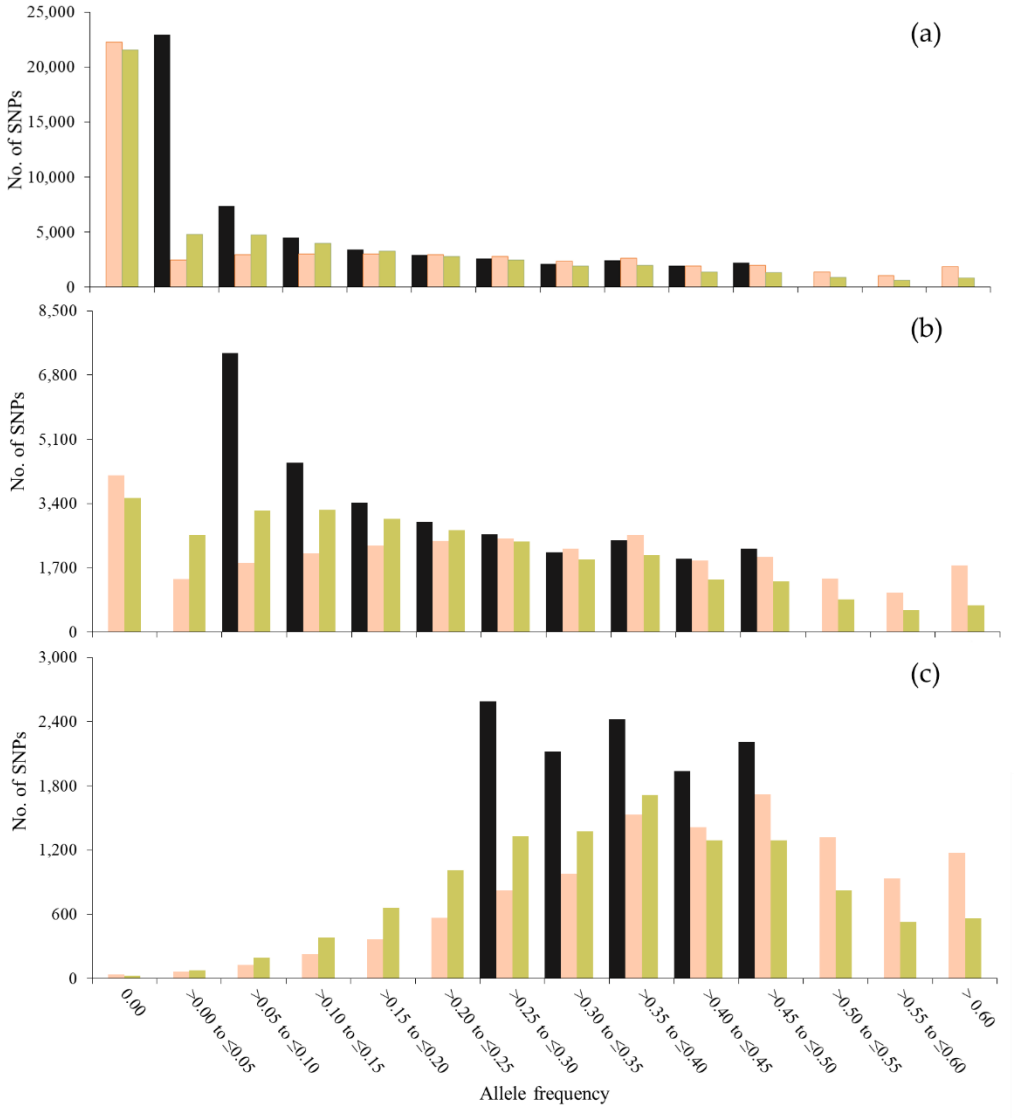

Figure 1. Number of SNPs for each class of allele frequency of the allele that was minor at generation 0 (gray bars) and the frequency of this allele after 10 (solid pattern) and 30 (stripe pattern) generations, when contributions are optimized using Li & Horvitz (S_{O_LH}, in orange) and VanRaden (S_{O_VR}, in green) coancestry matrices computed with SNPs with MAF > 0.00 (a), MAF > 0.05 (b) and MAF > 0.25 (c) in a population of 100 individuals.

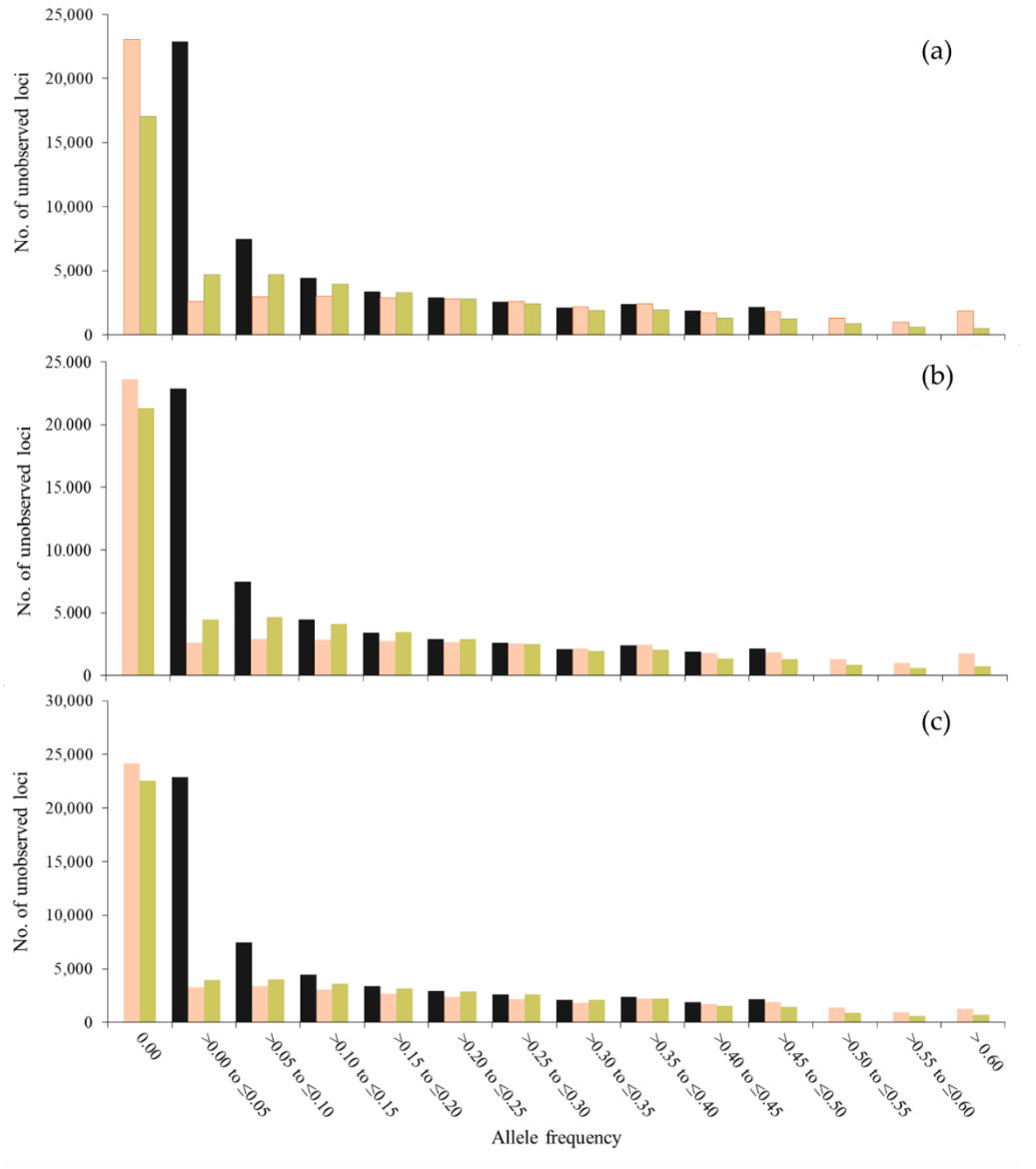

Figure 2. Number of unobserved loci for each class of allele frequency of the allele that was minor at generation 0 (gray bars) and the frequency of this allele after 10 (solid pattern) and 30 (stripe pattern) generations, when contributions are optimized using Li & Horvitz (S_{O_LH}, in orange) and VanRaden (S_{O_VR}, in green) coancestry matrices computed with SNPs with MAF > 0.00 (a), MAF > 0.05 (b) and MAF > 0.25 (c) in a population of 100 individuals.

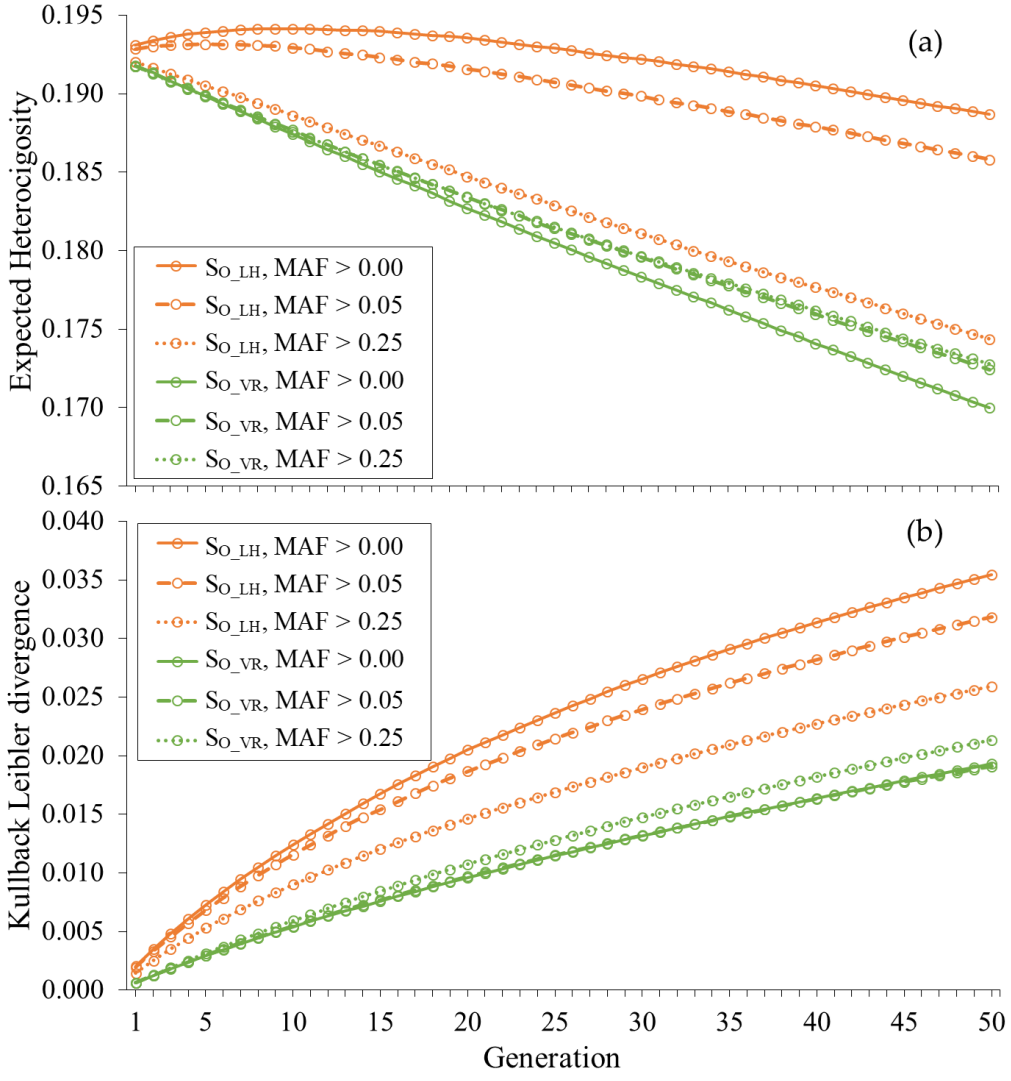

Figure 3. Expected heterozygosity (a) and Kullback Leibler divergence (b) for unobserved loci across generations when contributions are optimized using Li & Horvitz (S_{O_LH}) and VanRaden (S_{O_VR}) coancestry matrices computed with SNPs with MAF > 0.00, MAF > 0.05 and MAF > 0.25 in a population of 100 individuals.

3.2 Expected heterozygosity and Kullback–Leibler divergence for populations of size N = 20

Table 3 shows results from the different strategies (S_E, S_{O_LH} and S_{O_VR}) for populations of size N = 20. As it happened for populations of N = 100, i) S_{O_LH} led to higher H_e than S_{O_VR} and S_E; and ii) S_{O_VR} maintained allele frequencies closer to those in t = 0 than S_{O_LH}. However, differences among strategies were smaller for populations

of $N = 20$. For instance, for $N = 20$, H_e in $t = 10$ was less than 1% higher when managing with S_{O_LH} than when managing with S_{O_VR}, while for $N = 100$ this percentage was about 4%. For *KL*, the highest difference between strategies was 0.0027 units with $N = 20$ and 0.0127 units with $N = 100$. However, with $N = 20$, contrary to what happened with $N = 100$, S_{O_LH} managed to keep frequencies closer to the initial frequencies than S_E in the last generations ($t \geq 30$).

In populations of size $N = 20$, individuals are more closely related than in populations of size $N = 100$ and the genetic variability is smaller. Thus, most (if not all) individuals were selected to be parents of the next generation with all management strategies across generations. It should be noted that the number of loci segregating in $t = 0$, when management started, was substantially smaller when simulating populations of size $N = 20$. In order to investigate if the differences observed between $N = 20$ and $N = 100$ are a consequence of the different number of loci segregating in $t = 0$, a scenario with $N = 100$ starting with the same number of SNPs as in the scenario with $N = 20$ (about 40,000 SNPs) was simulated. The results indicated that the differences between scenarios with different N were due to the population size and not to the different number of loci (results not shown).

Standard errors (computed across replicates) ranged from 1.15×10^{-4} to 3.37×10^{-4} for H_e and from 10×10^{-4} to 1.72×10^{-4} for *KL*.

3.3 Effective population size

Table 4 shows estimates of N_e across generations for the different scenarios simulated. For $N = 100$, estimates of N_e were around 200 individuals under strategies S_E and S_{O_VR}. This is the expected value for N_e when contributions are equalized since N_e is approximately equal to $2N$. However, under strategy S_{O_LH}, estimates of N_e were unreasonable as they took negative values in the initial generations. In later generations, N_e became positive but did not reach a stable value. For $N = 20$, estimates under strategies S_E and S_{O_VR} were around 40 individuals, as expected. Estimates of N_e under strategy S_{O_LH}, were between 6% and 50% higher than under strategy S_E.

Table 3. Expected heterozygosity (H_e, in %) and Kullback Leibler divergence for unobserved loci (KL x 10^2), number of selected candidates (N_S), and number of SNPs (S) and unobserved loci (U) segregating across generations (t) when contributions are equalized (S_E) and when they are optimized using Li & Horvitz (S_{O_LH}) and VanRaden (S_{O_VR}) coancestry matrices computed with SNPs with MAF > 0.00 in a population of 20 individuals.

	S_E					S_{O_LH}*					S_{O_VR}*				
t	H_e	KL	N_S	S	U	H_e	KL	N_S	S	U	H_e	KL	N_S	S	U
1	23.35	0.27	20	38,995	38,955	+0.04	+0.05	−1	−193	−233	+0.03	0.00	0	+31	+134
2	23.06	0.52	20	37,093	37,050	+0.06	+0.07	−1	−275	−335	+0.01	0.00	0	+52	+155
3	22.76	0.76	20	35,522	35,472	+0.10	+0.09	−1	−356	−410	−0.02	+0.01	0	−12	+104
4	22.48	0.99	20	34,166	34,119	+0.07	+0.11	−1	−390	−442	−0.02	−0.01	0	−16	+94
5	22.19	1.20	20	33,016	32,978	+0.08	+0.13	−1	−456	−528	−0.03	0.00	0	−69	+37
10	20.79	2.17	20	28,782	28,692	+0.17	+0.18	−1	−533	−563	−0.07	−0.03	−1	−269	−62
15	19.52	3.00	20	25,844	25,763	+0.24	+0.17	−1	−497	−563	−0.03	−0.07	−1	−400	−206
20	18.33	3.75	20	23,512	23,434	+0.37	+0.13	−1	−336	−424	−0.01	−0.12	−1	−429	−247
30	16.02	5.13	20	19,854	19,795	+0.79	−0.02	−2	+81	−59	+0.04	−0.25	−2	−469	−337
40	14.03	6.26	20	17,044	17,002	+1.15	−0.16	−1	+545	+377	+0.18	−0.43	−2	−432	−309
50	12.32	7.23	20	14,853	14,811	+1.39	−0.27	−1	+787	+592	+0.19	−0.52	−2	−433	−322

* S_{O_LH} and S_{O_VR} values are those deviated from S_E.

Table 4. Effective population size (N_e) across generations (t) when contributions are equalized (S_E) and when they are optimized using Li & Horvitz (S_{O_LH}) and VanRaden (S_{O_VR}) coancestry matrices in populations of different size (N).

	$N = 100$			$N = 20$		
t	S_E	S_{O_LH}	S_{O_VR}	S_E	S_{O_LH}	S_{O_VR}
1	188.21	–111.90	195.55	36.92	42.27	40.40
5	199.07	–855.78	197.46	36.78	41.24	34.31
10	191.56	–5,777.32	193.05	38.54	40.81	41.77
15	203.50	1,855.71	194.54	36.65	45.41	43.18
20	202.62	1,033.03	201.52	40.61	47.25	40.02
25	190.44	636.00	209.85	40.20	47.08	42.02
30	193.58	670.07	209.79	36.45	53.03	38.57
35	193.30	524.97	206.03	33.41	50.28	44.62
40	204.95	601.67	212.53	36.94	47.91	49.68
45	207.44	703.31	205.00	37.52	48.50	40.09
50	206.86	481.08	213.02	41.99	46.20	38.53

4. Discussion

Using computer simulations, this study has compared two different management strategies in terms of two important criteria in genetic conservation programs; i.e., genetic diversity (H_e) maintained and changes in allele frequencies. Both strategies optimize contributions for maintaining diversity but differ in the genomic coancestry matrix used in the optimization ($\boldsymbol{\theta}_{\mathbf{LH}}$ in strategy S_{O_LH} and $\boldsymbol{\theta}_{\mathbf{VR}}$ in strategy S_{O_VR}). Also, as a benchmark, the simplest management strategy proposed to maintain genetic diversity that implies equalizing the contributions of all candidates (strategy S_E) was evaluated.

The evolution of changes in allele frequencies was evaluated using the *KL* divergence criterion. The greater the value of *KL*, the greater the divergence of frequencies with respect to the frequencies in the base population. When the strategies were compared using the *KL* criterion, it was clear that strategy S_{O_LH} gives higher values than strategy S_{O_VR} indicating that the latter is able to maintain allele frequencies closer to the original frequencies (lower *KL* values). On the other hand, with strategy S_{O_LH} the population evolves differently as it pushes frequencies towards 0.5 and thus changes the

genetic composition of the population more than strategy S_{O_VR}.

Pushing frequencies towards 0.5 as strategy S_{O_LH} does, leads to higher genetic variability when measured as expected heterozygosity. Thus, the hypothesis raised by Gómez-Romano *et al*. (2016) that using matrix $\boldsymbol{\theta_{LH}}$ in OC designed for maintaining genetic diversity achieves better the objective (i.e., higher H_e) than using matrix $\boldsymbol{\theta_{VR}}$ but using the latter maintains allele frequencies closer to the initial frequencies, is confirmed. This was observed both in populations with N = 20 and in populations with N = 100 although the differences between both strategies were smaller with N = 20. This is because individuals in the smaller populations are more closely related and there are less options to choose among individuals and strategies behave more similarly.

Strategy S_{O_VR} was only slightly more efficient for maintaining frequencies than strategy S_E. This strategy tends to reduce the change in allele frequencies, which implies a reduced genetic drift (Meuwissen *et al*., 2020). The magnitude of drift is minimized when N_e equals approximately 2N, and it is well known that, when managing the population using pedigree information, this is achieved by equalizing contributions (Falconer & Mackay, 1996; Fernández & Caballero, 2001). The small advantage of S_{O_VR} in terms of maintaining frequencies over S_E arises from the fact that the former uses realized relationships and detects real differences between individuals while S_E assumes homogeneous relationships. Contrarily, S_{O_LH} does not minimize drift but maximizes H_e by shifting frequencies towards 0.5. Thus, results from S_{O_LH} are quite different to those obtained under S_E in terms of the number of selected candidates and their optimal contributions.

Given that strategy S_{O_LH} brings the frequencies towards 0.5, H_e increased in the initial generations and this led to negative estimates of N_e in the largest population (N = 100). As generations go by, N_e becomes positive but with unrealistic very high values without attaining an asymptotic value. This was also observed by Toro *et al*. (2020) who questioned the meaning of N_e when genomic coancestry matrices are used in OC. They showed an unpredictable behavior for N_e when using the similarity genomic matrix of Nejati-Javaremi *et al*. (1997), which has a correlation of 1 with the $\boldsymbol{\theta_{LH}}$ matrix used here (Caballero & Toro, 2000; **Chapter 1**; Villanueva *et al*., 2021). However, our results

shown that when using $\boldsymbol{\theta_{VR}}$ in OC, estimates of N_e were close to the expected value when equalizing contributions (approximately 2N). As it has been discussed above, results from strategy S_{O_VR} were very similar to those from strategy S_E given that both tend to minimize drift. For the smallest population considered (N = 20) estimates of N_e were close to 2N not only with S_{O_VR} but also with S_{O_LH}. In such a small population, there is less options to choose among individuals and most of them are selected to contribute (Table 3). Thus, the three strategies investigated led to similar results.

Strategy S_{O_LH} led to higher H_e but also to a higher loss of segregating loci than strategy S_{O_VR}. In the largest population (N = 100), the percentage of alleles lost for unobserved loci at t = 1 was 13% and 9% with S_{O_LH} and S_{O_VR}, respectively (Table 1). The difference in both management strategies in terms of number of alleles lost could be due to the different numbers of individual selected to contribute to the next generation that was lower with S_{O_LH}. It must be emphasized that the mean coancestry of each individual with all the candidates (including the individual); i.e., the marginal of the coancestry matrix, is a useful concept for understanding the different numbers selected with both strategies. This is because the marginal of the coancestry matrix is a measure of the 'relevance' of each individual, in terms of the degree of genetic information shared with the rest and the optimal solutions will depend on all relationships between candidates. Its value is the same for all candidates when considering $\boldsymbol{\theta_{VR}}$. Then, all candidates are equally useful and should be selected as it was observed minimizing the global coancestry through OC using $\boldsymbol{\theta_{VR}}$ (strategy S_{O_VR}). However, when considering $\boldsymbol{\theta_{LH}}$, the average coancestry of individuals AA is lower than that of individuals BB, since individuals AA harbor genetic information that is underrepresented (i.e., they carry the rarer allele) and should be favored for selection and contributions. Therefore, OC using $\boldsymbol{\theta_{LH}}$ minimize the objective function when selecting the same number of AA and BB candidates. This leads to an increase in the frequency of allele A (actually to 0.5 in a single generation in this example with only one locus) while frequencies stay unchanged when using $\boldsymbol{\theta_{VR}}$.

Fernández *et al*. (2004) claimed that OC management using coancestry matrices based on allele sharing moves frequencies to intermediate values and reduces the

probability of losing alleles. In fact, these authors observed that strategies that maximize heterozygosity, by managing contributions from parents, keep levels of allelic diversity as high as strategies that maximize allelic diversity itself. Their results were obtained when applying OC using the similarity genomic matrix of Nejati-Javaremi *et al*. (1997), calculated with up to 40 multiallelic markers, but the same could be expected when using $\boldsymbol{\theta_{LH}}$ given that correlation between both matrices is 1. However, we have obtained solutions which maintain genetic diversity (H_e) but result in a higher number of fixed loci and this could be due to the different numbers of markers used in both studies.

To understand these contrasting results, we carried out extra simulations to compare observed with expected values for the number of fixed loci under both management strategies (i.e., S_{O_LH} and S_{O_VR}). In this extra scenario a population with $N = 20$ individuals was managed during 4 generations, with different numbers of SNPs used for the calculation of the coancestry matrices (20 and 1,000). A single chromosome was simulated. The expected number of fixed SNPs (ES_f) was estimated using the solutions that came out of each optimization before generating the offspring, following Fernández *et al*. (2004). Thus, ES_f was computed as $\sum_{k=1}^{2}\prod_{i=1}^{N} prob_{ki}$, where $prob_{ki}$ is the probability of individual i not transmitting allele k. If parent i carries a unique type of allele (that is, homozygous for the h allele) and leaves descendants, $prob_{ki}$ is 0 if $k = h$ and 1 if $k \neq h$. If it carries two different alleles (that is, heterozygous), the probability is $prob_{ki} = (0.5)^{c_i}$ where ci is the number of offspring to be contributed by parent i. ES_f value can be averaged then across loci. Table 5 shows that expected and observed numbers of SNPs becoming fixed each generation were close. When using only 20 SNPs, even though only 7-8 individuals are selected with S_{O_LH}, the expected (observed) number of SNPs that become fixed is lower than with S_{O_VR}. However, when the number of SNPs used was increased the trend reversed and the expected (and observed) number of fixed SNPs become lower for S_{O_VR} than for S_{O_LH} even when the number of selected individuals increases for S_{O_LH}. The explanation for this performance could be that with many markers, S_{O_LH} is able to find a solution with higher mean H_e by keeping loci with high MAF and allowing SNPs with rare alleles to become fixed.

Table 5. Number of selected candidates (N_S), and expected (ES_f) and observed number of fixed SNPs (S_f) across generations (t) when contributions are optimized using Li & Horvitz's (S_{O_LH}) and VanRaden's (S_{O_VR}) coancestry matrices computed with two different number of SNPs (S), for a population of 20 individuals.

		S_{O_LH}			S_{O_VR}		
t	S	N_S	ES_f	S_f	N_S	ES_f	S_f
1	20	7	0.3	0	20	0.3	0
2		7	0.7	0	13	0.8	1
3		8	0.8	0	13	1.4	1
4		8	0.9	0	12	1.7	1
1	1,000	15	21.7	21	20	17.6	18
2		16	38.9	37	19	34.6	33
3		15	54.6	52	19	50.9	47
4		15	68.6	64	18	66.3	60

Results have shown that the differences in maintained diversity (H_e) and divergence from the original frequencies (KL) between strategies S_{O_LH} and S_{O_VR} decreased when using only SNPs with a minimum MAF (MAF > 0.05 or MAF > 0.25) for computing the coancestry matrices. As mentioned above, S_{O_LH} promotes the contribution of individuals carrying rare alleles, as their coancestries with the rest of the population are smaller, and thus increases the frequencies of rare alleles. When the minimum MAF permitted increases, the number of rare alleles decreases, and the differences between the average coancestries between pairs of individuals decrease. In such situation, S_{O_LH} does not prioritize too much the contributions from any individual and leads to solutions that imply a higher number of candidates selected. Consequently, results are closer to those obtained with strategy S_{O_VR}. Moreover, when using only SNPs with high MAF in $t = 0$ (i.e., initial frequencies are close to 0.5), the performance of S_{O_VR} (i.e., keeping those initial frequencies) is similar to the performance of S_{O_LH} (moving them to intermediate values). These observations are in agreement with results from **Chapter 1** and Villanueva *et al*. (2021) who found that the correlation between VanRaden's and Li & Horvitz's coefficients increases with increasing the MAF of the

SNPs used.

Here, we have optimized contributions of parents for minimizing the loss of variability and then changes in frequencies have been evaluated. On the other hand, Saura *et al*. (2008) optimized contributions of parents for minimizing changes in allele frequencies and then the loss of genetic variability was evaluated. An alternative to both approaches could be to consider simultaneously the control of variability and the allele frequency changes. Similarly to the OC algorithm designed for maximizing genetic gain while restricting the rate of inbreeding (Meuwissen, 1997; Grundy *et al*., 1998; Woolliams *et al*., 2015) or for maximizing the phenotypic level for a trait of interest while restricting the loss in variability when creating base populations (Fernández *et al*., 2014) one could develop an algorithm for minimizing the loss of variability while restricting the change in frequencies or alternatively, for minimizing frequency changes while restricting the loss of variability. The specific objective would depend on the particular interest of the managers of the program. This kind of approach was followed by Fernández *et al*. (2006) in the context of optimizing the sampling strategy for establishing a gene bank. In particular, they developed an algorithm that simultaneously allows targeting frequencies for alleles at a particular locus while controlling the genetic diversity of other unlinked loci.

It could be also possible to combine both coancestry matrices ($\boldsymbol{\theta}_{\mathbf{LH}}$ and $\boldsymbol{\theta}_{\mathbf{VR}}$) in the objective function when the specific objective differs across genomic regions (i.e., in some regions the interest may be to maintain diversity and in other regions the interest may be to maintain frequencies). Maintaining diversity may be of interest for regions associated with inbreeding depression for fitness related traits and also for regions that harbor loci involved in general resistance to diseases (e.g., the major histocompatibility complex, MHC) as a high level of genetic diversity is desirable to ensure that the population can deal with potential new disease challenges (Gómez-Romano *et al*., 2016). Maintaining frequencies may be of interest in regions containing loci that have been under natural or artificial selection, and one wants to keep the genetic progress obtained. Gómez-Romano *et al*. (2016) showed that the OC method using a matrix equivalent to $\boldsymbol{\theta}_{\mathbf{LH}}$ is efficient in maintaining H_e in specific regions and simultaneously restrict the loss

of H_e in the rest of the genome. Their approach could be extended to include the use of θ_{VR} for minimizing the change in allele frequencies in some genomic regions.

The amount of genetic variability retained was measured as the expected heterozygosity (H_e). However, other measures such as allelic diversity can be used (Fernández *et al*., 2004; Caballero & Rodríguez-Ramilo, 2010). Allelic diversity is essential from an evolutionary perspective, since the limit of selection response is determined by the initial number of alleles (James, 1970; Hill & Rasbash, 1986). It is worth to note that strategy S_{O_VR} would be more efficient than strategy S_{O_LH}, not only to maintain allele frequency but also to maintain diversity when this is measured as the number of unobserved loci segregating. It is thus clear that the coancestry matrix to be used in OC when managing a particular genetic conservation programs would be case specific.

5. Conclusions

When applying strategy S_{O_LH}, more H_e is maintained than when applying strategy S_{O_VR} given that S_{O_LH} moves allele frequencies towards 0.5. However, S_{O_VR} maintained allele frequencies closer to those of the initial generation and more loci segregating than S_{O_LH}. Therefore, considering that conservation programs generally aim to increase genetic diversity, but it is also important to maintain population uniqueness, the choice of which genomic coancestry matrix is used in management may depend on which of these two goals is more important for each particular case. When a subset of SNPs with MAF > 0.05 or MAF > 0.25 is used to estimate coancestry matrices, the differences between both strategies in terms of both H_e and *KL* was reduced. The differences between strategies were smaller for populations of smaller size given that in a smaller population it is more difficult to differentiate between individuals.

Errata

In the published article corresponding to **Chapter 2**, there was an error in **Figure**

2. This was fixed in this document and, therefore, the corrected figure appears in page 90.

Author Contributions

Conceptualization, J.F., R.P.W. and B.V.; methodology, E.M.G, J.F and B.V..; software, E.M.G. and J.F; formal analysis, E.M.G.; writing—original draft preparation, E.M.G, J.F and B.V.; writing—review and editing, M.A.T and R.P.W.; supervision, B.V.; project administration, J.F. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The research leading to these results has received funding from the Ministerio de Ciencia, Innovación y Universidades, Spain (grant CGL2016-75904-C2-2-P). R. Pong-Wong is funded by the European Union's Horizon 2020 research and innovation program under the Grant Agreement n°772787 (SMARTER) and the Biotechnology and Biological Sciences Research Council through Institute Strategic Program Grant funding (BBS/E/D/30002275).

References

Caballero, A. & Rodríguez-Ramilo, S. T. (2010). A new method for the partition of allelic diversity within and between subpopulations. *Conservation Genetics*, 11, 2219–2229. https://doi.org/10.1007/s10592-010-0107-7.

Caballero, A. & Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3, 289–299. https://doi.org/10.1023/A:1019956205473.

de Cara, M.A.R., Fernández, J., Toro, M.A. & Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes.

Journal of Animal Breeding and Genetics, 128, 456–464. https://doi.org/10.1111/j.1439-0388.2011.00971.x.

de Cara, M.A.R., Villanueva, B., Toro, M.A. & Fernández, J. (2013a). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology*, 22, 6091–6099. https://doi.org/10.1111/mec.12560.

de Cara, M.A.R., Villanueva, B., Toro, M.A. & Fernández, J. (2013b). Purging deleterious mutations in conservation programs: Combining optimal contributions with inbred mattings. *Heredity*, 110, 530–537. https://doi.org/10.1038/hdy.2012.119.

Eynard, S. E., Windig, J. J., Hiemstra, S. J. & Calus, M. P. (2016). Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genetics Selection Evolution*, 48, 33. https://doi.org/10.1186/s12711-016-0210-4.

Falconer, D.S. & Mackay, F.C. (1996). Introduction to Quantitative Genetics. 4th ed. Longman Group Ltd, Harlow, Essex, England.

Fernández, J. & Caballero, A. (2001). Accumulation of deleterious mutations and equalization of parental contributions in the conservation of genetic resources. *Heredity*, 86, 480–488. https://doi.org/10.1046/j.1365-2540.2001.00851.x.

Fernández, J., Roughsedge, T., Woolliams, J. A. & Villanueva, B. (2006). Optimization of the sampling strategy for establishing a gene bank: storing PrP alleles following a scrapie eradication plan as a case study. *Animal Science*, 82, 813–821. https://doi.org/10.1017/ASC2006101.

Fernández, J., Toro, M.A. & Caballero, A. (2003). Fixed contributions designs *vs*. minimization of global coancestry to control inbreeding in small populations. *Genetics*, 165, 885–894. https://doi.org/10.1093/genetics/165.2.885.

Fernández, J., Toro, M. A. & Caballero, A. (2004). Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conservation Biology*, 18, 1358–1367. https://doi.org/10.1111/j.1523-1739.2004.00341.x.

Fernández, J., Toro, M. A., Sonesson, A.K. & Villanueva, B. (2014). Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Frontiers in Genetics*, 5, 414. https://doi.org/10.3389/fgene.2014.00414.

Forni, S., Aguilar, I. & Misztal, I. (2011). Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution*, 43, 1. https://doi.org/10.1186/1297-9686-43-1.

Frankham, R. (2008). Genetic adaptation to captivity in species conservation programs. *Molecular Ecology*, 17, 325–333. https://doi.org/10.1111/j.1365-294X.2007.03399.x.

Frankham, R., Ballou, J. D. & Briscoe D. A. (2010). *Introduction to conservation genetics*. 2nd ed. Cambridge: Cambridge University Press, United Kingdom.

Gómez-Romano, F., Villanueva, B., de Cara, M.A.R & Fernández, J. (2013). Maintaining genetic diversity using molecular coancestry: The effect of marker density and effective population size. *Genetics Selection Evolution*, 45, 38. https://doi.org/10.1186/1297-9686-45-38.

Gómez-Romano, F.; Villanueva, B.; Fernández, J.; Woolliams, J.A.; & Pong-Wong, R. (2016). The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution*, 48, 2. https://doi.org/10.1186/s12711-015-0172-y.

Grundy, B., Villanueva, B. & Wooliams, J.A. (1998). Dynamic selection procedures for constrained inbreeding and their consequences for pedigree development. *Genetics Research*, 72, 159–168. https://doi.org/10.1017/S0016672398003474.

Hill, W. G. & Rasbash, J. (1986). Models of long term artificial selection in finite population. *Genetics Research*, 48, 41–50. https://doi.org/10.1017/S0016672300024642.

James, J. W. (1970). The founder effect and response to artificial selection. *Genetics Research*, 16, 241–250.

Kullback, S. (1997). *Information theory and statistics*; Courier Dover Publications. Dover, New York.

Lacy, R. C. (2000). Should we select genetic alleles in our conservation breeding programs? *Zoo Biology*, 19, 279–282. https://doi.org/10.1002/1098-2361(2000)19:4<279::AID-ZOO5>3.0.CO;2-V.

Li, C.C. & Horvitz, D.G. (1953). Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics*, 5, 107–117.

Meuwissen, T.H.E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75, 934–940. https://doi.org/10.2527/1997.754934x.

Meuwissen, T. H. E., Sonesson, A. K., Gebregiwergis, G. & Woolliams, J. A. (2020). Management of genetic diversity in the era of genomics. *Frontiers in Genetics*, 11, 880. https://doi.org/10.3389/fgene.2020.00880.

Nejati-Javaremi, A., Smith, C. & Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science*, 75, 1738–1745. https://doi.org/10.2527/1997.7571738x.

Saura, M., Pérez-Figueroa, A., Fernández, J., Toro, M. A. & Caballero, A. (2008). Preserving population allele frequencies in ex situ conservation programs. *Conservation Biology*, 22, 1277–1287. https://doi.org/10.1111/j.1523-1739.2008.00992.x.

Toro, M.A., Barragán, C., Óvilo, C., Rodrigáñez, J., Rodríguez, C. & Silió, L. (2002). Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetic*, 3, 309–320. https://doi.org/10.1023/A:1019921131171.

Toro, M. A., Villanueva, B. & Fernández, J. (2020). The concept of effective population size loses its meaning in the context of optimal management of diversity using molecular markers. *Journal of Animal Breeding and Genetics*, 137, 345–355. https://doi.org/10.1111/jbg.12455.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of*

Dairy Science, 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E. *et al*. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics Selection Evolution*, 53, 42. https://doi.org/10.1186/s12711-021-00635-0.

Woolliams, J.A., Berg, P., Dagnachew, B.S. & Meuwissen, T.H.E. (2015). Genetic contributions and their optimisation. *Journal of Animal Breeding and Genetics*, 132, 89–99. https://doi.org/10.1111/jbg.12148.

Yang, J., Benyamin, B., Mcevoy, B.P., Gordon, S., Henders, A.K., Nyholt, D. R. *et al.* (2010). Common SNPs explain a large proportion of heritability for human height. *Nature Genetics*, 42, 565–569. https://doi.org/10.1038/ng.608.

CHAPTER 3

Maintenance of genetic diversity in subdivided populations using genomic coancestry matrices

Elisabet Morales-González[1,*], Beatriz Villanueva[1], Miguel Á. Toro[2] and Jesús Fernández[1]

[1]Departamento de Mejora Genética Animal, INIA-CSIC, Ctra. de La Coruña, km 7.5, 28040 Madrid, Spain.

[2]Departamento de Producción Agraria, ETSI Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, 28040 Madrid, Spain

*Corresponding author.

The content of this chapter has been published in *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13781

Abstract

For both undivided and subdivided populations, the consensus method to maintain genetic diversity is the Optimal Contribution (OC) method. For subdivided populations, the method determines the optimal contribution of each candidate to each subpopulation to maximize the global genetic diversity (which implicitly optimizes migration between subpopulations) while balancing the relative levels of coancestry between and within subpopulations. Inbreeding can be controlled by increasing the weight given to the within-subpopulation coancestry (λ). Here we extend the original OC method for subdivided populations that used pedigree-based coancestry matrices, to the use of more accurate genomic matrices. Global levels of genetic diversity, measured as expected heterozygosity and allelic diversity, their distributions within and between subpopulations and the migration pattern between subpopulations, were evaluated via stochastic simulations. The temporal trajectory of allele frequencies was also investigated. The genomic matrices investigated were i) the matrix based on deviations of the observed number of alleles shared by two individuals from the expected number under Hardy-Weinberg equilibrium; and ii) a matrix based on a genomic relationship matrix. The matrix based on deviations led to higher global and within-subpopulation expected heterozygosities, lower inbreeding and similar allelic diversity than the second genomic and pedigree-based matrices when a relative high weight was given to the within-subpopulation coancestries ($\lambda \geq 5$). Under this scenario, allele frequencies only moved away slightly from the initial frequencies. Therefore, the recommended strategy is to use the former matrix in the OC methodology giving a high weight to the within-subpopulation coancestry.

Keywords: allele frequency changes, allelic diversity, expected heterozygosity, genomic coancestry, subdivided populations, optimal contributions

1. Introduction

Most populations of endangered species are subdivided into disconnected breeding groups. In nature, the main causes of subdivision are the deterioration of habitats, with the consequent isolation of different subpopulations, and/or the creation of artificial barriers (e.g., roads, railways or fences). In ex-situ conservation programs of wild and domestic species, the fragmentation may be intentional for logistic reasons (e.g., limited space and facilities to keep the population in a single location or ease to manage small subpopulations) or because the subdivision has a clear biological reason, as different subpopulations may be characterized by local adaptations which should be preserved. In any case, the subdivision has the advantage of reducing the risk of extinction of the global population due to different hazards (e.g., fires, predator's attacks and infectious disease outbreaks), as such events could cause the extinction of a particular subpopulation while keeping safe the global population. In fact, one of the criteria of both the Food and Agriculture Organization (FAO) and the IUCN (International Union for Conservation of Nature) for considering a species or breed critically endangered is that it is concentrated in a restricted area (FAO, 2013; IUCN Standards and Petitions Committee, 2022).

On the other hand, the fragmentation of a population constitutes a danger because it implies a reduced census (and, consequently, a reduced effective population size) in each subpopulation. This could lead to high levels of genetic drift within subpopulations with the consequent increase in inbreeding and loss of genetic diversity at the local level (Falconer & Mackay, 1996). This is despite the fact that, at the global population level, genetic drift could be low. Consequently, when a population has been subdivided, it is convenient to favor gene flow between the different subpopulations to reduce the increase in inbreeding and the risk of extinction.

For an undivided population, the consensus method to maintain genetic diversity is the Optimal Contribution (OC) method that, in the context of conservation programs, determines the optimal number of offspring (contribution) that each breeding candidate should produce to maximize genetic diversity, measured as expected heterozygosity. This is achieved by minimizing the global coancestry between candidates (Villanueva *et al*.,

2004; Fernández *et al*., 2003). Fernández *et al*. (2008) extended the OC method to optimally manage subdivided populations. Their method determines the optimal contribution of each candidate to every subpopulation in the next generation to maximize the global genetic diversity while balancing the relative levels of coancestry between and within subpopulations by including specific weights to each term. This implies that, when the within-subpopulation term is given a high weight, the levels of inbreeding within subpopulations can be better controlled (i. e., the inbreeding levels reached are lower). It must be highlighted that the process implicitly leads to the optimization of the migration pattern between subpopulations in order to achieve the intended distribution of diversity. Fernández *et al*. (2008) demonstrated, by computer simulations, that their method maintains higher levels of genetic diversity in the global population and equal or lower inbreeding in each subpopulation than the reference methodology of One Migrant Per Generation (OMPG). The OMPG method (Mills & Allendorf, 1996; Wang, 2004) is based on the island model derived by Wright (1931) and was the standard approach to manage subdivided populations in the past. The efficiency of the dynamic method of Fernández *et al*. (2008) has been demonstrated not only with simulated but also with real data (Caballero *et al*., 2010; Ávila *et al.*, 2011) and has been implemented in the software METAPOP2 (López-Cortegano *et al*., 2019).

As originally proposed, the method of Fernández *et al*. (2008) used pedigree information to compute coancestries and optimize contributions. However, pedigree-based genetic relationships between individuals in wild populations and in many domestic populations are unknown, especially those between individuals belonging to different subpopulations. In such situations, relationships can be estimated using genetic marker information. In fact, for undivided populations it has been shown that, when the number of markers is large enough, the use of molecular-based (genomic) coancestries in the OC method leads to a more efficient maintenance of genetic diversity than the use of pedigree-based coancestries (de Cara *et al*., 2011; 2013; Gómez-Romano *et al*., 2013). The increase in the efficiency of the management is due to the fact that genomic coefficients measure the actual proportion of loci that two particular individuals have in common (i.e., they give the realized relationships) while pedigree-based coefficients give

only expectations of these proportions that can differ from the exact proportions. The potential benefit of using genomic coancestries in the management of subdivided populations is worth investigating to determine if the comparison with pedigree-based management produces the same results as in the undivided population scenario.

Several authors (Lacy, 2000; Frankham, 2008; Saura *et al*., 2008; Meuwissen *et al*., 2020) argue that maximizing the global genetic diversity, per se, may have negative consequences given that the population genetic composition is modified. For instance, in the context of an ex-situ conservation program, changing the genetic composition of the population can affect its survival once it is reintroduced to the wild. Management aimed at maximizing diversity may lead to increased frequencies of deleterious recessive alleles and may also disrupt positive interactions between loci which have occurred due to many generations of natural selection (Schoen *et al*., 1998; Fernández & Caballero, 2001; Saura *et al*., 2008). It is thus also worthwhile to investigate potential changes in the genetic composition (i.e., in allele frequencies) of the population when applying the OC methodology.

Several genomic coancestry matrices have been proposed (Villanueva *et al*., 2021) and recent studies with undivided populations have showed that their use in OC can have different consequences in terms of the genetic diversity maintained and the evolution of allele frequencies (Gómez-Romano *et al*., 2016; Meuwissen *et al*., 2020; **Chapter 2**). In particular, results of these previous studies indicate that the use of measures of coancestry based on alleles sharing (e.g., Nejati-Javaremi *et al*., 1997) in OC resulted in a higher genetic diversity (measured as expected heterozygosity) than the use of coancestry computed from realized relationship matrices commonly used in genetic evaluations in animal breeding (VanRaden, 2008). However, the use of the latter maintained allele frequencies closer to those in the base population. Recently, Meuwissen *et al*. (2020) have shown that the use of VanRaden's matrices manages drift and limits changes in allele frequency at the expense of a higher rate of increase in homozygosity. It is expected that the use of the different genomic matrices would have also different consequences when managing subdivided populations, not only on the global levels of diversity and the fate of the allelic frequencies, but also on the distribution of diversity

across subpopulations.

The objective of this study was to evaluate, via stochastic simulations, the efficiency of using genomic coancestry matrices in the OC method for maintaining genetic diversity in subdivided populations. The global genetic diversity, its distribution within and between subpopulations and the migration flow between subpopulations were evaluated. Also, the trajectory of allele frequencies under this management method was investigated. Results obtained when using genomic information were compared with those obtained using pedigree information.

2. Materials and Methods

All scenarios simulated involved the management of a subdivided population composed of five subpopulations, mimicking the structure of the captive breeding program of the Iberian lynx (Kleinman-Ruiz *et al*., 2019). Management was carried out using the Fernández et al. (2008) development of OC for subdivided populations, directed to the maintenance of genetic diversity in the global population (measured as expected heterozygosity) while restricting the increase in inbreeding within subpopulations by giving different weights to the within-subpopulation coancestry term. Management was carried out for ten discrete generations to evaluate the different scenarios in the short and long term. Matings were performed within subpopulations (i.e., offspring was always generated from couples belonging to the same subpopulation). Subsequently, offspring migrated to other subpopulations (if required) following the solutions arising from the optimization. Results from the implementation of the method using different genomic coancestry matrices were compared with those using the pedigree-based coancestry matrix. The conservation program started from a base population made up of 100 individuals, which was created in several steps. Firstly, a population at mutation-drift equilibrium was generated. Secondly, the population was expanded in order to have enough individuals for sampling 100 replicates. Thirdly, in some scenarios, extra random mating generations were performed in order to create subpopulations genetically differentiated before starting the management. The detailed steps taken in the simulations

are given below. The simulations were carried out using in-house Fortran 90 codes.

2.1 *Generation of the base population*

As stated above, the creation of the base population was carried out in several steps. In a first step, a population of size $N = 100$ was simulated during 10,000 generations of random mating in order to create enough levels of linkage disequilibrium between markers used in the management and other loci in the genome where diversity also needs to be maintained (see below). Using a larger N would have generated unrealistic low levels of linkage disequilibrium. The genome was composed of 20 chromosomes of one Morgan each. Two types of biallelic loci (500,000 SNPs and 500,000 'unobserved loci' per chromosome) were simulated. Both types of loci were evenly distributed and interspersed (i.e., SNPs and unobserved alternated along a particular chromosome). SNPs and unobserved loci differed simply in their subsequent use. SNP loci were used for computing the genomic coancestry matrices involved in the management and the unobserved loci were used for calculating the different parameters evaluated (i.e., they were used for monitoring purposes). Thus, the effect of using different genomic coancestry matrices in OC can be evaluated on the whole genome and not only on the loci used in the management (i.e., the SNPs). At the beginning of the process, all loci were fixed. The mutation rate per locus and generation (μ) was 2.5×10^{-6} for all loci. When producing the gametes, the number of crossovers per chromosome was drawn from a Poisson distribution with mean equal to one. Crossovers were randomly distributed without interference. At the end of the process, the expected heterozygosity measured at both types of loci had stabilized, approaching thus a mutation-drift equilibrium.

In a second step, the population was expanded during four generations in order to create enough individuals to sample 100 different replicates with 100 individuals each. During this expansion, each individual was randomly mated to eight different individuals and each mating produced one offspring. Thus, the number of individuals was quadrupled each generation. At the end of this process, the population was composed by 25,600 individuals (half females and half males).

Each replicate was created by initially sampling 100 individuals (half of each sex) at random from the expanded population. Then two different population structures were simulated. These mainly reflected contrasting levels of differentiation between subpopulations in the base population (t = 0 thereafter) where management started (Figure 1):

1. Scenarios 'Equal' (E). In these scenarios, all subpopulations were equally related, and all individuals had equal or very similar inbreeding coefficients at t = 0. This was a consequence of randomly distributing individuals among subpopulations directly from the expanded population. There was a total of five subpopulations comprising ten males and ten females each (Figure 1a). At t = 0 pedigree-based inbreeding coefficients were zero. Pedigree-based coancestry coefficients between and within subpopulations were also zero except self-coancestries which had a value equal to 0.5 (consequently, the average coancestry within subpopulations was one divided by twice the number of individuals in the subpopulation, i.e., 1/40 in our simulations). At this generation, the genomic coefficients of coancestry between and within subpopulations and the genomic inbreeding coefficients within subpopulations were very similar across subpopulations. Only loci segregating in a particular replicate at t = 0 were used in the management (SNPs) and in the calculation of the parameters evaluated (unobserved loci). The average number of SNPs and unobserved loci still segregating at t = 0 across replicates in the global population was 48,830 and 48,720, respectively. The expected heterozygosity for the global population computed with all loci (SNPs and unobserved loci) still segregating was 0.193.

2. Scenarios 'Unequal' (U). In these scenarios, one of the subpopulations was generated to be genetically differentiated and more inbred than the other four at t = 0. To obtain this specific structure, the 100 individuals initially sampled from the expanded population, were divided in two groups. One of the groups was composed by ten males and ten females and the other group was composed by the rest of individuals (40 males and 40 females). Then, five discrete generations of random mating were carried out within each group, keeping size and sex ratio constant (Figure 1b). Afterwards, the largest group was divided into four subpopulations of equal size (i.e.,

ten males and ten females each). These four subpopulations (subpopulations 2 to 5) together with the subpopulation isolated before (subpopulation 1) constituted the base population (t = 0) from which management started. Pedigree was recorded during the five random mating generations and, therefore, pedigree-based coancestries and inbreeding coefficients at t = 0 had nonzero values. Notice that, in these scenarios, the pedigree-based and the genomic coancestries between an individual of subpopulation 1 and individuals from any of the other subpopulations were higher than the coancestries between individuals of subpopulations 2 to 5. This translates into a greater genetic differentiation between subpopulation 1 and the rest of the subpopulations. Similarly, inbreeding coefficients were higher in subpopulation 1 than in the rest. As before, only loci segregating in a particular replicate at t = 0 were used and on average, they were 48,519 SNPs and 48,376 unobserved loci. The expected heterozygosity for the global population computed with all loci (SNPs and unobserved loci) still segregating was 0.188.

2.2 Management method

In all scenarios, management was performed following the methodology proposed by Fernández *et al*. (2008). Briefly, the aim of the methodology is to determine the contributions (i.e., the number of offspring from each potential parent) that maximize the global amount of diversity (measured as expected heterozygosity) in the next generation. As we deal with subdivided populations, this diversity can be partitioned into within- and between-subpopulation diversity. Consequently, the objective function to be minimized also includes two terms: one related to the coancestry coefficient between subpopulations (B) and another term related to the coancestry coefficient within-subpopulations (W). Additionally, these terms can be weighted differentially by including a factor (λ). Specifically, the formulation would be $B + \lambda W$, where $B = \sum_{k=1}^{n} \sum_{l \neq k}^{n} \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} c_{ik} c_{jl}$, $W = \sum_{k=1}^{n} \sum_{i=1}^{N} \sum_{j=1}^{N} f_{ij} c_{ik} c_{jk}$, λ is a weighting factor, n is the number of subpopulations, N is the number of individuals in the global population, f_{ij} is the coancestry coefficient between individuals i and j, and c_{ik} is the contribution of candidate i to subpopulation k (i.e., the number of offspring generated by that candidate

to be raised in subpopulation *k* and, consequently, restricted to be a positive integer). Thus, this formulation reflects that, when dealing with a structured (subdivided) population, the contribution of a particular individual can be partitioned into its contribution to each subpopulation. Note that B is the term corresponding to the coancestry of candidates generating offspring to be allocated to different subpopulations (and, thus, it is proportional to the diversity between subpopulations in the next generation), W is the term corresponding to the coancestry of candidates generating offspring to be reared in the same subpopulation (and, thus, it is proportional to the diversity within subpopulations) and λ is a factor balancing the relative importance of within-subpopulation coancestry, and consequently, the level of inbreeding for each subpopulation. In this study, the relative weight given to within-subpopulation coancestry (λ) took values of one, five or ten. The higher the value of λ, the lower the expected levels of within-subpopulation coancestry and inbreeding.

The methodology also allows to restrict the number of migrants by imposing the constraint $\sum_{k=1}^{N}\sum_{l\neq k}^{n} c_{ik} \leq 2nM$, where M is the maximum number of individuals allowed to move (on average) from one subpopulation to another subpopulation per generation. In all scenarios this value was restricted to one, implying that the maximum total number of migrants allowed per generation was five. This is a rate acceptable for most conservation programs when considering the logistic problems that the movement of individuals may have (Mills & Allendorf, 1996; Wang, 2004). Constraints to guarantee that the subpopulation sizes and the sex ratio were kept constant across generations were also applied. In addition, given that that breeding was imposed to be performed within subpopulations, the sum of contributions (number of offspring) of females breeding in subpopulation *k* to subpopulation *l* was forced to be equal to the sum of contributions of males breeding in subpopulation *k* to subpopulation *l*. The optimization was performed using a simulated annealing algorithm that is described in detail in Fernández & Toro (1999). All restrictions were satisfied by penalizing solutions not fitting them during the search performed by the algorithm.

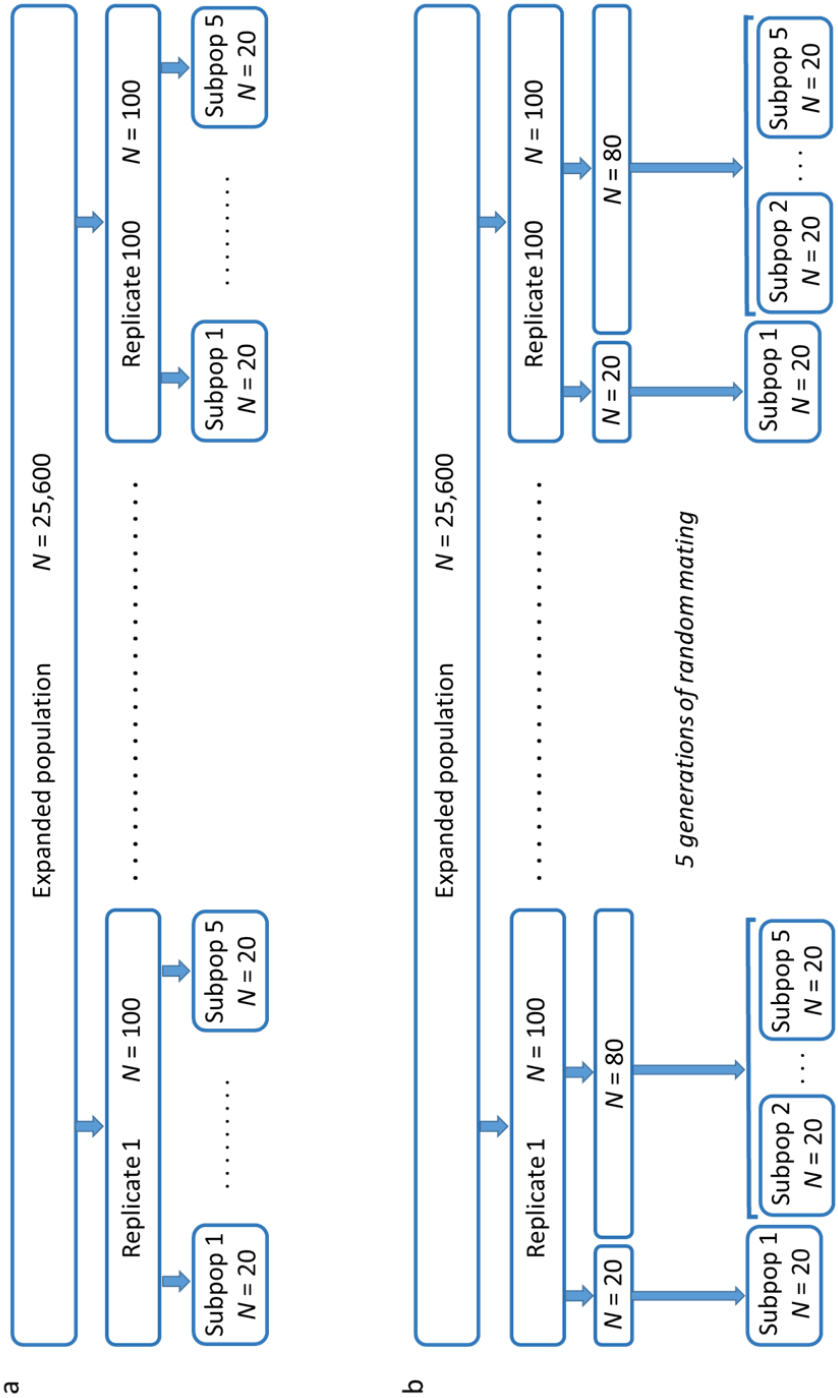

Figure 1. Diagram of the steps given for generating the base population in scenarios with different population structures: Equal (a, all subpopulations were equally related) and Unequal (b, subpopulation 1 was genetically differentiated and more inbred than the other four subpopulations).

Management was implemented using three estimates of coancestry (*f*), including one estimate derived from the pedigree (f_P) and two estimates derived from genomic information (i.e., from the SNPs segregating at $t = 0$). The two genomic coancestry coefficients used were:

1. $f_{L\&H}$: the coancestry coefficient describing the excess of observed number of alleles shared by two individuals relative to the expected number under Hardy-Weinberg equilibrium (Li & Horvitz, 1953; Toro *et al*., 2002). Specifically, the coancestry coefficient between individuals *i* and *j* was computed as

$$f_{L\&H(i,j)} = \frac{\sum_{k=1}^{S} f_{OBS(i,j)k} - S + 2\sum_{k=1}^{S} p_k(1-p_k)}{2\sum_{k=1}^{S} p_k(1-p_k)},$$

   where $f_{OBS(i,j)}$ is the proportion of alleles shared by both individuals, *S* is the number of SNPs and p_k is the frequency of the reference allele of SNP *k* at $t = 0$.
2. f_{VR2}: the coancestry coefficient computed from the genomic relationship matrix obtained using method two described in VanRaden (2008) and proposed by Amin *et al*. (2007). Specifically, the coancestry coefficient between individuals *i* and *j* was computed as

$$f_{VR2(i,j)} = \frac{1}{2S}\sum_{k=1}^{S} \frac{(x_{ki} - 2p_k)(x_{kj} - 2p_k)}{2p_k(1-p_k)},$$

   where x_{ki} is the genotype of individual *i* for SNP *k*, coded as zero, one or two for genotypes *AA*, *AB* and *BB*, respectively, and p_k is the allele frequency as defined for $f_{L\&H}$.

These coefficients have been widely used in the literature, but under different names (see Table 1 of Villanueva *et al*., 2021). Here, we used the terminology given in Villanueva *et al*. (2021). Matrices constructed with coefficients f_{PED}, $f_{L\&H}$ and f_{VR2} will be referred to as $\boldsymbol{\theta}_{\mathbf{PED}}$, $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$, respectively. $\boldsymbol{\theta}_{\mathbf{VR2}}$ differs from $\boldsymbol{\theta}_{\mathbf{L\&H}}$ in that with $\boldsymbol{\theta}_{\mathbf{VR2}}$ rare alleles contribute more to the coancestry coefficient than common alleles (Gómez-Romano *et al*., 2016; **Chapter 2**). In fact, the correlation between VanRaden's and Li & Horvitz's coefficients increases when only SNPs with high MAF are used (**Chapter 1** and **2**; Villanueva *et al*., 2021).

Coefficients $f_{L\&H}$ and f_{VR2} depend on allele frequencies at $t = 0$. In the case of a subdivided population, it is unclear which frequencies should be used as they could be those in the global population or the averages of frequencies of the two subpopulations to which two particular individuals belong. In scenarios E, allele frequencies were similar for all subpopulations and, therefore, only global frequencies were used to compute $f_{L\&H}$ and f_{VR2}. In scenarios U, both approaches (global or average subpopulation frequencies) were considered.

Summarizing, the scenarios simulated are combinations of four factors: i) the population structure (E or U); ii) the weight given to the within-subpopulation coancestry (λ); iii) the coancestry matrix used in the optimization ($\boldsymbol{\theta_{PED}}$, $\boldsymbol{\theta_{L\&H}}$ or $\boldsymbol{\theta_{VR2}}$); and iv) the frequencies used when computing $\boldsymbol{\theta_{L\&H}}$ and $\boldsymbol{\theta_{VR2}}$. The different scenarios are summarized in Table 1.

2.3 Parameters estimated

Management scenarios were compared in terms of the genetic diversity retained in the global population and its distribution between and within subpopulations. All parameters were computed using the unobserved loci. The genetic diversity in the global population (H_T) was measured as the expected heterozygosity. The genetic diversity within subpopulations (H_S) was measured as the average expected heterozygosity across subpopulations. The expected heterozygosity was computed each generation as $1 - \sum_{k=1}^{L}\sum_{l=1}^{2} p_{kl}^2$, where L is the number of loci and p_{kl} is the frequency of allele l of locus k (calculated for the global population or for subpopulations). The genetic diversity between subpopulations was calculated as $D = H_T - H_S$. This parameter reflects the degree of differentiation (distance) between subpopulations and is equal to the Nei's minimum genetic distance (e.g., Toro & Caballero, 2005).

Another measure of genetic diversity used was the number of segregating loci at a given generation t, both at the global level and within subpopulations. It is given as the percentage of loci that continued segregating at t relative to those segregating at $t = 0$. Note that the number of segregating loci is a measure of allelic diversity when biallelic loci are used.

Table 1. Scenarios simulated (marked with *) that varied in combinations of the initial population structure, the weight given to the within-subpopulation coancestry (λ), the coancestry matrix used in the optimization and the frequencies used when computing the genomic coancestry matrices $\boldsymbol{\Theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\Theta}_{\mathbf{VR2}}$.

			Coancestry matrix[c]		
Population structure[a]	λ	Frequencies used[b]	$\boldsymbol{\theta}_{\mathbf{PED}}$	$\boldsymbol{\theta}_{\mathbf{L\&H}}$	$\boldsymbol{\theta}_{\mathbf{VR2}}$
Equal	1	Global	*	*	*
	5	Global	*	*	*
	10	Global	*	*	*
Unequal	1	Global	*	*	*
	5	Global	*	*	*
	10	Global	*	*	*
	1	Subpopulations		*	*
	5	Subpopulations		*	*
	10	Subpopulations		*	*

[a]Equal: all subpopulations were equally related; Unequal: one of the subpopulations was genetically differentiated and more inbred than the other four.
[b]Global: using initial frequencies in the global population; Subpopulations: using the average initial frequencies of the subpopulations involved in the calculation of a particular coancestry coefficient.
[c]$\boldsymbol{\theta}_{\mathbf{PED}}$: pedigree-based matrix; $\boldsymbol{\theta}_{\mathbf{L\&H}}$: Li & Horvitz genomic matrix; $\boldsymbol{\theta}_{\mathbf{VR2}}$: VanRaden genomic matrix.

The average molecular inbreeding was computed as the observed homozygosity (i.e., the proportion of homozygous loci), in the global population (F_T) and within a particular subpopulation i (F_{Si}).

The different scenarios were also compared in terms of the evolution across generations of the average frequency of the minor allele (measured at the global population level but also within subpopulations) and in terms of the migration pattern. Specifically, the migration pattern was focussed on the number of migrants sent to and received by subpopulation 1 because, as stated above, this subpopulation was genetically differentiated (and more inbred) at $t = 0$ in scenarios U. Results presented for all parameters are averages across the 100 replicates.

3. Results

Increasing λ from one to five led to different results for expected heterozygosity, allelic diversity, levels of inbreeding and changes in the average allele frequency. However, increasing λ from five to ten led to almost the same outcomes. For this reason, only the results for $\lambda = 1$ and $\lambda = 5$ are shown.

3.1 Genetic diversity

When management was based on $\boldsymbol{\theta}_{\mathbf{PED}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$, both the expected heterozygosity in the global population (H_T) and within subpopulations (H_S) decreased across generations in all scenarios (Table 2). The opposite trend was observed for the genetic distance between subpopulations (D) indicating that the subpopulations diverged over time. While H_T was insensitive to the value of λ, H_S was slightly lower (and D was slightly higher) for the lowest λ.

On the other hand, when the management was based on $\boldsymbol{\theta}_{\mathbf{L\&H}}$, H_T increased across generations, and consequently reached values higher than those achieved when the management was based on $\boldsymbol{\theta}_{\mathbf{PED}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$ after a single generation of management. Contrarily, H_S decreased across generations when using $\boldsymbol{\theta}_{\mathbf{L\&H}}$ in the optimization. For $\lambda = 1$, this decrease was greater than when using $\boldsymbol{\theta}_{\mathbf{PED}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$. Thus, using $\boldsymbol{\theta}_{\mathbf{L\&H}}$ with $\lambda = 1$ led to higher differentiation among subpopulations (i.e., higher D) than using other coancestry matrices. However, for $\lambda = 5$, H_S decreased less when using $\boldsymbol{\theta}_{\mathbf{L\&H}}$ than when using $\boldsymbol{\theta}_{\mathbf{PED}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$ and still kept higher levels of H_T. Therefore, it seems that the use of $\boldsymbol{\theta}_{\mathbf{L\&H}}$ with $\lambda = 5$ in OC could be the strategy to follow as it leads to higher levels of genetic diversity both in the global population and within subpopulations than the use of $\boldsymbol{\theta}_{\mathbf{PED}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$. For $\lambda = 5$, D values were similar when using different coancestry matrices.

Table 3 shows the evolution of the allelic diversity (measured as the percentage of unobserved loci segregating at a given generation) in the global population (LT), within subpopulation 1 (LS1), and averaged across subpopulations 2 to 5 (LS2-5). LT barely decreased in the global population throughout the management period regardless of the coancestry matrix and the λ used. For $\lambda = 1$, LS2-5 decreased faster when using

Table 2. Average expected heterozygosity in the global population (H_T) and within (H_S) and between (D) subpopulations across generations (t) when contributions are optimized using pedigree-based ($\boldsymbol{\theta}_{\mathbf{PED}}$), Li & Horvitz ($\boldsymbol{\theta}_{\mathbf{L\&H}}$) and VanRaden ($\boldsymbol{\theta}_{\mathbf{VR2}}$) coancestry matrices for two different weights given to the within-subpopulation coancestry (λ) and two different population structures. Matrices $\boldsymbol{\Theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\Theta}_{\mathbf{VR2}}$ were computed using the initial allele frequencies in the global population.[a]

Population structure	λ	t	$\boldsymbol{\theta}_{\mathbf{PED}}$			$\boldsymbol{\theta}_{\mathbf{L\&H}}$			$\boldsymbol{\theta}_{\mathbf{VR2}}$		
			H_T	H_S	D	H_T	H_S	D	H_T	H_S	D
Equal	1	0	0.192	0.189	0.004	0.192	0.189	0.004	0.192	0.189	0.004
		1	0.192	0.187	0.005	0.193	0.184	0.010	0.192	0.187	0.005
		5	0.190	0.180	0.010	0.195	0.174	0.021	0.190	0.179	0.011
		10	0.188	0.174	0.014	0.196	0.169	0.027	0.188	0.173	0.015
	5	0	0.192	0.189	0.004	0.192	0.189	0.004	0.192	0.189	0.004
		1	0.192	0.187	0.005	0.193	0.187	0.006	0.192	0.187	0.005
		5	0.190	0.180	0.010	0.194	0.183	0.010	0.190	0.181	0.009
		10	0.188	0.177	0.011	0.194	0.183	0.011	0.188	0.177	0.010
Unequal	1	0	0.188	0.180	0.008	0.188	0.180	0.008	0.188	0.180	0.008
		1	0.188	0.178	0.010	0.189	0.176	0.013	0.187	0.178	0.009
		5	0.186	0.173	0.013	0.190	0.167	0.024	0.186	0.172	0.013
		10	0.184	0.169	0.015	0.191	0.162	0.030	0.183	0.167	0.016
	5	0	0.188	0.180	0.008	0.188	0.180	0.008	0.188	0.180	0.008
		1	0.188	0.179	0.009	0.189	0.179	0.009	0.187	0.179	0.008
		5	0.186	0.176	0.010	0.189	0.178	0.011	0.185	0.176	0.010
		10	0.183	0.172	0.011	0.189	0.177	0.012	0.183	0.172	0.010

[a] Standard errors ranged from 6.58×10^{-5} to 4.87×10^{-4} for H_T and H_S.

$\boldsymbol{\theta}_{\mathbf{L\&H}}$ than when using $\boldsymbol{\theta}_{\mathbf{PED}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ in OC. Differences in LS2-5 across scenarios using different matrices almost disappeared for $\lambda = 5$. Given that all subpopulations were equally related initially in scenarios E, the percentage of loci that remained segregating at t = 0 and at subsequent generations in subpopulation 1 (LS1) was the same as in the rest of the subpopulations (LS2-5). However, in scenarios U, LS1 at t = 0 was lower than LS2-5, as expected. Management over generations reduced the differences between LS1

and LS2-5. This reduction was faster with λ = 5 and at t = 10, LS1 and LS2-5 were very similar regardless of the matrix used.

Summarizing, allelic diversity in the global population remained almost at its initial levels in all scenarios. However, H_T was always higher when using $\boldsymbol{\theta}_{\mathbf{L\&H}}$ than when using $\boldsymbol{\theta}_{\mathbf{PED}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$. The advantage of $\boldsymbol{\theta}_{\mathbf{L\&H}}$ held for $\lambda = 5$ when considering genetic diversity within subpopulations, as a similar level of allelic diversity and more expected heterozygosity (H_S) was retained when using this matrix.

3.2 Inbreeding

Table 4 shows inbreeding for the global population (F_T), subpopulation 1 (F_{S1}) and the average for subpopulations 2 to 5 ($F_{S2\text{-}5}$) across generations for the different scenarios. Note that F_T is simply the average inbreeding across subpopulations. As expected, strategies that led to higher/lower H_S (Table 2) led to lower/higher inbreeding (Table 4). It must be recalled that, as matings were at random in each subpopulation, the expected ($1 - H_S$) and observed (F_S) homozygosity must be similar.

Management using $\boldsymbol{\theta}_{\mathbf{L\&H}}$ reduced the levels of inbreeding (F_T, F_{S1} and $F_{S2\text{-}5}$) in the first and second (not shown) generations for any value of λ and for both population structures but for $\lambda = 1$, after the initial decrease there was a faster increase than when using $\boldsymbol{\theta}_{\mathbf{PED}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$. Consequently, at $t > 2$ global and subpopulation inbreeding levels were higher with $\boldsymbol{\theta}_{\mathbf{L\&H}}$ than with $\boldsymbol{\theta}_{\mathbf{PED}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$. However, for $\lambda = 5$ the rate of increase in global and subpopulation inbreeding was faster with $\boldsymbol{\theta}_{\mathbf{PED}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ than with $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and, thus, management using the latter led to less inbreeding. As expected, inbreeding was very similar in all subpopulations under scenarios E and it was lower for $\lambda = 5$ than for $\lambda = 1$ (Table 4). In scenarios U, the difference in inbreeding between subpopulation 1 (initially more inbred) and the rest of subpopulations was effectively reduced by the management, and this reduction was faster for $\lambda = 5$.

Table 3. Average allelic diversity (measured as the percentage of loci segregating) in the global population (L_T), in subpopulation 1 (L_{S1}) and average percentage in subpopulations 2 to 5 ($L_{S2\text{-}5}$) across generations (t) when contributions are optimized using pedigree-based ($\boldsymbol{\theta}_{\mathbf{PED}}$), Li & Horvitz ($\boldsymbol{\theta}_{\mathbf{L\&H}}$) and VanRaden ($\boldsymbol{\theta}_{\mathbf{VR2}}$) coancestry matrices for two different weights given to the within-subpopulation coancestry (λ) and two different population structures. L_{S1} and $L_{S2\text{-}5}$ values obtained with $\boldsymbol{\theta}_{\mathbf{PED}}$ and L_T, L_{S1} and $L_{S2\text{-}5}$ values obtained with $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ are those deviated from L_T obtained with $\boldsymbol{\theta}_{\mathbf{PED}}$. Matrices $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ were computed using the initial allele frequencies in the global population.[a]

Population structure	λ	t	$\boldsymbol{\theta}_{\mathbf{PED}}$			$\boldsymbol{\theta}_{\mathbf{L\&H}}$			$\boldsymbol{\theta}_{\mathbf{VR2}}$		
			L_T	L_{S1}	$L_{S2\text{-}5}$	L_T	L_{S1}	$L_{S2\text{-}5}$	L_T	L_{S1}	$L_{S2\text{-}5}$
Equal	1	0	100.0	–14.6	–14.7	+0.0	–14.6	–14.7	+0.0	–14.6	–14.7
		1	99.9	–19.2	–19.3	–0.2	–25.6	–25.8	+0.0	–19.3	–19.3
		5	99.6	–27.9	–27.9	–0.3	–37.3	–36.8	+0.0	–28.6	–28.5
		10	99.3	–32.0	–32.3	–0.4	–40.2	–40.4	+0.0	–34.0	–33.7
	5	0	100.0	–14.6	–14.7	+0.0	–14.6	–14.7	+0.0	–14.6	–14.7
		1	99.9	–19.2	–19.3	+0.0	–20.4	–20.6	+0.0	–19.2	–19.2
		5	99.6	–27.4	–27.4	–0.1	–28.8	–28.8	+0.0	–27.1	–27.2
		10	99.3	–31.2	–31.0	–0.1	–31.7	–31.5	+0.0	–30.8	–30.8
Unequal	1	0	100.0	–39.4	–23.2	+0.0	–39.4	–23.2	+0.0	–39.4	–23.2
		1	99.9	–38.8	–26.7	–0.1	–40.9	–30.8	+0.0	–37.9	–26.1
		5	99.7	–37.0	–32.3	–0.3	–42.4	–39.5	+0.0	–37.3	–32.7
		10	99.5	–37.4	–35.5	–0.3	–43.6	–42.6	–0.1	–38.6	–36.6
	5	0	100.0	–39.4	–23.2	+0.0	–39.4	–23.2	+0.0	–39.4	–23.2
		1	99.9	–36.1	–25.6	+0.0	–36.6	–26.6	+0.0	–36.2	–25.6
		5	99.7	–31.3	–31.6	–0.1	–33.6	–32.8	+0.0	–32.7	–31.6
		10	99.4	–34.1	–34.3	–0.1	–35.7	–35.2	+0.0	–34.5	–34.7

[a] Standard errors for the number of unobserved loci segregating were less than 1.47 x 10^{-2}.

3.3 Change in allele frequencies

Under scenarios U, subpopulation 1 started management (t = 0) with a MAF (minimum allele frequency) lower than other subpopulations (Figure 2) due to the greater genetic drift that suffered as it was isolated from the rest during the five previous generations. When using $\theta_{L\&H}$, the average MAF in subpopulation 1 and in the global population always increased across generations. This is due to the greater efficiency of $\theta_{L\&H}$ to maintain expected heterozygosity, which takes the highest value at intermediate frequencies. Contrarily, when using θ_{PED} and θ_{VR2}, the average MAF decreased in the global population, but this decrease was less pronounced than the increase observed with $\theta_{L\&H}$; i.e., θ_{PED} and θ_{VR2} maintained frequencies closer to the initial values than $\theta_{L\&H}$. The difference between MAF in the global population and in subpopulation one became smaller over time, especially with λ = 5 and when using θ_{PED} or θ_{VR2}.

3.4 Migrants

In scenarios E, on average, each subpopulation sent one individual to (and received one individual from) another subpopulation across generations (results not shown). For λ = 1, this was also the case in scenarios U when using $\theta_{L\&H}$ or θ_{VR2} computed from the global initial frequencies or θ_{PED} (Figures 3a and c, dotted lines).

However, for λ = 5 (Figures 3a and c, dashed lines), whatever the matrix used in OC, subpopulation 1 always sent three or four migrants to other subpopulations in the first generation and, as the generations went by, the number of migrants sent decreased (one migrant after four or five generations). Also, for λ = 5 the migrants received by subpopulation 1 from the rest was initially on average slightly higher than one and after few generations (about five) stabilized around one. These outcomes are a reflection of the balance between maximizing genetic diversity and controlling within-subpopulation inbreeding implicit in the method. Note that subpopulation 1 was the most inbred but also was the most genetically differentiated (i.e., it harbored particular genetic information) and, thus, moving individuals from subpopulation 1 helped to reduce inbreeding in subpopulations 2 to 5. The differential migration rate toward subpopulation 1 led to similar expected heterozygosities (Table 2), same percentage of loci segregating (Table

Table 4. Average molecular inbreeding in the global population (F_T), subpopulation 1 (F_{S1}) and subpopulations 2 to 5 ($F_{S2\text{-}5}$) across generations (t) when contributions are optimized using pedigree-based ($\boldsymbol{\theta}_{\mathbf{PED}}$), Li & Horvitz ($\boldsymbol{\theta}_{\mathbf{L\&H}}$) and VanRaden ($\boldsymbol{\theta}_{\mathbf{VR2}}$) coancestry matrices for two different weights given to the within-subpopulation coancestry (λ) and two different population structures. Matrices $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ were computed using the initial allele frequencies in the global population.[a]

Population structure	λ	t	$\boldsymbol{\theta}_{\mathbf{PED}}$			$\boldsymbol{\theta}_{\mathbf{L\&H}}$			$\boldsymbol{\theta}_{\mathbf{VR2}}$		
			F_T	F_{S1}	$F_{S2\text{-}5}$	F_T	F_{S1}	$F_{S2\text{-}5}$	F_T	F_{S1}	$F_{S2\text{-}5}$
Equal	1	0	0.807	0.807	0.807	0.807	0.807	0.807	0.807	0.807	0.807
		1	0.807	0.807	0.807	0.804	0.805	0.804	0.807	0.807	0.807
		5	0.814	0.813	0.814	0.815	0.815	0.815	0.813	0.813	0.813
		10	0.819	0.819	0.819	0.822	0.822	0.822	0.819	0.820	0.819
	5	0	0.807	0.807	0.807	0.807	0.807	0.807	0.807	0.807	0.807
		1	0.807	0.807	0.807	0.805	0.805	0.805	0.807	0.807	0.807
		5	0.813	0.812	0.813	0.808	0.809	0.808	0.812	0.812	0.812
		10	0.816	0.816	0.817	0.809	0.809	0.809	0.815	0.815	0.815
Unequal	1	0	0.814	0.825	0.811	0.814	0.825	0.811	0.814	0.825	0.811
		1	0.815	0.826	0.812	0.812	0.824	0.810	0.815	0.827	0.812
		5	0.820	0.828	0.819	0.822	0.830	0.822	0.820	0.828	0.819
		10	0.825	0.829	0.824	0.829	0.833	0.829	0.825	0.830	0.824
	5	0	0.814	0.825	0.811	0.814	0.825	0.811	0.814	0.825	0.811
		1	0.814	0.825	0.811	0.813	0.823	0.810	0.815	0.825	0.812
		5	0.817	0.818	0.817	0.814	0.815	0.814	0.817	0.818	0.817
		10	0.821	0.821	0.821	0.815	0.815	0.815	0.821	0.820	0.821

[a] Standard errors ranged from 1.45×10^{-4} to 5.16×10^{-4}.

3) and similar levels of inbreeding (Table 4) in all subpopulations at $t = 10$.

3.5 Effect of using different initial allele frequencies when computing $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$

Using the initial allele frequencies of subpopulations to compute $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ led to similar allelic diversity (data not shown) but, in general, to lower expected heterozygosity (H_T and H_S) and higher inbreeding (F_{S1} and $F_{S2\text{-}5}$) than using initial

frequencies in the global population (Table 5). Genetic differences between subpopulations (*D*) increased when using local frequencies. The largest difference between using global and subpopulation frequencies was for θ_{VR2}, particularly with λ = 5. It was interesting to note that management based on θ_{VR2} computed using allele frequencies of subpopulations led to the same results for different values of λ (Table 5 and Figures 2, 3b and 3d).

Figure 2 shows that management with θ_{VR2} computed with subpopulation frequencies led to greater changes in MAF than when the matrix was computed using global frequencies (except in subpopulation 1 with λ = 1). At a global level, the use of $\theta_{L\&H}$ was more insensitive to the frequencies used, although this was not the case for subpopulation 1. Specifically, the use of subpopulation frequencies made the MAF of subpopulation 1 rise less, which implies that the allele frequencies were kept closer to those in the base population than when using global frequencies (Figures 2a and 2c). Also, for the global population, frequencies were closer to the initial values with $\theta_{L\&H}$ than with θ_{VR2} when using subpopulation frequencies.

The migration flow greatly changed when the frequencies of subpopulations were used to estimate the genomic coancestry matrices (Figure 3). In scenarios using θ_{VR2}, subpopulation 1 sent and received one migrant on average in all generations for any λ. However, in scenarios using $\theta_{L\&H}$ and λ = 1, subpopulation 1 sent four or five migrants per generation without receiving any contribution from other subpopulations for the whole period of management. For λ = 5, the initial large contribution of subpopulation 1 (i.e., five migrants sent to other subpopulations) gradually decreased with time, stabilizing around two migrants. In parallel, a small number of individuals (average < 1) was received by subpopulation 1 in all generations in scenarios using $\theta_{L\&H}$ and λ = 5.

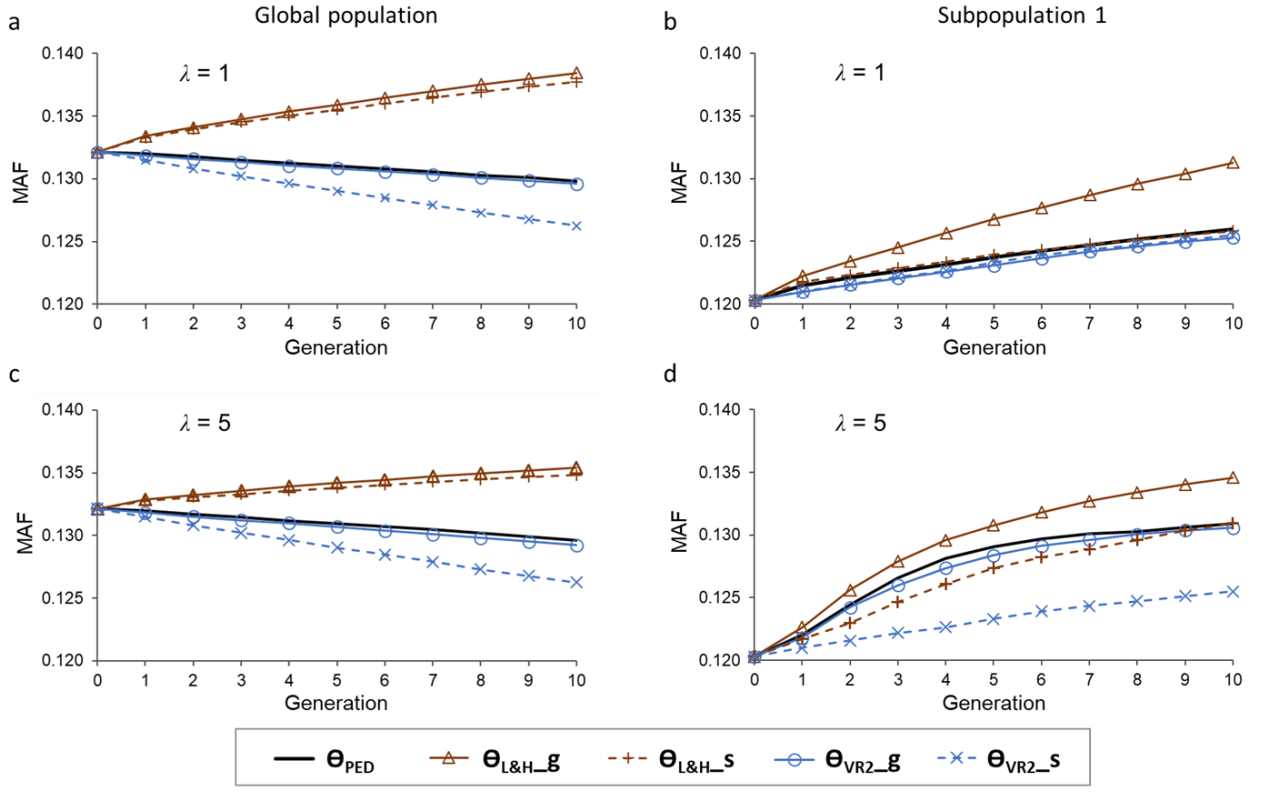

Figure 2. Average frequency of the minor allele (MAF) in the global population (left panels) and in subpopulation 1 (right panels) across generations when contributions are optimized using pedigree-based ($\boldsymbol{\theta}_{\mathbf{PED}}$), Li & Horvitz ($\boldsymbol{\theta}_{\mathbf{L\&H}}$) and VanRaden ($\boldsymbol{\theta}_{\mathbf{VR2}}$) coancestry matrices and two different weights are given to the within-subpopulation coancestry ($\lambda = 1$ and $\lambda = 5$), for scenarios with Unequal population structure. Matrices $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\Theta}_{\mathbf{VR2}}$ were computed using global (subscript '_g') or subpopulation initial allele frequencies (subscript '_s').

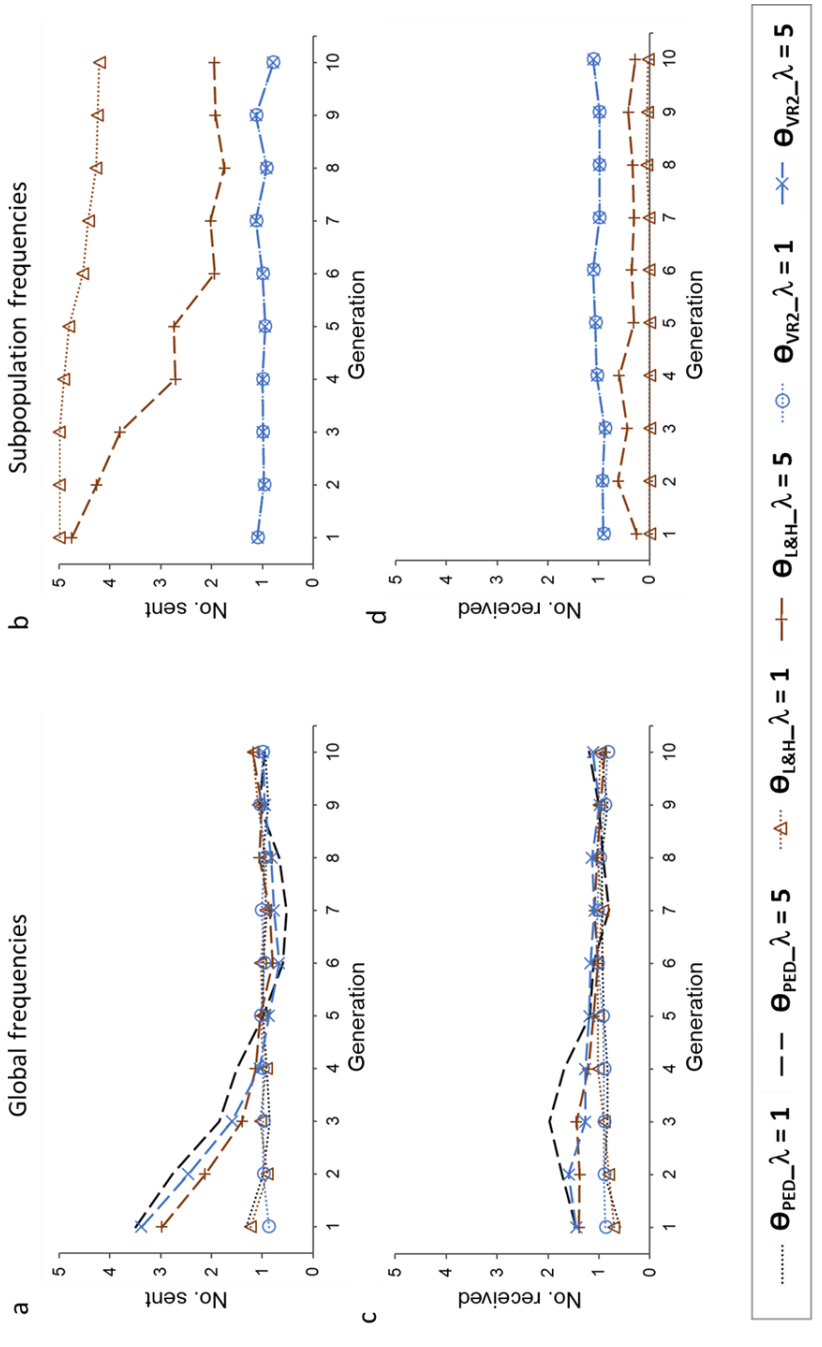

Figure 3. Number of individuals sent to and received by subpopulation 1 across generations when contributions are optimized using pedigree-based ($\boldsymbol{\theta}_{\mathbf{PED}}$), Li & Horvitz ($\boldsymbol{\theta}_{\mathbf{L\&H}}$) and VanRaden ($\boldsymbol{\theta}_{\mathbf{VR2}}$) coancestry matrices and two different weights are given to the within-subpopulation coancestry ($\lambda = 1$ and $\lambda = 5$), for scenarios with Unequal population structure. Matrices $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ were computed using global (left panels) or subpopulation initial allele frequencies (right panels).

Table 5. Average expected heterozygosity in the global population (H_T) and within (H_S) and between (D) subpopulations, and average inbreeding in subpopulation 1 (F_{S1}) and in subpopulations 2 to 5 ($F_{S2\text{-}5}$) across generations (t) when contributions are optimized using pedigree-based ($\boldsymbol{\theta}_{\mathbf{PED}}$), Li & Horvitz ($\boldsymbol{\theta}_{\mathbf{L\&H}}$) and VanRaden ($\boldsymbol{\theta}_{\mathbf{VR2}}$) coancestry matrices for two different weights given to the within-subpopulation coancestry (λ) for scenarios with Unequal population structure. Matrices $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ were computed using initial average subpopulation allele frequencies. Values in brackets are those deviated from results obtained when matrices $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ were computed using the initial allele frequencies in the global population (Tables 2 and 4), in percentage.[a]

		$\boldsymbol{\theta}_{\mathbf{L\&H}}$					$\boldsymbol{\theta}_{\mathbf{VR2}}$				
λ	t	H_T	H_S	D	F_{S1}	$F_{S2\text{-}5}$	H_T	H_S	D	F_{S1}	$F_{S2\text{-}5}$
1	0	0.188	0.180	0.008	0.825	0.811	0.188	0.180	0.008	0.825	0.811
		(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)
	1	0.189	0.176	0.013	0.825	0.809	0.187	0.175	0.012	0.828	0.813
		(+0.0)	(+0.0)	(+0.0)	(+0.1)	(–0.1)	(+0.0)	(–1.7)	(+33.3)	(+0.1)	(+0.1)
	5	0.190	0.162	0.027	0.840	0.825	0.182	0.160	0.022	0.836	0.829
		(+0.0)	(–3.0)	(+12.5)	(+1.2)	(+0.4)	(–2.2)	(–7.0)	(+69.2)	(+1.0)	(+1.2)
	10	0.191	0.148	0.042	0.854	0.841	0.177	0.150	0.027	0.844	0.840
		(+0.0)	(–8.6)	(+40.0)	(+2.5)	(+1.4)	(–3.3)	(–10.2)	(+68.8)	(+1.7)	(+1.9)
5	0	0.188	0.180	0.008	0.825	0.811	0.188	0.180	0.008	0.825	0.811
		(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)	(+0.0)
	1	0.188	0.179	0.009	0.826	0.810	0.187	0.175	0.012	0.828	0.813
		(–0.5)	(+0.0)	(+0.0)	(+0.4)	(+0.0)	(+0.0)	(–2.2)	(+50.0)	(+0.4)	(+0.1)
	5	0.188	0.177	0.011	0.821	0.814	0.182	0.160	0.022	0.836	0.829
		(–0.5)	(–0.6)	(+0.0)	(+0.7)	(+0.0)	(–1.6)	(–9.1)	(+120.0)	(+2.2)	(+1.5)
	10	0.188	0.176	0.012	0.825	0.815	0.177	0.150	0.027	0.844	0.840
		(–0.5)	(–0.6)	(+0.0)	(+1.2)	(+0.0)	(–3.3)	(–12.8)	(+170.0)	(+2.9)	(+2.3)

[a] Standard errors ranged from 1.00 x 10^{-4} to 6.69 x 10^{-4} for H_T and H_S and from 1.00 x 10^{-4} to 1.41 x 10^{-4} for F_{S1} and $F_{S2\text{-}5}$.

4. Discussion

This study has compared the use of different coancestry matrices (pedigree-based and genomic matrices) in the management of a subdivided population through OC

methodology in the framework of a conservation program. Comparisons were made in terms of the genetic diversity maintained in the total population, its distribution between and within-subpopulations, and in terms of the average global and within-subpopulation molecular inbreeding. Results showed that management based on matrices $\boldsymbol{\theta}_{\mathbf{PED}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ led to similar outcomes, and that the use of $\boldsymbol{\theta}_{\mathbf{L\&H}}$ led to higher global genetic diversity than the use of $\boldsymbol{\theta}_{\mathbf{PED}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$ for any weight given to subpopulation diversity (i.e., for any value of λ). Moreover, in scenarios where more weight was given to the within-subpopulation coancestry ($\lambda \geq 5$), the use of $\boldsymbol{\theta}_{\mathbf{L\&H}}$ also led to higher local genetic diversity, lower inbreeding levels and similar allelic diversity. Using local allele frequencies to construct the genomic coancestry matrices instead of the global frequencies implied, in general, lower genetic diversity and higher inbreeding.

Results for the different parameters have been given for the unobserved loci which are not used in calculating coancestry and, thus, they are not directly under selection. However, unobserved loci are in linkage disequilibrium with the markers used in the management and this disequilibrium drives changes in the same direction in both types of loci (de Cara *et al*. 2011; Gómez-Romano *et al*., 2013; **Chapter 2**; Toro *et al*., 2020; Woolliams & Meuwissen, 2022).

In population genetics and conservation biology studies using neutral molecular markers, genetic diversity is usually measured as expected heterozygosity (Nei, 1973) or as allelic diversity, i.e., the mean number of alleles per marker (Toro *et al*., 2009; Allendorf *et al*., 2013), which is equivalent to the percentage of segregating loci for biallelic markers. Most conservation studies dealing with managing and monitoring genetic diversity have focused on heterozygosity because high levels of heterozygosity also imply high levels of additive genetic variance and, thus, high potential responses to selection (Falconer & Mackay, 1996). In addition, the heterozygosity is inversely related to inbreeding and inbreeding depression. Consequently, the OC methodology has been directed to the maximization of the expected heterozygosity by using coancestries between candidates. However, allelic diversity is also a very relevant parameter in conservation genetics (Nei *et al*., 1975; Luikart *et al*., 1998; Vilas *et al*., 2015). Thus, both expected heterozygosity and allelic diversity should be evaluated and accounted for

in the management of populations. In addition, it can be argued that in a conservation program genetic variability should be preserved as closely as possible to that of the original population. Thus, management leading to changes in allele frequencies could be undesirable, especially in ex-situ conservation programs (e.g., Saura *et al.*, 2008; Toro *et al*., 2020).

Different genomic coancestry matrices have been described in the literature (e.g., Villanueva *et al*., 2021) including those used here ($\theta_{L\&H}$ and θ_{VR2}). One of the simplest genomic matrix is the matrix of Nejati-Javaremi *et al*. (1997) where the coancestry between two individuals is computed as the proportion of alleles shared by both individuals. This matrix (also called IBS matrix or similarity matrix) has been used in previous studies applying the OC method for managing genetic conservation programs (de Cara *et al*., 2011; 2013; Gómez-Romano *et al*., 2013; Eynard *et al*., 2016). Also, the software Metapop2 (López-Cortegano *et al*., 2019) implements such a matrix calculated from multiallelic markers for the management of subdivided populations. Although this matrix has not been considered here, it has a correlation of one with $\theta_{L\&H}$ (Villanueva et al. 2021) and therefore, the same results are expected from the use of both coancestry matrices.

In the context of subdivided populations, our study has showed that the use of $\theta_{L\&H}$ in OC was able to maintain higher expected heterozygosity and similar allelic diversity than θ_{VR2}. For undivided populations and using a reduced number of multiallelic markers in the management, Fernández *et al*. (2004) also found that OC was able to give higher expected heterozygosity and the same allelic diversity than a method specifically aimed at maximizing the latter. However, the use of matrices constructed from a large number of biallelic markers (SNPs) in OC, can lead to different outcomes as shown by Meuwissen *et al*. (2020) and **Chapter 2**. These studies compared the use of genomic $\theta_{L\&H}$ (a matrix with a correlation of one with the matrix used by Fernández *et al*., 2004) and θ_{VR2} in OC and showed that $\theta_{L\&H}$ led to higher levels of expected heterozygosity but to lower levels of allelic diversity than θ_{VR2}. In **Chapter 2** argued that, when the average H_e is maximized, the low allelic diversity found may be a consequence of the fact that rare alleles have little effect on H_e; i.e. many loci can be fixed without a reduction in H_e,

provided the remaining loci increase their MAF to get closer to intermediate values.

In order to understand the performance of OC in subdivided populations when using different matrices, we need to consider the fact that in this case global diversity is partitioned into within (H_S) and between subpopulations (D) diversity. From an exclusively theoretical point of view, the subdivision of populations can be beneficial for preserving global genetic diversity (Falconer & Mackay, 1996) as, in the long term, the highest diversity is achieved by maintaining many isolated lines hoping that, by drift, different alleles will be fixed in each of them. This is what happened when using $\boldsymbol{\theta_{L\&H}}$ and $\lambda = 1$ (i.e., no special weight was given to the within-subpopulation diversity) that led to a higher differentiation between subpopulations (higher D) and also the highest global expected heterozygosity. However, allelic diversity was lower than when using $\boldsymbol{\theta_{VR2}}$. The same results (higher expected heterozygosity and lower allelic diversity with $\boldsymbol{\theta_{L\&H}}$) were observed for undivided populations in studies comparing these two matrices (Meuwissen *et al*., 2020; **Chapter 2**). However, when λ was increased to 5, $\boldsymbol{\theta_{L\&H}}$ was able not only to give the highest heterozygosity but also to give levels of allelic diversity similar to those obtained with $\boldsymbol{\theta_{VR2}}$. At the global population level, the maintenance of the proportion of segregating loci is a reflection of the increased differentiation between subpopulations through the management that makes the fixation of the same allele in all subpopulations unlikely (i.e., for a particular locus, both alleles are likely to be kept in the global population).

Although theoretically subdivision may lead to the maintenance of higher levels of genetic diversity, a high degree of isolation implies higher levels of inbreeding in each subpopulation. The effect of inbreeding depression on fitness will result in an increased risk of extinction of a particular subpopulation, with a net loss of genetic diversity (Charlesworth & Willis, 2009). This problem can be tackled by using an increased weight on the maintenance of within-subpopulation diversity that, consequently, would reduce inbreeding. Doing so, here we have shown that it is possible to keep higher levels of both global and within-subpopulation diversity and lower levels of inbreeding by imposing $\lambda = 5$ when using $\boldsymbol{\theta_{L\&H}}$. Therefore, it seems that this strategy should be chosen when managing subdivided populations.

In some situations, keeping some degree of differentiation between subpopulations could be advantageous if the interest is to maintain their particular genetic singularity arising from, for example, local adaptations of each subpopulation. In this case, an explicit restriction on the minimum levels of *D*, F_{ST} (Wright´s fixation index) or any other measure of genetic differentiation could be imposed in the optimization either at the global level or at the subpopulations level, as suggested by Fernández *et al*. (2008). Actually, F_{ST} is related to coancestry through the expression $F_{ST} = (\tilde{f} - f)/(1 - f)$, where $\tilde{f}$ is the mean coancestry within subpopulations and f is the global coancestry (e.g., Caballero & Toro, 2002). Consequently, the restriction on a specific value for the differentiation between subpopulations could be perfectly integrated in the general framework of the OC methodology for subdivided populations.

Besides the main objective of preserving genetic diversity and avoiding inbreeding in conservation programs, it may also be desirable that certain characteristics previously selected naturally or artificially, are maintained under relaxed selection during the management period. For this reason, some authors have proposed to focus on maintaining allele frequencies as close as possible to those of the original population (e.g., Saura *et al*., 2008). The matrices used here for management ($\boldsymbol{\theta}_{\mathbf{PED}}$, $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$) influences the trajectory of the allele frequencies in undivided populations, as shown by Meuwissen *et al*. (2020) and in **Chapter 2**. In particular, they observed that matrices $\boldsymbol{\theta}_{\mathbf{PED}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$ led to smaller frequency changes than $\boldsymbol{\theta}_{\mathbf{L\&H}}$ so it can be argued that $\boldsymbol{\theta}_{\mathbf{L\&H}}$ is not suitable for conservation which aims to maintain the original allele frequencies. In this study, we have observed the same pattern for subdivided populations when genomic matrices were computed using global allele frequencies and no extra weight was applied to the within subpopulation coancestry.

The genomic matrices used here (i.e., $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{VR2}}$) depend on the allele frequencies in the base population. When the main objective of the conservation program is to maintain the global diversity of the population, it may be sensible to use the allele frequencies of the entire population to compute these matrices. However, when the main objective is to maintain the singularity of each subpopulation (i.e., when subdivision makes biological sense due to different adaptations or genetic characteristics), using the

initial local frequencies could be a better approach. Management based on $\theta_{L\&H}$ always results in allele frequency changes toward 0.5 and then using subpopulation or global frequencies led to very similar global heterozygosity for any value of λ. Also, very similar within- and between-subpopulations heterozygosities and inbreeding were obtained for $\lambda = 5$ (Table 5). However, with $\lambda = 1$, a higher genetic distance between subpopulations and a higher inbreeding were observed when using subpopulation frequencies, especially in the last generations where migration was reduced (Figure 3). This also happened in all scenarios with management based on θ_{VR2} (higher differentiation and inbreeding when using subpopulation frequencies). Management based on θ_{VR2} tends to reduce genetic drift and thus to maintain allele frequencies close to those at $t = 0$ (Meuwissen *et al*., 2020; **Chapter 2**). Because each subpopulation had different initial allele frequencies, the management using subpopulation frequencies would tend to reduce the flow between subpopulations. Consequently, differentiation among subpopulations and within-subpopulation inbreeding are higher than when using global frequencies, for any value of λ. Thus, the initial expectation of a better control of allelic frequencies deviation by using local (subpopulation) frequencies in the computation of θ_{VR2} was not observed in our results. Moreover, the rest of parameters tested (inbreeding, H_e and segregating loci) were also worse than when using the global initial frequencies. This is due to the fact that rare alleles within each subpopulation are more likely to be lost when using subpopulation frequencies due to a decreased effective population size and increased genetic drift. In fact, with θ_{VR2}, the average MAF decreased substantially more when using subpopulation frequencies and current frequencies moved away from the original frequencies even more than with $\theta_{L\&H}$ (Figure 2).

As indicated above, when a population has been subdivided, it is convenient to favor the gene flow between the different subpopulations to reduce the increase in inbreeding in each of them, as it has been claimed in the past (Falconer & Mackay 1996; Frankham *et al*., 2010). In the scenarios with no extra weight on subpopulation diversity (i.e., $\lambda = 1$) the mixture between subpopulations was carried out in a uniform way; that is, the same number of migrants on average was sent to and received (in this case one) by any subpopulation. This was true irrespective of the coancestry matrix used and even

for U scenarios, where subpopulation 1 was more inbred and more differentiated than the rest. In the latter scenarios and for $\lambda = 1$, it seems that there was an equilibrium between the need of reducing the high inbreeding of subpopulation 1 and the need of promoting the maintenance of the specific genetic diversity that it harbored. However, with $\lambda \geq 5$, reducing the high inbreeding of subpopulation 1 became the priority and, thus, the number of migrants directed to this subpopulation was initially higher. The migrant flow pattern was equal for any of the three coancestry matrices used in the OC management when the global initial frequencies were used to compute $\theta_{\mathbf{L\&H}}$ and $\theta_{\mathbf{VR2}}$. If subpopulation 1 were genetically different from the rest of the subpopulations but not so inbred, the pattern of migration would probably change with an initial tendency of moving individuals preferably from subpopulation 1 to the other subpopulations.

In this study we imposed a restriction on the maximum number of migrants allowed per generation (one migrant per subpopulation). Allowing a higher number of migrants would probably lead to lower inbreeding and to the homogenization of the genetic composition of all subpopulations in fewer generations. However, this may be not a realistic scenario due to the cost and risk of moving animals between subpopulations in some scenarios. Firstly, for many species it is an expensive procedure that also implies administrative burden. Moreover, the transportation of animals can cause them stress that can induce maladaptation to the new site and even an increased probability of dying along the way. In any case, Wang (2004) shown that relatively small migratory flows (of the order of one migrant per generation and subpopulation; i.e., the OMPG method) are sufficient to maintain levels of inbreeding at acceptable levels. Thus, in this study we limited to five (the number of subpopulations) the number of migrants per generation to make the present results comparable with OMPG as the classical management method applied before the development of OC. Fernández *et al*. (2008) showed that OC performs better than the OMPG method when relying on pedigree data, as higher levels of diversity were retained and lower levels of inbreeding were generated. Although the OMPG method has not been considered in the present study, we can state that molecular implementation of OC in subdivided populations performs better than OMPG given that the use of marker-based management has led to better results than pedigree-based

management in our simulations, and the latter outperform OMPG (Fernández *et al*. 2008) as said before. Nevertheless, the OC methodology is flexible and the number of migrants can be increased to the level that could be reasonable for each particular species and conservation program. Moreover, in the original derivation of the OMPG method, some degree of differentiation between subpopulations was intended to be maintained. If this is the case, the OC methodology could be easily modified to impose a restriction on the minimum value of the differentiation (measured as genetic distance or F_{ST}), as the objective function to be optimized implicitly includes the calculation of the genetic diversity between populations.

The most likely implementation of the OC method used here is in the management of ex-situ conservation programs that comprise different centers with captive animals. The chosen scenario in this study roughly mimics the real structure of 'The Iberian Lynx Ex situ Conservation Program' (https://www.lynxexsitu.es/programa-en.php)**.** This program involves five centers of similar capacity where the managers aim at having the same numbers of males and females at the different centers. Movements of animals between centers is limited for logistic reasons to levels comparable to those imposed in our simulations (i.e., one migrant per generation). Currently, the management (i.e., contributing parents, mating pattern and translocation of animals between centers) is designed following the principles stablished in this study but relying on pedigree information (Kleinman-Ruiz *et al*., 2019). Genomic resources have been recently developed for this species (Abascal *et al*., 2016; Kleinman-Ruiz *et al*., 2017) and the plan is to implement these resources not only in the ex-situ program but also in the in-situ program. Beyond the improvement in the management of captive animals from the use of genomic information as shown in this study (e.g. increased H_e and allelic diversity and decreased inbreeding), routine genotyping of captive and wild animals will improve the coordination between ex-situ and in-situ programs and will allow a more accurate management. With routine genotyping, information from an increased number of animals will be able to be included in the management, since it will be possible to estimate the relationships between wild and captive animals. This would allow a more precise control of movements of individuals between wild populations (i.e., translocations) to reorganize

the diversity and avoid the rise of inbreeding in particular areas. Genomic information would also help to drive the decisions on the breeding in captive populations (i.e., which animals to breed) accounting for the genetic information which is already present (or lacking) in wild populations in order to release the more adequate individuals. This scenario also applies to many other species. Therefore, the methodology used in this study could have a positive impact in these programs.

As a general conclusion, our results show that using matrix $\boldsymbol{\theta_{L\&H}}$ could be the best option for managing subdivided populations as it leads to higher global diversity and lower inbreeding. Moreover, the global allele frequencies should be used to compute the genomic coancestry matrices since higher levels of diversity and lower inbreeding are obtained than when using subpopulation frequencies.

Data Accessibility Statement

All software and scripts necessary to run the simulations are available on request.

Author contribution

EM-G: Software development, formal analysis, investigation, writing - original draft; BV: Supervision, project administration, funding acquisition, writing – review & editing; MAT: Investigation, writing – review & editing; JF: Study design, software development, supervision, project administration, funding acquisition, writing – review & editing.

Acknowledgments

This research was funded by MCIN/ AEI /10.13039/501100011033 (Project PID2020-114426GB-C22).

References

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M. *et al*. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 17, 251. https://doi.org/10.1186/s13059-016-1090-1.

Allendorf, F. W., Luikart, G. & Aitken, S. N. (2013). *Conservation and the genetics of populations*. John Wiley & Sons, Hoboken.

Amin, N., van Duijn, C. M. & Aulchenko Y. S. (2007). A genomic background based method for association analysis in related individuals. *PLoS ONE*, 2, e1274. https://doi.org/10.1371/journal.pone.0001274.

Ávila, V., Fernández, J., Quesada, H. & Caballero, A. (2011). An experimental evaluation with *Drosophila melanogaster* of a novel dynamic system for the management of subdivided populations in conservation programs. *Heredity*, 106, 765–774. https://doi.org/10.1038/hdy.2010.117.

Caballero, A., Rodríguez-Ramilo, S. T., Avila, V. & Fernández, J. (2010). Management of genetic diversity of subdivided populations in conservation programmes. *Conservation Genetics*, 11, 409–419. https://doi.org/10.1007/s10592-009-0020-0.

Caballero, A. & Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3, 289–299. https://doi.org/10.1023/A:1019956205473.

Charlesworth, D. & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10, 783–796. https://doi.org/10.1038/nrg2664.

de Cara, M.A.R., Fernández, J., Toro, M.A. & Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics*, 128, 456–464. https://doi.org/10.1111/j.1439-0388.2011.00971.x.

de Cara, M.A.R., Villanueva, B., Toro, M.A. & Fernández, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology*, 22, 6091−6099. https://doi.org/10.1111/mec.12560.

Eynard, S. E., Windig, J. J., Hiemstra, S. J. & Calus, M. P. (2016). Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genetics Selection Evolution*, 48, 33. https://doi.org/10.1186/s12711-016-0210-4.

Falconer, D.S. & Mackay, F.C. (1996). Introduction to Quantitative Genetics. 4th ed. Longman Group Ltd, Harlow, Essex, England.

FAO. (2013). *In vivo conservation of animal genetic resources*. FAO Animal Production and Health Guidelines. No. 14. Rome.

Fernández, J. & Caballero, A. (2001). Accumulation of deleterious mutations and equalization of parental contributions in the conservation of genetic resources. *Heredity*, 86, 480−488. https://doi.org/10.1046/j.1365-2540.2001.00851.x.

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. https://doi.org/10.1046/j.1439-0388.1999.00196.x

Fernández, J., Toro, M.A. & Caballero, A. (2003). Fixed contributions designs *vs*. minimization of global coancestry to control inbreeding in small populations. *Genetics*, 165, 885−894. https://doi.org/10.1093/genetics/165.2.885.

Fernández, J., Toro, M. A. & Caballero, A. (2004). Managing individuals' contributions to maximize the allelic diversity maintained in small, conserved populations. *Conservation Biology*, 18, 1358−1367. https://doi.org/10.1111/j.1523-1739.2004.00341.x.

Fernández, J., Toro, M. A. & Caballero, A. (2008). Management of subdivided populations in conservation programs: development of a novel dynamic system. *Genetics*, 179, 683−692. https://doi.org/10.1534/genetics.107.083816.

Frankham, R. (2008). Genetic adaptation to captivity in species conservation programs. *Molecular Ecology*, 17, 325−333. https://doi.org/10.1111/j.1365-

294X.2007.03399.x.

Frankham, R., Ballou, J. D. & Briscoe D. A. (2010). *Introduction to conservation genetics*. 2nd ed. Cambridge: Cambridge University Press, United Kingdom.

Gómez-Romano, F., Villanueva, B., de Cara, M.A.R & Fernández, J. (2013). Maintaining genetic diversity using molecular coancestry: The effect of marker density and effective population size. *Genetics Selection Evolution*, 45, 38. https://doi.org/10.1186/1297-9686-45-38.

Gómez-Romano, F.; Villanueva, B.; Fernández, J.; Woolliams, J.A.; & Pong-Wong, R. (2016). The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution*, 48, 2. https://doi.org/10.1186/s12711-015-0172-y.

IUCN (2022). Guidelines for Using the IUCN Red List Categories and Criteria. Version 15.1. Prepared by the *Standards and Petitions Committee*. IUCN, Gland, Switzerland and Cambridge, U.K. Downloadable from https://www.iucnredlist.org/documents/RedListGuidelines.pdf.

Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., Fernández, J. & Godoy, J. A. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, 18, 556. https://doi.org/10.1186/s12864-017-3946-5.

Kleinman-Ruiz, D., Soriano, L., Casas-Marce, M., Szychta, C., Sánchez, I., Fernández, J. & Godoy, J. A. (2019). Genetic evaluation of the Iberian lynx ex situ conservation programme. *Heredity*, 123, 647–661. https://doi.org/10.1038/s41437-019-0217-z.

Lacy, R. C. (2000). Should we select genetic alleles in our conservation breeding programs? *Zoo Biology*, 19, 279–282. https://doi.org/10.1002/1098-2361(2000)19:4<279::AID-ZOO5>3.0.CO;2-V.

Li, C.C. & Horvitz, D.G. (1953). Some methods of estimating the inbreeding coefficient. *American journal of human genetics*, 5, 107–117.

López-Cortegano, E., Pérez-Figueroa, A. & Caballero, A. (2019). Metapop2: Re-implementation of software for the analysis and management of subdivided populations using gene and allelic diversity. *Molecular Ecology Resources*, 19, 1095–1100. https://doi.org/10.1111/1755-0998.13015

Luikart, G., Allendorf, F. W., Cornuet, J. M. & Sherwin, W. B. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, 89, 238–247. https://doi.org/10.1093/jhered/89.3.238.

Meuwissen, T. H. E., Sonesson, A. K., Gebregiwergis, G. & Woolliams, J. A. (2020). Management of genetic diversity in the era of genomics. *Frontiers in Genetics*, 11, 880. https://doi.org/10.3389/fgene.2020.00880.

Mills, L. S. & Allendorf, F. W. (1996). The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, 10, 1509–1518. https://www.jstor.org/stable/2387022.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences U.S.A.*, 70, 3321–3323. https://doi.org/10.1073/pnas.70.12.3321.

Nei, M., Maruyama, T. & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 29, 1–10. https://doi.org/10.2307/2407137.

Nejati-Javaremi, A., Smith, C. & Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science*, 75, 1738–1745. https://doi.org/10.2527/1997.7571738x.

Saura, M., Pérez-Figueroa, A., Fernández, J., Toro, M. A. & Caballero, A. (2008). Preserving population allele frequencies in ex situ conservation programs. *Conservation Biology*, 22, 1277–1287. https://doi.org/10.1111/j.1523-1739.2008.00992.x.

Schoen, D. J., David, J. L. & Bataillon, T.M. (1998). Deleterious mutation accumulation and the regeneration of genetic resources. *Proceedings of the National Academy of Sciences U.S.A.*, 95, 394–399.

Toro, M.A., Barragán, C., Óvilo, C., Rodrigáñez, J., Rodríguez, C. & Silió, L. (2002). Estimation of coancestry in Iberian pigs using molecular markers. *Conservation Genetic*, 3, 309–320. https://doi.org/10.1023/A:1019921131171.

Toro, M. A. & Caballero, A. (2005). Characterization and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360, 1367–1378. https://doi.org/10.1098/rstb.2005.1680.

Toro, M.A., Fernández, J. & Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science*, 120, 174–195. https://doi.org/10.1016/j.livsci.2008.07.003.

Toro, M. A., Villanueva, B. & Fernández, J. (2020). The concept of effective population size loses its meaning in the context of optimal management of diversity using molecular markers. *Journal of Animal Breeding and Genetics*, 137, 345–355. https://doi.org/10.1111/jbg.12455.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

Vilas, A., Pérez-Figueroa, A., Quesada, H. & Caballero, A. (2015). Allelic diversity for neutral markers retains a higher adaptive potential for quantitative traits than expected heterozygosity. *Molecular Ecology*, 24, 4419–4432. https://doi.org/10.1111/mec.13334.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E. *et al*. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics Selection Evolution*, 53, 42. https://doi.org/10.1186/s12711-021-00635-0.

Villanueva, B., Pong-Wong, R., Woolliams, J.A. & Avendaño, S. (2004). Managing genetic resources in selected and conserved populations. In: Simm, G., Villanueva, B., Sinclair, K.D., Townsend, S. (eds.). *Farm Animal Genetic Resources*. BSAS Nottingham University Press, Nottingham, United Kingdom.

113–132.

Wang, J. (2004). Application of the one-migrant-per-generation rule in conservation and management. *Conservation Biology*, 18, 332–343. https://doi.org/10.1111/j.1523-1739.2004.00440.x.

Woolliams, J. A. & Meuwissen, T.H.E. (2022) Genetic management meets genomics. In *Proceedings of the 12th World Congress of Genetics Applied to Livestock Production*, Rotterdam, Netherlands; Available online: https://www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/61_015.pdf (accessed on 27 July 2022).

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159. https://doi.org/10.1093/genetics/16.2.97.

GENERAL DISCUSSION

The main objective of this thesis was to evaluate the efficiency of using genomic coancestry matrices in the management of populations aimed at minimizing the loss of genetic diversity, and also to evaluate the effect of such management in the evolution of allele frequencies. To complete this goal, in **Chapter 1** we made use of turbot genotypes to compare six different genomic coancestry matrices in terms of their efficiency in retaining genetic diversity (measured as expected heterozygosity), when implementing the OC method in a single generation. The matrices compared were those based on: i) the proportion of alleles shared by two individuals ($\boldsymbol{\theta_{SIM}}$); ii) deviations of the observed number of alleles shared by two individuals from the expected number under Hardy-Weinberg equilibrium ($\boldsymbol{\theta_{L\&H}}$); iii) the realized relationship matrix obtained by VanRaden's method 1 ($\boldsymbol{\theta_{VR1}}$); iv) the realized relationship matrix obtained by VanRaden's method 2 ($\boldsymbol{\theta_{VR2}}$); v) the realized relationship matrix obtained by Yang´s method ($\boldsymbol{\theta_{YAN}}$); and vi) IBD segments ($\boldsymbol{\theta_{SEG}}$). In **Chapter 2**, we used computer simulations to compare $\boldsymbol{\theta_{L\&H}}$ and $\boldsymbol{\theta_{VR2}}$ when used in OC, in the short and long term. Populations were simulated for 50 generations and comparisons were performed not only in terms of genetic diversity measured as expected heterozygosity but also in terms of genetic diversity measured as allelic diversity, and changes in allele frequencies. The reason to choose these two matrices was to test the hypothesis of Gómez-Romano *et al.* (2016b) who suggested that while OC using $\boldsymbol{\theta_{L\&H}}$ (or $\boldsymbol{\theta_{SIM}}$ as both matrices correlate perfectly) favors solutions that tend to move allele frequencies towards 0.5 and, therefore, to increase genetic diversity, OC using realized relationship matrices would lead to solutions that tend to keep allele frequencies closer to those in the original population (i.e., allele frequencies would tend to be unchanged). Matrix $\boldsymbol{\theta_{VR2}}$ was chosen over $\boldsymbol{\theta_{VR1}}$ and $\boldsymbol{\theta_{YAN}}$ because it led to more extreme results in **Chapter 1**. Matrix $\boldsymbol{\theta_{SEG}}$ was not considered in **Chapters 2** and **3** because results using this matrix greatly depend on the choice of the parameters to define a segment (e.g., minimum SNP density required, maximum distance allowed between two consecutive segments, and minimum number of SNPs in a segment). In **Chapter 3** the simulations performed in **Chapter 2** were extended to subdivided populations. As in the previous chapters, the evaluation of the genomic matrices was carried out in terms of the genetic diversity maintained when used

in OC. However, in this case, genetic diversity (measured either as expected heterozygosity or as allelic diversity) was evaluated at the global population level, and also between and within subpopulations. Also, the average global molecular inbreeding and the inbreeding within each subpopulation, the migration flow between subpopulations and the trajectory of allele frequencies were compared across strategies using both matrices.

Our results evidenced large differences in the magnitude of the six different coancestry coefficients compared in **Chapter 1**. Large differences have been also found in previous studies comparing genomic measures of coancestry (Rodríguez-Ramilo *et al*., 2015; Eynard *et al*., 2016; Gómez-Romano *et al*., 2016b). In particular, f_{SIM} was much larger than other coancestry coefficients. This was expected given that f_{SIM} simply measures observed similarity and does not distinguish between IBS and IBD while frequency-based coefficients ($f_{L\&H}$, f_{VR1}, f_{VR2}, f_{YAN} and $f_{L\&H}$) and f_{SEG} attempt to measure IBD (Gusev *et al*., 2009; de Cara *et al*., 2013; Toro *et al*., 2014;). Although there were large differences in the magnitude of the different coancestry coefficients, the pairwise correlations between them were relatively high (Figure 4 in **Chapter 1**). One of the lowest correlations (0.72) was that between $f_{L\&H}$ and f_{VR2} (i.e., the coefficients investigated in detail in **Chapters 2** and **3**) so the management using $\boldsymbol{\theta_{L\&H}}$ and $\boldsymbol{\theta_{VR2}}$ is expected to lead to different results as shown in **Chapters 2** and **3**.

When allele frequencies in the base population (assumed to be constituted by non-inbred and unrelated individuals) are known, frequency-based coefficients are expected to provide unbiased estimates of the average IBD measures of relatedness relative to the base population (VanRaden, 2008; Caballero *et al*., 2022). In **Chapter 1**, only genotype data from two generations (parents and offspring) were available and thus the base population considered was the parental generation. This was the reason for the four average global frequency-based coancestries being close to zero. This was also the case for $f_{L\&H}$ and f_{VR2} in **Chapter 2** in the initial generations but not in later generations as IBD developed.

Commercial SNP arrays are usually designed with an underrepresentation of SNPs with extreme allele frequencies (i.e., SNPs with low MAF are usually filtered out).

This increases the power for detecting SNP effects on traits of interest, but it is a suboptimal approach when the purpose is to estimate genetic diversity because the ascertainment bias resulting from the selection of SNPs to be included in the array can led to an overestimation of both observed and expected heterozygosity (Geibel *et al*., 2021). In contrast, whole genome data provides more complete genetic information of individuals, including rare variants that are not fully covered by SNP arrays. Using computer simulations, Pérez-Enciso (2014) found that although, as expected, sequence data result in the most accurate estimates of genetic relationships (provided enough coverage is obtained), high-density genotyping can also result in highly accurate estimates. However, in the context of OC management aimed at maximizing diversity, Eynard *et al*. (2015) showed considerable losses of genetic diversity when using SNP arrays data rather than sequences in Holstein cattle. They observed a uniform distribution of MAF for SNP variants and a L shaped distribution for sequence variants (see Figure 1 in Eynard *et al*., 2015). Thus, all classes of MAF were equally represented on the array, while low MAF classes were overrepresented in the sequence data. In **Chapter 1**, genotypes for 18,097 SNPs for a turbot population were used to compute different coancestry matrices. Genotypes were obtained by genotyping-by-sequencing using a 2b-RAD-sequencing approach and the MAF distribution actually had a L shape (see Figure 2 in **Chapter 1**). In **Chapters 2** and **3**, where SNP genotypes were simulated, the MAF distributions for the SNP also followed a L shape**.** Thus, we can expect that the heterozygosities presented in this thesis are not overestimated.

Using arrays with an underrepresentation of SNPs with extreme allele frequencies when managing populations through OC can also have an effect on the potential maximum genetic diversity maintained. In **Chapter 2**, it was showed that a substantial amount of expected heterozygosity was lost across generations when only SNPs with MAF above a particular threshold were used to compute $\theta_{\mathbf{L\&H}}$ (see Figure 3 in **Chapter2**). However, the opposite occurred when using $\theta_{\mathbf{VR2}}$ (i.e., higher heterozygosity was observed when rare alleles were discarded). Management using $\theta_{\mathbf{VR2}}$ give more emphasis to rare alleles and, thus, when they are removed, the method can focus on increasing diversity and not on conserving those rare alleles. In contrast,

management using $\boldsymbol{\theta_{L\&H}}$ loses the opportunity of detecting those alleles and increasing their frequencies, so it achieves lower levels of H_e. Thus, management after removing SNPs with low MAF leads to more similar results using both matrices than when rare alleles are not discarded (**Chapter 2**). These observations are in agreement with results from Villanueva *et al*. (2021), who found that the correlation between inbreeding coefficients $F_{L\&H}$ and F_{VR2} increases with increasing MAF of the SNPs used.

In an undivided population, the use of $\boldsymbol{\theta_{L\&H}}$ in OC led to higher levels of H_e than the use of $\boldsymbol{\theta_{VR2}}$ (see Table 3 in **Chapter 1** and Table 1 in **Chapter 2**) by moving allele frequencies to intermediate values. These frequency changes may be undesirable for example if the population has a particular adaptation to its environment. Actual frequencies in highly threatened populations may be a consequence of adaptation or simply genetic drift. Thus, in order to decide which is the most appropriate matrix to be used in the management, some knowledge on the historical evolution of the population is needed. Evidence of natural selection and adaptation to different environmental variables should be available if the objective chosen is to maintain the initial frequencies. Adaptation is particularly evident in the case of clines, which are directional patterns of phenotypic or genetic change across environmental gradients. Genetic latitudinal clines have been widely studied in *Drosophila melanogaster* (Oakeshott *et al*., 1982; Berry & Kreitman, 1993; Umina *et al*., 2005; Durmaz *et al*., 2019), but also in other species (Sotka *et al*., 2004; Rosenberg *et al*., 2005; Lehnert *et al*., 2018). However, in species under conservation programs this is difficult to verify because usually there are not enough locations to find a pattern in the clines. Consequently, in the absence of a clear evidence of a local adaptation, the practical approach could be to maximize diversity to give the population the ability of responding to natural or artificial selection and the possibility of adapting to a wide range of scenarios. In this case, therefore, the chosen strategy would be to use the $\boldsymbol{\theta_{L\&H}}$ matrix for OC management.

Interestingly, in **Chapter 3** we showed that the conflict with $\boldsymbol{\theta_{L\&H}}$ (higher H_e but higher frequency changes and lower allelic diversity) can be not a problem in subdivided populations. By increasing the weight given to the within-subpopulation coancestry we can still achieve the highest global and within subpopulations heterozygosity (and, thus,

the lowest inbreeding) while maintaining allele frequencies close to those of the base population. Moreover, under this scenario, the levels of allelic diversity (i.e., number of segregating alleles) were similar with $\theta_{L\&H}$ and θ_{VR2}. In order to achieve a balance between both objectives (i.e., large H_e with minimal frequency change) in an undivided population, an OC strategy using $\theta_{L\&H}$ to maximize genetic diversity could be applied but imposing a constraint on the magnitude of the change in allele frequencies expected for each feasible solution, as proposed by Saura *et al.* (2008).

An additional advantage of using genomic coancestries lies in the fact that they can be compute for particular regions of the genome, contrarily to what happen to genealogical coancestries, which provide expectations for the whole genome. Consequently, genomic information allows us to focus the management on specific regions (Roughsedge *et al*. 2008; Gómez-Romano *et al*. 2016a). This approach could be used (in both undivided and subdivided populations) when the objective is to maximize genetic diversity in certain regions of the genome. For example, there are regions that have accumulated more inbreeding than the rest of the genome for different reasons (e.g., selection). In these regions, it is especially important to minimize further losses of genetic diversity. Regions associated with inbreeding depression for fitness related traits also deserves a special treatment to avoid the generation of inbreeding. Another example refers to regions that harbor loci involved in general resistance to disease (e.g., the major histocompatibility complex, MHC) where a high level of genetic diversity is desirable to ensure that the population can deal with potential new disease challenges. Gomez-Romano *et al*. (2016a) showed, through computer simulations, that the OC method, using semi-definite programming and θ_{SIM} calculated with SNPs mapped to specific regions, is efficient in maintaining (and even increasing) heterozygosity in those regions while imposing a restriction on the increase in inbreeding in the rest of the genome. However, in other regions of the genome, alleles could be at frequencies determined by adaptation to the environment and could be desirable to keep them constant. The same approach (Gomez-Romano *et al*., 2016a) could be used to impose a restriction on the change in allele frequencies rather than on inbreeding using the formulations of Fernández *et al.* (2006) and Saura *et al*. (2008). Another option to achieve the same objective could be to

compute $\theta_{L\&H}$ with SNPs mapped to specific regions where the objective is to maximize genetic diversity and θ_{VR2} with SNPs mapped to specific regions where the objective is to maintain allele frequencies, and combine both matrices in the optimization.

When designing the conservation program for a particular population that is split in two or more genetically differentiated subpopulations, an important initial decision to make is whether the subpopulations should be kept isolated (for example if they harbor local adaptations) or, on the contrary, should be mixed. If the decision is the former, a separate program with OC management as described in **Chapters 1** and **2**, should be established for each subpopulation. Even in the cases where the decision is to carry out a joint management of the whole population, it may be still advisable to keep different centers (i.e., subpopulations) for logistic reasons (see **Chapter 3**). Under this scenario, the use of the OC method is again the best option for the maintenance of genetic diversity (as in undivided populations) and is implementation is straightforward as described by Fernández et al. (2008) and extended in **Chapter 3**. A clear example of this scenario is the case of the *ex-situ* conservation program of *Iberian lynx*. When the program began, animals were distributed in only two remnant isolated populations located in Sierra Morena (Andalusia, Spain): Doñana and Andújar (Abascal *et al*., 2016; Kleinman-Ruiz *et al*., 2019). A study using mitochondrial sequences and microsatellite markers documented low genetic diversity, high inbreeding levels, and high genetic differentiation between both subpopulations (Casas-Marce *et al*., 2013). However, these genetic patterns were the result of the recent decline and fragmentation of the global lynx population, and the differentiation between sites where not the results of local adaptation but to drift (Peña *et al*., 2006; Jiménez *et al*., 2008; Palomares *et al*., 2012; Ruiz-López *et al*., 2012). Therefore, the conservation strategy recommended was to establish an integrated genetic management of both subpopulations, including some flow between them. In fact, the captive populations was founded with animals of both remnant subpopulations in the wild and distributed in five breeding centers. This situation is ideal for the management of the global population applying the OC method for subdivided populations. At the beginning of the program only pedigree data were available (Kleinman-Ruiz *et al*., 2019) so at that time the method of Fernández *et al* (2008) was

used. However, genomic resources have been developed for this species (Abascal *et al*., 2016; Kleinman-Ruiz *et al*., 2017) and the plan is to use them not only in the *ex-situ* program but also in the *in-situ* program. Therefore, it is possible to begin applying the strategy proposed in **Chapter 3**.

In **Chapter 3**, the genetic distance between subpopulations was expressed as the Nei's minimum genetic distance (*D*). Another option would be to use of the Wright's fixation index (F_{ST}) as suggested in previous studies (Meirmans & Hedrick, 2011; Whitlock, 2011; Wang, 2012). However, both F_{ST} and *D* are directly related given that $F_{ST} = D/H_T$ (Wright, 1969; Caballero & Toro 2002; Caballero *et al*., 2010). Thus, the trend in the differentiation between subpopulations (i.e., if this differentiation increases or decreases) and the conclusions from this chapter are expected to be the same using *D* or F_{ST}. We believe that F_{ST} would be a better measure when comparing two populations with different H_T. However, in our study, rather than populations we compared management strategies for a given population and, thus, comparisons were carried out at the same level of H_T.

In most conservation programs the limiting factor is the reduced budget available. Thus, the implementation of molecular tools is somehow compromised if the advantages from the use of these tools (i.e., higher diversity maintained and/or lower inbreeding generated) do not compensate the extra costs. In this thesis, direct comparisons between pedigree and molecular management were only made for the subdivided population scenario (**Chapter 3**). We showed that the results when using $\boldsymbol{\theta_{PED}}$ and $\boldsymbol{\theta_{VR2}}$ in OC were very similar and that the differences in absolute values when using $\boldsymbol{\theta_{PED}}$ and $\boldsymbol{\theta_{L\&H}}$ were small. These small differences in between $\boldsymbol{\theta_{PED}}$ and $\boldsymbol{\theta_{L\&H}}$ (for example, differences in observed homozygosity, H_o, were between $1.0\text{x}10^{-3}$ and $6.0\text{x}10^{-3}$) may cast some doubts about the convenience of genotyping. However, if comparisons are made relative to the standard deviation of H_o then there is a reduction of around two standard deviations when using $\boldsymbol{\theta_{PED}}$. Therefore, the benefits obtained with $\boldsymbol{\theta_{L\&H}}$ are really relevant and genotyping may be worthy.

The results presented in this thesis could be also valuable for taking the right

decisions when creating gene banks. Gene banks, in the form of reproductive material (sperm, ova and embryos), act as reservoirs of the genetic diversity of living populations and thus they can provide a valuable tool for reducing the risks that animal genetic resources are facing (Johnston & Lacy, 1995; Frankham *et al*., 2010; FAO 2012; Eynard *et al*., 2018). Gene banks can be used to complement management strategies applied to these populations and also provide insurance against particular risks leading to the loss of genetic diversity such as diseases outbreaks or natural disasters. The information contained in the banks also represent useful material for conducting research. However, gene banks are expensive and require a strong commitment in both their foundation and maintenance and thus they must be rigorously planned, starting off with a clear definition of the objectives of the bank.

When the objective of the bank is to store as much genetic diversity as possible, Engelsma *et al.* (2011) showed, using Holstein-Friesian cattle data that the preferred strategy is to apply the OC method using a genomic relationship matrix. The sampling material determined by OC methodology based on genomic data resulted in a higher conserved diversity than when OC was based on pedigree data, although differences were small. The genomic matrix they used was θ_{SIM} which is based on the proportion of alleles shared by two individuals (Nejati-Javaremi *et al*., 1997). The same results could be expected when using $\theta_{L\&H}$ given that the correlation between both matrices is one.

However, it could be argued that rather than maximizing diversity in the bank, the preferred objective would be to store all alleles that are segregating in the population. This was a relevant issue, for example, for the semen banks constructed in different countries to alleviate the risks inherent to national scrapie eradication programs (Dawson *et al*., 1998; Brandsma *et al*., 2004; Drögemfller *et al*., 2004; Barillet *et al*., 2002). These programs were based on selection for increased frequency of the resistant ARR allele and removal of the most susceptible allele; i.e., the VRQ. Roughsedge *et al*. (2006) identified three primary risks when carrying out such eradication programs: i) the possibility that a new disease, more threatening than scrapie, appears and the allele being removed from sheep populations may be the allele conferring resistance; ii) the loss of favorable attributes from sheep populations if there is a unfavorable association between the

resistant PrP allele variant and other important traits in sheep production; and iii) the potential genetic bottleneck caused by limiting selection for breeding to animals of particular genotypes. Given these risks, semen banks were constructed to archive the alleles being eradicated in order to have the potential for restoring them if needed in the future (Fernández *et al*., 2006; Roughsedge *et al*., 2006). Thus, in this case, when creating the bank based on OC methodology, the use of θ_{VR2} could be more desirable than the use of $\theta_{L\&H}$ given that θ_{VR2} tends to maintain allele frequencies closer to the original values better than $\theta_{L\&H}$.

This thesis has focused on the optimization of contributions through the OC method aimed exclusively at maximizing genetic diversity of populations and thus, it has focused on conservation programs. However, the OC method can be also used in the context of genetic improvement programs where the objective is maximizing genetic gain while restricting the increase of inbreeding. A lot of research was carried out in the past on OC theory for selection programs using pedigree data to compute genetic relationships (see review by Woolliams *et al*., 2015) but few studies have dealt with genomic data. Recently, Meuwissen *et al*. (2020) compared different genomic coancestry matrices used in OC and observed that $\theta_{L\&H}$ maintained more genetic diversity and θ_{VR2} maintained allele frequencies closer to the original values and higher genetic gain. However, more research is needed for fully understand the performance of the different matrices when selection is applied. Another interesting area of future research would be to evaluate the different matrices when using OC in selected populations that are subdivided.

References

Abascal, F., Corvelo, A., Cruz, F., Villanueva-Cañas, J. L., Vlasova, A., Marcet-Houben, M. *et al*. (2016). Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology*, 17, 251. https://doi.org/10.1186/s13059-016-1090-1.

Barillet, F., Andréoletti, O., Palhiere, I., Aguerre, X., Arranz, J. M., Minery, S. *et al*. (2002, August). Breeding for scrapie resistance using PrP genotyping in the

French dairy sheep breeds. In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*. 31, 683–6.

Berry, A. & Kreitman, M. (1993). Molecular analysis of an allozyme cline: alcohol dehydrogenase in Drosophila melanogaster on the east coast of North America. *Genetics*, 134, 869–893. https://doi.org/10.1093/genetics/134.3.869.

Brandsma, J. H., Janss, L. L. G. & Visscher, A. H. (2004). Association between PrP genotypes and littersize and 135 days weight in Texel sheep. *Livestock Production Science*, 85, 59–64. https://doi.org/10.1016/S0301-6226(03)00116-7.

Caballero, A., Fernández, A., Villanueva, B. & Toro, M. A. (2022). A comparison of marker-based estimators of inbreeding and inbreeding depression. *Genetics Selection Evolution*, 54, 82. https://doi.org/10.1186/s12711-022-00772-0.

Caballero, A., Rodríguez-Ramilo, S. T., Avila, V. & Fernández, J. (2010). Management of genetic diversity of subdivided populations in conservation programmes. *Conservation Genetics*, 11, 409–419. https://doi.org/10.1007/s10592-009-0020-0.

Caballero, A. & Toro, M. A. (2002). Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics*, 3, 289–299. https://doi.org/10.1023/A:1019956205473.

Casas-Marce, M., Soriano, L., López-Bao, J. V. & Godoy, J. A. (2013). Genetics at the verge of extinction: insights from the Iberian lynx. *Molecular Ecology*, 22, 5503–5515. https://doi.org/10.1111/mec.12498.

Dawson, M., Hoinville, L. J., Hosie, B. D. & Hunter, N. (1998). Guidance on the use of PrP genotyping as an aid to the control of clinical scrapie. *Veterinary Record*, 142, 623-625.

de Cara, M.A.R., Villanueva, B., Toro, M.A. & Fernández, J. (2013). Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology*, 22, 6091–6099. https://doi.org/10.1111/mec.12560.

Durmaz, E., Rajpurohit, S., Betancourt, N., Fabian, D. K., Kapun, M., Schmidt, P. & Flatt, T. (2019). A clinal polymorphism in the insulin signaling transcription factor *foxo* contributes to life-history adaptation in *Drosophila*. *Evolution*, 73, 1774–1792. https://doi.org/10.1111/evo.13759.

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. https://doi.org/10.1111/j.1439-0388.2011.00936.x.

Eynard, S. E., Windig, J. J., Hiemstra, S. J. & Calus, M. P. (2016). Whole-genome sequence data uncover loss of genetic diversity due to selection. *Genetics Selection Evolution*, 48, 33. https://doi.org/10.1186/s12711-016-0210-4.

Eynard, S. E., Windig, J. J., Hulsegge, I., Hiemstra, S. J. & Calus, M. P. (2018). The impact of using old germplasm on genetic merit and diversity—A cattle breed case study. *Journal of Animal Breeding and Genetics*, 135, 311–322. https://doi.org/10.1111/jbg.12333.

Eynard, S. E., Windig, J. J., Leroy, G., van Binsbergen, R. & Calus, M. P. (2015). The effect of rare alleles on estimated genomic relationships from whole genome sequence data. *BMC Genetics*, 16:24. https://doi.org/10.1186/s12863-015-0185-0.

FAO. (2012). Cryoconservation of animal genetic resources. FAO Commission on genetic resources for food and agriculture assessments, Rome, Italy.

Fernández, J., Roughsedge, T., Woolliams, J. A. & Villanueva, B. (2006). Optimization of the sampling strategy for establishing a gene bank: storing PrP alleles following a scrapie eradication plan as a case study. *Animal Science*, 82, 813–821. https://doi.org/10.1017/ASC2006101.

Fernández, J., Toro, M. A. & Caballero, A. (2008). Management of subdivided populations in conservation programs: development of a novel dynamic system. *Genetics*, 179, 683-692. https://doi.org/10.1534/genetics.107.083816.

Frankham, R., Ballou, J. D. & Briscoe D. A. (2010). *Introduction to conservation genetics*. 2nd ed. Cambridge: Cambridge University Press, United Kingdom.

Geibel, J., Reimer, C., Weigend, S., Weigend, A., Pook, T. & Simianer, H. (2021). How array design creates SNP ascertainment bias. *PLoS ONE*, 16, e0245178. https://doi.org/10.1371/journal.pone.0245178.

Gómez-Romano, F.; Villanueva, B.; Fernández, J.; Woolliams, J.A.; & Pong-Wong, R. (2016a). The use of genomic coancestry matrices in the optimisation of contributions to maintain genetic diversity at specific regions of the genome. *Genetics Selection Evolution*, 48, 2. https://doi.org/10.1186/s12711-015-0172-y.

Gómez-Romano, F., Villanueva, B., Sölkner, J., de Cara, M. Á. R., Mészáros, G., Pérez O'Brien, A. M. & Fernández, J. (2016b). The use of coancestry based on shared segments for maintaining genetic diversity. *Journal of Animal Breeding and Genetics*, 133, 357–365. https://doi.org/10.1111/jbg.12213.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L. *et al*. (2009). Whole population, genome-wide mapping of hidden relatedness. *Genome research*, 19, 318–326. https://doi.org/10.1101/gr.081398.108.

Jiménez, M. Á., Sánchez, B., Alenza, M. D. P., García, P., López, J. V., Rodriguez, A. *et al*. (2008). Membranous glomerulonephritis in the Iberian lynx (*Lynx pardinus*). *Veterinary Immunology and Immunopathology*, 121, 34–43. https://doi.org/10.1016/j.vetimm.2007.07.018.

Johnston, L. A. & Lacy, R. C. (1995). Genome resource banking for species conservation: selection of sperm donors. *Cryobiology*, 32, 68–77. https://doi.org/10.1006/cryo.1995.1006.

Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B. *et al*. (2017). Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. *BMC Genomics*, 18, 556. https://doi.org/10.1186/s12864-017-3946-5.

Kleinman-Ruiz, D., Soriano, L., Casas-Marce, M., Szychta, C., Sánchez, I., Fernández,

J. & Godoy, J. A. (2019). Genetic evaluation of the Iberian lynx ex situ conservation programme. *Heredity*, 123, 647–661. https://doi.org/10.1038/s41437-019-0217-z.

Lehnert, S. J., DiBacco, C., Jeffery, N. W., Blakeslee, A. M., Isaksson, J., Roman *et al*. (2018). Temporal dynamics of genetic clines of invasive European green crab (*Carcinus maenas*) in eastern North America. *Evolutionary Applications*, 11, 1656–1670. https://doi.org/10.1111/eva.12657.

Meirmans, P. G. & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, 11, 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x.

Meuwissen, T. H. E., Sonesson, A. K., Gebregiwergis, G. & Woolliams, J. A. (2020). Management of genetic diversity in the era of genomics. *Frontiers in Genetics*, 11, 880. https://doi.org/10.3389/fgene.2020.00880.

Nejati-Javaremi, A., Smith, C. & Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *Journal of Animal Science*, 75, 1738–1745. https://doi.org/10.2527/1997.7571738x.

Oakeshott, J. G., Gibson, J. B., Anderson, P. R., Knibb, W. R., Anderson, D. G. & Chambers, G. K. (1982). Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution*, 36, 86–96. https://doi.org/10.2307/2407970.

Palomares, F., Godoy, J. A., López-Bao, J. V., Rodríguez, A., Roques, S., Casas-Marce, M. *et al*. (2012). Possible extinction vortex for a population of Iberian lynx on the verge of extirpation. *Conservation Biology*, 26, 689–697. https://doi.org/10.1111/j.1523-1739.2012.01870.x.

Pérez-Enciso, M. (2014). Genomic relationships computed from either next-generation sequence or array SNP data. *Journal of Animal Breeding and Genetics*, 131, 85–96. https://doi.org/10.1111/jbg.12074.

Peña, L., Garcia, P., Jiménez, M. Á., Benito, A., Alenza, M. D. P. & Sánchez, B. (2006).

Histopathological and immunohistochemical findings in lymphoid tissues of the endangered Iberian lynx (*Lynx pardinus*). *Comparative Immunology, Microbiology and Infectious Diseases*, 29, 114–126. https://doi.org/10.1016/j.cimid.2006.01.003.

Rodríguez-Ramilo, S.T., Fernández, J., Toro, M.A., Hernández, D., Villanueva, B. (2015). Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in the Spanish Holstein Population. *PLoS ONE,* 10, e0124157. https://doi.org/10.1371/journal.pone.0124157.

Roughsedge, T., Pong-Wong, R., Woolliams, J. A. & Villanueva, B. (2008). Restricting coancestry and inbreeding at a specific position on the genome by using optimized selection. *Genetics Research*, 90, 199–208. https://doi.org/10.1017/S0016672307009214.

Roughsedge, T., Villanueva, B. & Woolliams, J. A. (2006). Determining the relationship between restorative potential and size of a gene bank to alleviate the risks inherent in a scrapie eradication breeding programme. *Livestock Science*, 100, 231–241. https://doi.org/10.1016/j.livprodsci.2005.09.005.

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K. & Feldman, M. W. (2005). Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genetics*, 1, e70. https://doi.org/10.1371/journal.pgen.0010070.

Ruiz-López, M. J., Gañan, N., Godoy, J. A., Del Olmo, A., Garde, J., Espeso *et al*. (2012). Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *Conservation Biology*, 26, 1121–1129. https://doi.org/10.1111/j.1523-1739.2012.01916.x.

Saura, M., Pérez-Figueroa, A., Fernández, J., Toro, M. A. & Caballero, A. (2008). Preserving population allele frequencies in ex situ conservation programs. *Conservation Biology*, 22, 1277–1287. https://doi.org/10.1111/j.1523-1739.2008.00992.x.

Sotka, E. E., Wares, J. P., Barth, J. A., Grosberg, R. K. & Palumbi, S. R. (2004). Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology*, 13, 2143–2156. https://doi.org/10.1111/j.1365-294X.2004.02225.x.

Toro, M.A., Villanueva, B. & Fernández, J. (2014). Genomics applied to management strategies in conservation programmes. *Livestock Science*, 166, 48–53. https://doi.org/10.1016/j.livsci.2014.04.020.

Umina, P. A., Weeks, A. R., Kearney, M. R., McKechnie, S. W. & Hoffmann, A. A. (2005). A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. *Science*, 308, 691–693. https://doi.org/10.1126/science.1109523.

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E. *et al*. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics Selection Evolution*, 53, 1–17. https://doi.org/10.1186/s12711-021-00635-0.

Wang, J. (2012). On the measurements of genetic differentiation among populations. *Genetics Research*, 94, 275–289. https://doi.org/10.1017/S0016672312000481.

Whitlock, M. C. (2011). G'_{ST} and D do not replace F_{ST}. *Molecular Ecology*, 20, 1083–1091. https://doi.org/10.1111/j.1365-294X.2010.04996.x.

Woolliams, J.A., Berg, P., Dagnachew, B.S. & Meuwissen, T.H.E. (2015). Genetic contributions and their optimisation. *Journal of Animal Breeding and Genetics*, 132, 89–99. https://doi.org/10.1111/jbg.12148.

Wright, S. (1969). Evolution and the Genetics of Populations, Vol. 2: The Theory of Gene Frequencies. University of Chicago press.

GENERAL CONCLUSIONS

1. The magnitude of the different coancestry measures in the turbot population analyzed differed greatly. These differences can be explained by how far in the past each coefficient assumes the base population. Pairwise correlations between the different coancestry coefficients were relatively high (> 0.7). However, correlations between inbreeding coefficients (i.e., self-coancestries) varied greatly ranging from negative (– 0.4) to high positive values (1.0).
2. Managing the turbot population for a single generation with the Optimal Contribution method using $\boldsymbol{\theta}_{\mathbf{SIM}}$ $\boldsymbol{\theta}_{\mathbf{L\&H}}$ and $\boldsymbol{\theta}_{\mathbf{SEG}}$ retained more genetic diversity, measured as expected heterozygosity, than using $\boldsymbol{\theta}_{\mathbf{VR1}}$, $\boldsymbol{\theta}_{\mathbf{VR2}}$ or $\boldsymbol{\theta}_{\mathbf{YAN}}$. The higher the diversity achieved the lower was the number of individuals selected to contribute to the next generation.
3. Managing undivided populations across 50 generations with the Optimal Contribution method using $\boldsymbol{\theta}_{\mathbf{L\&H}}$ retained more expected heterozygosity, lower allelic diversity and more changes in alleles frequencies than using $\boldsymbol{\theta}_{\mathbf{VR2}}$. Therefore, the choice of which genomic coancestry matrix is used in the management may depend on which of these two goals is more important for each particular case.
4. The differences in expected heterozygosity, allelic diversity and changes in allele frequencies when using $\boldsymbol{\theta}_{\mathbf{L\&H}}$ or $\boldsymbol{\theta}_{\mathbf{VR2}}$ in OC increased with increasing population size and decreased when SNPs with low minimum allele frequency were removed.
5. When, in subdivided populations, the same weight is given to the within- and between- coancestry, we obtained the same patterns as in undivided population.

Namely, managing subdivided populations across 10 generations with the Optimal Contribution method using $\theta_{L\&H}$ retained more expected heterozygosity, lower allelic diversity and more changes in allele frequencies than using θ_{VR2}. However, when a larger weight is given to the increase of within-subpopulation coancestry, the use of $\Theta_{L\&H}$ could be the best option, since it led to higher levels of expected heterozygosity both in the global population and within subpopulations, similar loss of allelic diversity and similar changes in allele frequencies than the use of θ_{VR2}.

6. Managing subdivided populations using $\theta_{L\&H}$ the rate of increase in global and within each subpopulation inbreeding was lower than using θ_{VR2} when a weight is given to the increase of within-subpopulation coancestry.

ANNEX I:

Communications at congresses related to this thesis

International congresses

Morales-González, E., Saura, M., Fernández, A., Fernández, J., Cabaleiro, S., Martinez, P., Villanueva, B. (2018). Efficiency of different genomic coancestry matrices to maximize genetic variability in turbot selective breeding programs. 69th Annual Meeting of the European Federation of Animal Science (EAAP), Dubrovnik, Croatia, August 27-31. Oral presentation.

Villanueva, B., Saura, M., Caballero, A., Santiago, E., **Morales, E.**, Fernández, A., Fernández, J., Cabaleiro, S., Martínez, P., Millán, A., Palaiokostas, C., Kocour, M., Houston, R., Prchal, M., Bargelloni, L., Kostas, T. (2018). The importance of ensuring genetic variability when establishing selection programmes in aquaculture. AQUA 2018 (World Aquaculture Society), Montpellier, France, August 25-29. Oral presentation.

Morales-González, E., Fernández, J., Pong-Wong, R., Villanueva, B. (2019). Changes in allelic frequencies when different genomic coancestry matrices are used for maintaining genetic diversity. 37th International Society for Animal Genetics Conference (ISAG), Lérida, Spain, July 4-12. Poster

Morales-González, E.; Fernández, J.; Saura, M.; Fernández, A.; Pong-Wong, R.; Toro, M.A.; Cabaleiro, S.; Martínez, P., Villanueva, B. (2022). A comparison of genomic coancestry matrices for maintaining genetic variability using simulation and turbot data. 6º Genomics in Aquaculture Symposium, Granada, Spain, May 4-6. Oral presentation.

Morales-González, E.; Villanueva, B.; Toro, M.Á., Fernández, J. (2022). Maintenance of genetic diversity in subdivided populations using different genomic coancestry matrices. Proceedings of the 12th World Congress of Genetics Applied to Livestock Production, Rotterdam, The Netherlands, July 3–8. Oral presentation.

National congresses

Morales-González, E., Saura, M., Fernández, A., Fernández, J., Cabaleiro, S., Martinez, P., Villanueva, B. (2018). Evaluation of different genomic coancestry matrices to maintain genetic variability in a turbot selected population. 'XIX Reunión Nacional de Mejora Animal', León, Spain, June 14-15. Oral presentation.

Saura, M., Caballero, A., Santiago, E., **Morales, E.**, Fernández, A., Fernández, J., Cabaleiro, S., Martínez, P., Millán, A., Palaiokostas, C., Kocour, M., Houston, R., Prchal, M., Bargelloni, L., Kostas, T., Villanueva, B. (2018). The importance of ensuring genetic variability when establishing selection programmes in aquaculture. 'XIX Reunión Nacional de Mejora Animal', León, Spain, June 14-15. Oral presentation.

Morales-González, E.; Fernández, J.; Pong-Wong, R.; Toro, M.Á., Villanueva, B. (2021). Cambios en frecuencias alélicas cuando se utilizan diferentes matrices de parentesco genómico para mantener diversidad genética. 'XIX Jornadas Sobre Producción Animal' (Online), June 1-2. Poster.

Morales-González, E.; Villanueva, B.; Toro, M.Á., Fernández, J. (2022). Management of subdivided populations subject to conservation programs using genomic information. XX Reunión Nacional de Mejora Genética Animal, Madrid, Spain, June 1-3. Oral presentation.

Morales-González, E., Villanueva, B., Toro, M.A., Fernández, J. (2023). Management of subdivided populations subject to conservation programs using genomic information. 'XXIII Seminario de Genética de Poblaciones y Evolución', Las Caldas, Oviedo, Spain, January 18-20. Oral presentation.

ANNEX II:

Contributions to other SCI publications not part of this thesis

Saura, M., Caballero, A., Santiago, E., Fernández, A., **Morales-González, E**., Fernández, J., Cabaleiro, S., Millán, A., Millán A., Martínez, P., Palaiokostas, C., Kocour, M., Aslam, M. L., Houston R.D., Prchal M., Bargelloni L., Tzokas K., Haffray P., Bruant, J-S. & Villanueva, B. (2021). Estimates of recent and historical effective population size in turbot, seabream, seabass and carp selective breeding programmes. *Genetics Selection Evolution*, 53, 85. https://doi.org/10.1186/s12711-021-00680-9.

Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., **Morales-González, E**., Toro, M. A., & Pong-Wong, R. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics Selection Evolution*, 53, 42. https://doi.org/10.1186/s12711-021-00635-0.