Development and Application of Genomic Resources for Bighorn Sheep (*Ovis canadensis*)

by

Joshua Moses Miller

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctorate of Philosophy
in
Systematics and Evolution

Department of Biological Sciences
University of Alberta

© Joshua Moses Miller, 2015

Abstract

Since the mid-2000's there has been a major shift in molecular ecology to the use of genomic methodologies. These methods utilize genome-wide sampling of genetic variation and allow for consideration of questions that cannot be answered with a handful of microsatellite markers or a few gene sequences. However, the necessary resources for genomic analyses do not exist for many wild taxa. I developed such resources for the bighorn sheep (*Ovis canadensis*), and then applied these to several questions and analyses that can be conducted in the absence of a species-specific reference genome sequence. First, I used two parallel methodologies to rapidly discover genome-wide sets of single nucleotide polymorphisms (SNPs). Second, I used some of those loci as well as a large set of microsatellite markers to investigate how many loci and of what marker type would be needed to reflect genome-wide heterozygosity in two populations of bighorn sheep. This consideration is important for studies that wish to search for evidence of inbreeding depression in a population via heterozygosity fitness correlations (HFCs). Third, I performed a meta-analysis of 50 HFC studies to quantify the predicted magnitude of association between marker heterozygosity and inbreeding, and the number of markers that would have been needed to definitively detect such an association. Fourth, I conducted a genome wide association analysis to search for potential links between SNP variants and fitness related characteristics in a single population of bighorn sheep. I then checked the validity of the associations using an expanded set of individuals, and assessed if there have been changes in allele frequency over time. Finally, I constructed a draft whole genome sequence (WGS) from a single bighorn sheep via alignment to a domestic sheep genome as a reference. Together this work provides a robust set of genomic tools for research not only on bighorn sheep but other members of the genus *Ovis*, as well as guidance for those who wish to conduct HFC studies in any taxa.

Preface

This thesis is an original work by Joshua M Miller. However, much of the analyses would not have been possible without collaboration with the long-term research projects conducted at Ram Mountain Alberta Canada, and National Bison Range Montana USA as well as collaborators who provided sequence data or analytical guidance. Therefore, “we” is used throughout the text of the data chapters as a reflection of those involved. Portions of chapters use phenotypic data and tissue samples for DNA analyses collected previously under research protocols that were approved by the University of Alberta Animal Use and Care Committee, affiliated with the Canadian Council for Animal Care (Certificate 610901).

Portions of chapter 2 were published in two articles. The first article is Miller, Joshua M., Jocelyn Poissant, James W. Kijas, and D. W. Coltman. "A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep." Molecular Ecology Resources 11, no. 2 (2011): 314-322. Here I conducted data analyses and drafted the original manuscript. J.P., J.W.K., and D.W.C. provided DNA samples and funded the collection of SNP genotypes using the OvineSNP50 BeadChip. All authors contributed analytical guidance and provided input to the manuscript throughout its preparation. The second article is Coltman, David W., John T. Hogg, and Joshua M. Miller. "Genomic Resources Notes accepted 1 April 2013–31 May 2013." Molecular Ecology Resources 13, no. 5 (2013): 965-965. I conducted lab work prior to DNA sequencing and drafted the original manuscript. D.W.C. and J.T.H. provided DNA samples and funded the collection of SNP genotypes as well as contributed analytical guidance and provided input to the manuscript throughout its preparation.

A version of chapter 3 has been published as Miller, J. M., R. M. Malenfant, P. David, C. S. Davis, J. Poissant, J. T. Hogg, M. Festa-Bianchet, and D. W. Coltman. "Estimating genome-wide heterozygosity: effects of demographic history and marker type." Heredity 112, no. 3 (2013): 240-247. I conceived of the study, conducted data analyses, and drafted the original manuscript. R.M.M. assisted with data analyses, P.D. and C.S.D. contributed the theoretical background and concept formation. J.P., J.T.H., M.F-B., and D.W.C provided tissue samples and genetic data used in all analyses. All authors provided input to the manuscript throughout its preparation.

Aversion of chapter 4 has been published as Joshua M. Miller and David W. Coltman. "Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: a meta- analysis." *Molecular Ecology* 23, no. 8 (2014): 1899-1909. I conceived the study, collected and analyzed the data, and drafted the original manuscript. D.W.C. provided analytical guidance and input to the manuscript throughout its preparation.

A version of chapter 6 has been published as Joshua M Miller, Stephen S Moore, Paul Stothard, Xiaoping Liao, and David W. Coltman. "Harnessing cross-species alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (Ovis canadensis)" BMC Genomics 16, no. 1 (2015): doi: 10.1186/s12864-015-1618-x. I conducted bioinformatic analyses in conjunction with X.L. and analytical guidance from P.S. and DWC. Sequence data was provided by S.S.M. I drafted the original manuscript which all authors provided input on throughout its preparation.

Acknowledgments

I want to start off by thanking Dave for being a great supervisor. I will admit that I had my reservations about a mentor who admitted his style was “benign neglect”, but it turned out to be exactly what I needed. You have been extremely supportive of all my endeavors and side projects, and I cannot state how much it has meant that you were willing to send me to the four corners of the globe for training and collaboration.

I also want to highlight my committee members Corey and Paul. Corey, thank you for being an active (and mostly willing) ear for me to bounce ideas off and assist with all of the lab work. Also for helping me secure the TA-ship for 392/592 over three years, I truly enjoyed being able to develop my teaching skills and getting to chat for over six hours a week. Paul, thank you for providing bioinformatic guidance and constructive criticism throughout my degree.

In addition to my core committee there was a network of collaborators who provided data and analytical guidance. Especially Marco and Jack for giving me access to the long-term datasets for both Ram Mountain and National Bison Range populations. Much of the analyses in my thesis would not have been possible without this data and I am grateful for the trust and opportunity. Also for allowing a self-described lab person to visit both field sites and experience what it is like to collect the life history and morphological data that was shared with me in an Excel spreadsheet.

I would like to thank my funding sources as well. My salary for the past six years came from the University of Alberta, a National Science and Engeniring Research Council (NSERC) Vanier scholarship, the Killam Foundation, an Albera Innovates Technology Futures, an Andrew Stewart Memorial Graduate Prize, and a Mary Louise Imrie Graduate Student Award. My research has been funded from NSERC, Alberta Ingenuity Graduate Scholarship Research

Allowance, Alberta Conservation Association Grant in Biodiversity, and the Alberta Sport, Recreation, Parks, and Wildlife Foundation Development Initiatives Program.

As important as the funding was, so to was the fact that I had an amazing group of friends and colleagues around me. Lab members past and present, especially Aaron, Jocelyn, Rene, Jamie, Cathy, Jess R, Michelle, and Sim. As well as friends in other labs around the U of A: Josh P, Cahill, Kev, Dana, Julian, Jaz, Kim, Jill, and too many others to name. I can't express how much it meant that you took a kid from SF who liked white wine and watching TV and helped me become a part of the community in Edmonton (who now has a taste for beer and more of the outdoors). I have truly enjoyed my time here.

Then of course there is my family: my parents Dan and Dari, sister Lill, grandparents Uri, Aviva, and Shirley, as well as my aunts and uncles. You all have known my journey from a second grader with dyslexia who couldn't spell or read, to now completing a PhD. Everyone has been incredibly supportive as I continue down a path of esoteric research that is not easily explained to others.

And finally, a massive Thank You to Jess Haines. You have kept me grounded the last three years. I truly appreciate having someone at home to speak with about all aspects of life be it science, teaching, politics, or food. I know for a fact that it is unlikely I would have had the guts or motivation to go to as many cultural events in the city, let alone spend a week in Yellowknife doing science outreach. You make me a better, more well rounded person. Thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1

GENERAL INTRODUCTION

1.1 General introduction

Since the mid-2000's there has been a major shift in the field of molecular ecology: movement from "genetics" to "genomics". Though difficult to specifically define, genomics is a suite of laboratory and analytical methods that consider genome-wide sampling of markers or DNA sequence, as opposed to genetics that traditionally based inferences on tens of neutral markers (e.g. microsatellite loci) or a few candidate gene regions (Luikart *et al.* 2003; McMahon *et al.* 2014). This transition has been spurred by rapid developments in technology including high-throughput massively parallel DNA sequencing (Glenn 2011; Metzker 2010), laboratory methods that allow for high-throughput genotyping (Davey *et al.* 2011; Gunderson 2009; Shen *et al.* 2005), and increased bioinformatic capacity (Martin & Wang 2011; Miller *et al.* 2010; Narzisi & Mishra 2011; Nielsen *et al.* 2011).

With respect to data generation through high-throughput sequencing, there have been both declining costs as well as increased outputs of data. Take, for example, generating a human genome sequence. What once took a multi-national consortium over 10 years and hundreds millions of dollars (Collins *et al.* 1998) can currently be done by a single facility for a few thousand dollars in less than a week; and the field is moving towards a $1000, or even $100 genome sequence (Mardis 2006).

As such, when first developed, genomic methodologies were mostly restricted to humans, model species, and domestic organisms. However, as the costs have become less prohibitive genomic analyses of non-model and wild taxa have grown considerably (Ekblom & Galindo 2011; Ellegren 2014). This expansion is critical as it allows examination of species and populations that have not been the subject of artificial selection, thereby allowing consideration of a broader suite of research topics including: 1) determining gene content and genomic organization (Haussler *et al.* 2009; Yandell & Ence 2012); 2) generating accurate phylogenies (McCormack *et al.* 2013; Philippe & Telford 2006); 3) delineating (cryptic) populations with

fine resolution (Funk *et al.* 2012); 4) estimating demographic parameters and evolutionary histories (Harris & Nielsen 2013; Li & Durbin 2011); and 5) informing conservation and management (Allendorf *et al.* 2010; Ouborg *et al.* 2010; Shafer *et al.* 2015).

Part of the transition from "genetics" to "genomics" is development of genomic resources for a species of interest. This can include genome-wide marker sets or full genome sequence(s). The goal of my thesis was to develop and apply such genomic resources for one species: the bighorn sheep (*Ovis canadensis*).

Bighorn sheep are a mountain ungulate found in western North America from Baja California through the Canadian Rocky Mountains. They are one of six species within the genus *Ovis* and are part of a unique group the Pachyceriforms that is distinct from Eurasian sheep species including the wild relative of domestic sheep (Bunch *et al.* 2006; Rezaei *et al.* 2010). Within Pachyceriforms there are three monophyletic species: bighorn sheep, thinhorn sheep (*Ovis dalli*), and snow sheep (*Ovis nivicola*). Bighorn sheep are sister to thinhorn sheep with both species distributed in North America, while snow sheep are found in northeastern Asia. It is hypothesized that an ancestral Pachyceriform crossed the Barring Sea land bridge from Asia giving rise to the divergence between snow sheep and the North American sheep species (Bunch *et al.* 2006; Rezaei *et al.* 2010), subsequently bighorn and thinhorn sheep diverged within North America, possibly in different refugia during the last glacial maxima (Loehr *et al.* 2006).

Bighorn sheep have a complex demographic history having experienced intense hunting, local extirpations and disease-related die-offs, as well as translocations and reintroductions throughout their range (Berger 1990; Festa-Bianchet *et al.* 2014; Hedrick 2014; Johnson *et al.* 2011; Olson *et al.* 2013; Shackleton *et al.* 1999). Thus, there is interest in developing genomic resources to address a variety of evolutionary and conservation questions. In addition, there are several populations that have been the subjects of long-term individual based studies, allowing for examination of the genetic basis of phenotypic traits and individual variation in fitness (Hogg *et al.* 2006; Poissant *et al.* 2012). Generation of genomic resources for bighorn sheep may be

made easier from the close relationship between bighorn sheep and domestic sheep (*Ovis aries*; ~3 million years divergence) for which a variety of genomic resources have been developed (e.g. Jiang *et al.* 2014; Kijas *et al.* 2009).

1.2 Thesis objectives and data chapters

During the course of my PhD I conducted research that was inspired by or related to my thesis but not included in the chapters presented here. First, as an application of genomic resources in bighorn sheep I investigated of the genomic consequences of a genetic rescue in the population at National Bison Range (Montana, USA). Here I examined locus-specific effects of the rescue using a previously developed genome-wide set of microsatellite loci (n = 195). I showed that following the rescue many loci deviated from neutral patterns of inheritance, with the most common deviation indicative of directional selection for introduced alleles. Though the potential for outbreeding depression is a major concern when conducting a genetic rescue, I found no evidence of such effects in this population (Miller *et al.* 2012c).

Second, to more fully explore the utility of cross-species application of SNP chips, I investigated if there were consistent patterns of return in terms of genotyping success and polymorphism retention. To do so I found previously published cases where SNP chips developed for domestic ungulate species were applied to their wild relatives. I showed that across three different SNP chips, application to wild relatives resulted in linear decreases in call-rate (number of loci for which a genotype could be determined), but exponential decreases in the retention of polymorphisms as divergence time between the species for which the chip was developed and the one it was applied to increased. This knowledge will help researchers gauge expected success of application of these chips in taxa of interest before investing in the costs of genotyping (Miller *et al.* 2012a).

Finally, as a precursor to the construction of a draft nuclear genome sequence I constructed a full mitochondrial genome for bighorn sheep using similar techniques.

Specifically, aligning bighorn sequencing reads to a domestic sheep reference. This process confirmed the quality of the sequencing library and demonstrated that it is possible to construct a complete mitochondrial genome by "skimming reads" from a genomic sequencing library (Miller *et al.* 2012b).

The thesis is composed of five data chapters including one meta-analysis (Chapter 4). The order reflects possible developmental steps for a system moving from "genetics" to "genomics". I start with marker discovery then progresses to analyses conducted in the absence of a species-specific genome sequence. I close with the creation of a draft genome sequence that can serve a resource for future population genomic studies.

In **chapter 2**, I used two parallel methodologies to discover large numbers of SNP markers in bighorn sheep. The first was cross-species application of technology developed for domestic sheep, the Ovine SNP50 BeadChip. The second was via restriction-site associated DNA (RAD) sequencing.

In **chapter 3**, I examined the number of genetic markers (either SNP loci or microsatellite markers) that would be needed to reflect genome-wide heterozygosity in two populations of bighorn sheep. This consideration is important for studies that wish to search for evidence of inbreeding depression in a population via heterozygosity fitness correlations (HFCs). In this chapter I utilized the markers discoverer from the Ovine SNP50 BeadChip in chapter 2, as well as a set of over 200 microsatellite loci previously developed for bighorn sheep. I first tested if heterozygosity at the two marker types was correlated, an assumption for HFC analyses. I then went on to assess the correlation between different subsets of each marker type and "overall" heterozygosity to see the minimum number of markers that would be maximally informative about genome-wide heterozygosity.

In **chapter 4**, I built on the results of chapter 3 by conducting a meta-analysis examining power of previously conducted HFC studies. I started by investigating if a population's level of identity disequilibrium (a measure of the correlation in identity by descent among markers) was

related to the strength of the reported HFC. I then quantified the predicted magnitude of association between marker heterozygosity and inbreeding, and the number of markers that would have been needed to definitively detect such an association.

In **chapter 5**, I conducted a genome wide association analysis to search for potential links between SNP variants and fitness related characteristics (three morphological and five life history traits). I tested for associations using SNP genotypes from a new SNP chip, the Ovine Infinium®HD SNP BeadChip, and phenotypic records from the sheep at Ram Mountain, Alberta. I then screened candidate loci in an expanded set of individuals to check the validity of the associations, and assessed if there have been changes in allele frequency over time.

In **chapter 6**, I constructed a draft whole genome sequence (WGS) from a single bighorn sheep via alignment to a domestic sheep genome as a reference. I then called variants from the draft WGS and compared the accuracy of the SNP genotypes to ones called from the same individual on the Ovine Infinium®HD SNP BeadChip. Finally, I annotated SNPs based on annotations from the domestic sheep and compared gene ontology categories for non-synonymous and synonymous loci showing fixed differenced between the bighorn sheep draft WGS and the domestic sheep reference.

1.3 Bibliography

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Review Genetics* **11**, 697-709.

Berger J (1990) Persistence of different-sized populations: An empirical assessment of rapid extinctions in bighorn sheep. *Conservation Biology* **4**, 91-98.

Bunch T, Wu C, Zhang Y, Wang S (2006) Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. *Journal of Heredity* **97**, 21-30.

Collins FS, Patrinos A, Jordan E*, et al.* (1998) New Goals for the U.S. Human Genome Project: 1998-2003. *Science* **282**, 682-689.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* **12**, 499-510.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29**, 51–63.

Festa-Bianchet M, Pelletier F, Jorgenson JT, Feder C, Hubbs A (2014) Decrease in horn size and increase in age of trophy sheep in Alberta over 37 years. *The Journal of Wildlife Management* **78**, 133–141.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *TREE* **27**, 489–496.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**, 759-769.

Gunderson KL (2009) Whole-genome genotyping on bead arrays. In: *DNA Microarrays for Biomedical Research. Methods and Protocols* (ed. Dufva M), pp. 197-213. Humana Press, a part of Springer Science+Business Media, LLC.

Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics* **9**, e1003521.

Haussler D, O'Brien SJ, Ryder OA*, et al.* (2009) Genome 10K: A proposal to obtain whole-genome sequence for 10 000 vertebrate species. *Journal of Heredity* **100**, 659-674.

Hedrick PW (2014) Conservation genetics and the persistence and translocation of small populations: bighorn sheep populations as examples. *Animal Conservation* **17**, 106–114.

Hogg JT, Forbes SH, Steele BM, Luikart G (2006) Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1491-1499.

Jiang Y, Xie M, Chen W*, et al.* (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1168-1173.

Johnson HE, Mills LS, Wehausen JD, Stephenson TR, Luikart G (2011) Translating effects of inbreeding depression on component vital rates to overall population growth in endangered bighorn sheep. *Conservation Biology* **25**, 1240-1249.

Kijas JW, Townley D, Dalrymple BP*, et al.* (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* **4**, Article No.: e4668.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496.

Loehr J, Worley K, Grapputo A*, et al.* (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology* **19**, 419-430.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet* **4**, 981-994.

Mardis ER (2006) Anticipating the $1,000 genome. *Genome biology* **7**, 112.

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nat Rev Genet* **12**, 671-682.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526–538.

McMahon BJ, Teeling EC, Höglund J (2014) How and why should we implement genomics into conservation? *Evolutionary Applications* **7**, 999–1007.

Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46.

Miller J, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. *Genomics* **95**, 315-327.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012a) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Malenfant RM, Moore SS, Coltman DW (2012b) Short reads, circular genome: Skimming solid sequence to construct the bighorn sheep mitochondrial genome. *Journal of Heredity* **103**, 140-146.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012c) Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Molecular Ecology* **21**, 1583–1596.

Narzisi G, Mishra B (2011) Comparing de novo genome assembly: the long and short of it. *PloS one* **6**, e19175.

Nielsen R, Paul J, Albrechtsen A, Song Y (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**, 443-451.

Olson ZH, Whittaker DG, Rhodes OE (2013) Translocation history and genetic diversity in reintroduced bighorn sheep. *The Journal of Wildlife Management* **77**, 1553–1563.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics* **26**, 177-187.

Philippe H, Telford M (2006) Large-scale sequencing and the new animal phylogeny. *Trends in Ecology & Evolution* **21**, 614-620.

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2012) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity* **108**, 256–263.

Rezaei H, Naderi S, Chintauan-Marquier I*, et al.* (2010) Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution* **54**, 315-326.

Shackleton DM, Shank CC, Wikeem B (1999) Natural history of Rock Mountain and California bighorn sheep. In: *Mountain Sheep of North America* (eds. Valdez R, Krausman PR), pp. 78-138. The University of Arizona Press, Tuscon.

Shafer ABA, Wolf JBW, Alves PC*, et al.* (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**, 78–87.

Shen R, Fan J-B, Campbell D*, et al.* (2005) High-throughput SNP genotyping on universal bead arrays. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **573**, 70-82.

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet* **13**, 329-342.

Chapter 2

GENOME-WIDE SNP DISCOVERY IN TWO WILD SHEEP SPECIES: OVIS CANADENSIS & OVIS DALLI

Portions of this chapter have been published as:

Miller, Joshua M., Jocelyn Poissant, James W. Kijas, and D. W. Coltman. "A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep." *Molecular Ecology Resources* 11, no. 2 (2011): 314-322.

-&-

Coltman, David W., John T. Hogg, and Joshua M. Miller. "Genomic Resources Notes accepted 1 April 2013–31 May 2013." *Molecular Ecology Resources* 13, no. 5 (2013): 965-965.

2.1 Introduction

Single nucleotide polymorphisms (SNPs) are fast becoming the marker of choice for addressing a wide variety of evolutionary and population genetic questions (Morin *et al.*, 2004; Namroud *et al.*, 2008; Slate *et al.*, 2009; Stinchcombe & Hoekstra, 2008). SNPs offer several advantages over other markers, such as microsatellites or amplified length polymorphisms (AFLPs), including: their abundance in the genome, slower mutation rate and thus reduced levels of homoplasy, ease of genotyping through automation, and direct comparability between studies as calls are universal (Coates *et al.*, 2009; Morin *et al.*, 2004; Ryynänen *et al.*, 2007).

Recent attention has focused on using dense panels of SNPs spread throughout a genome to conduct association studies (Andersson, 2009; Hirschhorn & Daly, 2005; Karlsson *et al.*, 2007; McCarthy *et al.*, 2008). By correlating individual SNPs with a phenotype, association studies aim to identify genomic regions that influence trait variation, with the ultimate goal of identifying causal genes or even mutations. However, to optimally plan whole genome association studies, it is crucial to know the extent of linkage disequilibrium (LD) in the genome. LD is the non-random association of alleles between two loci. LD determines the number of markers needed to obtain adequate coverage in a GWAS study, as well as the precision one may hope to achieve once an association has been found. Species with extensive LD will require fewer markers than those with low levels of LD (Meadows *et al.*, 2008). Similarly, LD can be used to select the maximally informative loci from a dense panel of SNPs, thereby reducing genotyping requirements and avoiding redundancy (Carlson *et al.*, 2004; Stram, 2004). On the other hand, low levels of LD allow for finer mapping of a gene-region or gene underlying an association.

To date, most large-scale SNP resources have been developed for humans, model or domestic organisms (Feltus *et al.*, 2004; Frazer *et al.*, 2007; The International HapMap Consortium, 2007). However, researchers have begun to apply genomic tools developed in

domestic or model organisms to their wild relatives to address a variety of questions (eg. Gray *et al.*, 2009; Pertoldi *et al.*, 2010; Sacks & Louie, 2008). In this way, many wild species can be considered "genome enabled" (Kohn *et al.*, 2006) due to available genomic resources in a closely related species. Cross-species utilization of genomic tools has the potential to dramatically increase the resources available to researchers working on non-model organisms without the need for development of genomic sequencing libraries and *de novo* SNPs for every organism. Despite this potential, it remains uncertain how efficient such techniques would be as applied to progressively more phylogenetically divergent species from the one for which they were developed.

Two species that could benefit from expanded genomic resources are bighorn sheep (*Ovis canadensis*) and thinhorn sheep (*Ovis dalli*). Both are iconic North American fauna, valued by both naturalists and recreational hunters. However, over the past century there have been staggering population declines due to a combination of factors including hunting pressure, competition from domestic species, and epizootics (Valdez & Krausman 1999). In addition, for those populations subject to recreational hunting pressure there is evidence for phenotypic changes, notably a decline in horn size over time, as a result of unintentional selection (Coltman *et al.* 2003; Hengeveld & Festa-Bianchet 2011; Loehr *et al*. 2007). These patterns have led population managers to take a variety of steps to preserve both population sizes and genetic diversity including reintroductions, translocations, and population supplementations (sometimes termed a 'genetic rescue'). However, the genomic consequences of these management strategies have yet to be investigated owing to the lack of genetic resources.

Here we describe the discovery of genome-wide SNP markers using two methodologies: cross-species application of the OvineSNP50 BeadChip, and restriction-site associated DNA (RAD) sequencing (Baird *et al*., 2008). The SNP chip was applied to both bighorn and thinhorn sheep, while RAD sequencing was conducted on only bighorn sheep. We then use the chip

derived SNPs to estimate the extent of genome-wide LD in one population of bighorn sheep, and to test for power to discriminate between species and population of origin.

2.2 Methods and Materials

2.2.1 OvineSNP50 BeadChip Typing

2.2.1.1 Study Animals

The thinhorn sheep (n = 2) originated from a single population located at the Yukon-Charley Rivers National Park, Alaska. Exact population demographic data is not available for these samples and therefore we assume that they represent unrelated individuals sampled at random. Bighorn sheep included in this portion of the study originated from two populations, Ram Mountain (RM, n = 50) and Wyoming (n = 2). Population information was not available for the Wyoming samples, so again we assume that they represent two randomly sampled and unrelated individuals from a single herd. Ram Mountain Alberta, Canada, is situated about 30km east of the main Rocky Mountain range. This population has been continuously monitored since the early 1970s and the resulting pedigree and phenotypic information has been used in a number of ecological and quantitative genetic investigations (eg. Coltman *et al.*, 2003; Poissant *et al.*, 2008; Réale *et al.*, 2009). The collection of tissue samples for genetic analysis began in 1988. For the current genotyping effort, 50 animals (28 females and 22 males) born between 1988 and 2004 that survived to at least 2 years of age were selected. Genomic DNA was extracted using either Qiagen DNeasy kits or a standard phenol-chloroform protocol. Some animals were known to be related based on previous behavioral and genetic work (see Coltman et al. 2005 and reference therein for details). This includes 6 parent-offspring trios, 5 mother-lamb pairs, and 7 sire-lamb pairs.

2.2.1.2 SNP Genotyping

The OvineSNP50 BeadChip is an Illumina Infinium chip developed by the International Sheep Genomics Consortium (see http://www.sheephapmap.org/genseq.php). Briefly, SNPs were identified using a combination of Sanger resequencing and two next generation sequencers: Roche 454 FLX and Illumina Genome Analyzer. Depending on the sequencing method the DNA panel used for discovery consisted of nine (Sanger), six (454), or sixty (Illumina) primarily female domestic sheep of different breeds. SNPs from all three discovery methods were selected for genotyping based on minor allele frequency (MAF) and genomic location. The resulting chip has 49,034 SNPs which passed both the manufacturing process and rigorous quality controls.

Genomic DNA from all wild sheep was submitted to Illumina (California, USA) for commercial genotyping using the OvineSNP50 BeadChip. Raw signal intensities were converted into genotype calls using Illumina's Genome Studio software. The reliability of genotype calls was estimated using the GenCall (GC) score. GC scores were determined for each wild sheep genotype using SNP cluster information derived using 2593 domestic sheep samples. All genotype calls with GC score < 0.6 were removed from the dataset.

2.2.1.3 Summary Statistics

PLINK v1.07 (Purcell *et al.*, 2007) was used to generate summary statistics for both species including individual and locus specific call rates, assessment of the number of polymorphic sites, and calculation of minor allele frequencies (MAFs) at those sites. In addition, PLINK was used to conduct exact tests for deviation from Hardy-Weinberg equilibrium in the Ram Mountain population.

2.2.1.4 Patterns of Linkage Disequilibrium

We examined the pattern and extent of LD in the RM population using the metric r^2 calculated in Haploview (Barrett *et al.*, 2005). This metric was chosen because it is less biased

by rare alleles than other measures (Eberle *et al.*, 2006; VanLiere & Rosenberg, 2008). We restricted these analyses to SNPs with MAF ≥ 10% and > 90% genotyping rate across all individuals. Pairwise LD was calculated assuming all 50 sheep were unrelated.

Significance of LD between markers believed to be syntenic based on their position in the domestic sheep genome was determined using an empirical null distribution based on r^2 values between purportedly nonsyntenic SNP pairs (Heifetz et al. 2005). Since SNP synteny and order are unknown for bighorn sheep, this approach was based on the assumption that genome organization is conserved between domestic sheep and bighorn sheep. This assumption appears reasonable since the two species share a recent common ancestor (~3 million years ago, Bunch *et al.*, 2006; Rezaei *et al.*, 2009) and have the same karyotype (Bunch *et al.*, 1999). In addition, linkage mapping in bighorn sheep based on microsatellite loci originally mapped in domestic sheep (Poissant *et al.*, 2009) also suggests highly conserved synteny and marker order (Poissant *et al.* 2010).

To assess how far LD extends, the half-length of r^2 was calculated. Half-length was estimated as the distance (in bp) at which r^2 first fell below 50% of its maximal value (Reich *et al.*, 2001). To calculate half-lengths we averaged r^2 values for syntenic SNP pairs binned in non-overlapping 1 mega base (Mb) intervals for each chromosome (De La Vega *et al.*, 2005). In addition, a genome-wide half-length was estimated by averaging all r^2 values for SNP pairs assumed to be syntenic, binned in 1Mb intervals, across the different chromosomes. Since marker position is unknown in bighorn sheep and linkage mapping of the markers is not possible with the small number of individuals typed here, we used absolute base pair positions from the ovine genome assembly v1.0, an extension of the virtual sheep genome (Dalrymple *et al.*, 2007). The virtual sheep genome is based on sequence information from cattle but since ovine and bovine genomes are estimated to be approximately the same size (Gregory, 2010; Gregory *et al.*, 2007) and chromosome staining has shown a high level of similarity between the two taxa

(Iannuzzi & Meo, 1995) we do not expect there to be a significant systematic bias in our half-length estimations.

2.2.1.5 Discriminating Between Species & Detecting Population Differentiation

To assess the ability of the SNPs to distinguish between thinhorn and bighorn sheep, as well as between individuals originating from the two populations of bighorn sheep we used a principal component analysis conducted in the smartpca package, part of EIGANSTRAT version 3.0 (Price *et al.*, 2006). For this analysis we considered only those loci which were polymorphic within or between species and not on the X chromosome (n = 851); no outliers were removed during the calculations. In addition, Plink v1.07 was used to generate pairwise allele sharing (D_{st}) matrix comparing all individuals from both species. We quantified differences between species through a permutation test for between group identity-by-state (IBS) differences. In this test species identity was randomized between samples 10,000 times and between-group IBS was recalculated after each iteration.

We examined population differentiation within bighorn sheep using a second D_{st} matrix comparing individuals from the Ram Mountain and Wyoming populations at those loci which are polymorphic within bighorn sheep (n = 561). The permutation test then considered population locality, Ram Mountain or Wyoming, rather than species. In both cases R statistical software suite version 2.9.2 (R Development Core Team) was used to visualize the respective IBS matrices using the heatmap function.

2.2.2 RAD Sequencing

2.2.2.1 Study Animals & Sequence Library Preparation

For RAD sequencing we utilized samples from RM as well as a third population of bighorn sheep from National Bison Range (NBR), Montana USA. NBR is also the subject of a

long term monitoring project and was subject to a genetic rescue where 15 individuals were introduced to the population to increase genetic diversity after years of inbreeding (Hogg *et al.* 2006). This rescue resulted in drastic increases in both genetic diversity and life-history metrics (e.g. reproductive success).

For this marker discovery we selected four individuals from each population: three males and one female from RM; three males and one female from NBR. NBR samples included a transplanted individual, a descendant of the founding individuals, and two admixed progeny. For each individual whole genomic DNA was extracted using a standard phenol-chloroform procedure. We then quantified the DNA using both spectrophotometry (via NanoDrop; Thermo Fisher Scientific) and fluorometry (via Qubit 2.0; Life Technologies) and verified their quality on agarose gels. Samples were subsequently submitted to Floragenex (Oregon, USA) who generated the RAD library as in Baird *et al.* (2008). Briefly, samples were digested with Sbf1, individually labeled with barcoded adaptors, and then pooled for sequencing using an Illumina HiSeq2000.

2.2.2.2 Sequence Processing & SNP Calling

Floragenex provided services to run sequencing, quality control, perform alignments, and call SNP using a combination of proprietary pipelines and open source software. Briefly, after sequencing, reads were separated by individual and sequencing barcodes were removed. For RAD tag processing and SNP calling reads were aligned to the domestic sheep (*Ovis aries*) genome (version 2; http://www.livestockgenomics.csiro.au/sheep/oar2.0.php), which is thought to be highly syntenic with the bighorn sheep genome (Poissant *et al.* 2010).

SNP calling was based on output from the bowtie (version 0.11.3; Langmead *et al.* 2009) and samtools (0.1.12a; Li *et al.* 2009) algorithms and custom scripts to parse SNP information (Floragenex). Reference mapping with bowtie took sequence quality information into account, allowed for up to three mismatches between each read and the reference sequence and ignored

reads which mapped against more than one position in the genome while all other parameters remained at default. samtools tabulated SNP results for all individuals (assuming diploid individuals using the 'pileup' module and varFilter options), and we retained information on the number of reads covering each SNP (-D).

We restricted our analysis to those SNPs that were unambiguously mapped on the 27 chromosomes of the *Ovis aries* reference genome and could be genotyped in at least 75% of the individuals. In addition, we present only those loci that meet the criteria for printing on a SNP chip, specifically that the chosen SNP does not have other SNPs within 25 base pairs (-W 25) and that that we could print 50 base pairs of information on either side (-l 50). Note that SNP discovery included fixed differences between the domestic sheep reference genome as well as intraspecific polymorphisms.

2.3 Results

2.3.1 OvineSNP50 BeadChip Typing

The application of the OvineSNP50 Beadchip resulted in the successful genotyping of 48,004 and 48,230 loci in thinhorn sheep and bighorn sheep, respectively. Pooling data from both species we found 868 loci to be polymorphic. Of these, 54 were fixed differences between the two species (Supplementary Table 2-S1), 86 were polymorphic in both species, 484 were polymorphic only in bighorn sheep, and 244 were polymorphic only in thinhorn sheep (Supplementary Table 2-S2). Based on their position in domestic sheep, the polymorphic loci are likely distributed on all autosomes and the X chromosome.

Within bighorn sheep, we successfully genotyped 47,885 loci in the Ram Mountain population, 441 of which were polymorphic. Similarly, 48,124 loci were successfully genotyped in the Wyoming population, 308 of which were polymorphic. Of these loci, 181 had both alleles present in both populations, 127 were only polymorphic in the Wyoming population, and 260

were only polymorphic in Ram Mountain. We observed only two instances of fixed differences between populations (locus s56759.1 on chromosome 13 and s44723.1 on the X chromosome). Thus, application of the OvineSNP50 BeadChip to bighorn sheep identified 570 SNPs.

Observed heterozygosity in the Ram Mountain population ranged from 0.02 to 0.83 (mean ± SD = 0.30 ± 0.17) while expected heterozygosity ranged from 0.02 to 0.50 (0.30 ± 0.16). We found 14 markers to be out of Hardy-Weinberg equilibrium ($p < 0.05$). However, after Bonferroni correction only one SNP (s72530.1 on chromosome 3) still significantly deviated from equilibrium. The loci exhibited a wide but even frequency spectrum (Figure 2-1A), with 40% of loci being highly variable (MAF > 0.3) and another 40% exhibiting limited variability (MAF < 0.15). Of the polymorphic loci in the Ram Mountain population, 308 had minor allele frequencies ≥ 0.1 and call rate ≥ 90% and were used in subsequent calculations of LD. Based on the sheep genome assembly, adjacent polymorphic markers in this population were on average separated by 8.38 mega bases (Mb; Table 2-1, Figure 2-1B). Chromosomal distribution and summary statistics for these markers is presented in Table 2-1.

Mean (± SD) genome wide r^2 among syntenic markers was 0.042 ± 0.067 (Table 2-1), while average r^2 among non-syntenic SNP pairs was 0.030 ± 0.038. Of the 2,282 syntenic comparisons, 212 showed significant LD ($r^2 \geq 0.107$, $p \leq 0.05$) between marker pairs. The average r^2 value between significantly correlated syntenic SNP pairs was higher than that of significantly correlated non-syntenic pairs, 0.199 ± 0.117 versus 0.148 ± 0.045 respectively. LD for syntenic marker pairs decreased with increasing predicted physical distance (Figure 2-1C). Genome-wide average half-length was ~4.6Mb (Figure 2-1D).

Individual autosomes showed the same general trend of decreasing LD with increasing distance. Most had half-lengths of between 1 and 4Mb. Half-length was not examined for chromosome 22 due to lack of data. Chromosomes 24, 25, and X showed increasing LD with increasing distance (data not shown). We were concerned about the presence of multiple known family groups possibly inflating our estimates of LD so we repeated the calculations excluding

all offspring (n = 18). Excluding individuals resulted in minor changes in the number of SNPs meeting inclusion criteria (MAF ≥0.1, ≥90% of individuals genotyped) for a new dataset of 302 loci (Supplementary Table 2-S3). However, these calculations showed no major difference in magnitude or extent of LD: average genome wide $r^2 = 0.046 \pm 0.069$ and half-length ~4.2Mb.

Principal component analysis showed clear distinctions between bighorn and thinhorn sheep, as well as between the two populations of bighorn sheep (Figure 2-2A). The first principal component axis, which separates thinhorn from bighorn sheep, accounted for 33% of the variability in the dataset; while the second principal component axis, which distinguishes between Ram Mountain and Wyoming bighorns, accounted for an additional 8% of the variation. Patterns of IBS distinguished between the two species of wild sheep, as well as between bighorn sheep originating from Wyoming and those from Ram Mountain (Figure 2-2B). Analysis of the small number of thinhorn sheep means preliminary conclusions can be drawn concerning inter-species variation, however the number tested were insufficient to extensively examine variability within the population of thinhorn sheep from which these samples originated. Relatedness between thinhorn and bighorn sheep was low (mean $D_{st} \pm SD = 0.63 \pm 0.03$) compared to relatedness within bighorn sheep (mean $D_{st} = 0.87 \pm 0.03$).

Ram Mountain and Wyoming were still differentiated at the sub set of loci polymorphic within bighorn sheep (Figure 2-2B). Individuals from Ram Mountain were more closely related to individuals from their same population (mean $D_{st} = 0.81 \pm 0.02$) than to individuals from the Wyoming population (cross population mean $D_{st} = 0.68 \pm 0.02$). We also detected substructure within the Ram Mountain population (Figure 2-2C). This, the sub-structure roughly corresponds to known relationship categories or family groups within Ram Mountain (Figure 2-2D). For example, RM 3 is the sire of both of RM 35 and 32 (bottom right corner of Figure 2-2C). Mean D_{st} (± SD) between purportedly unrelated individuals was 0.81 ± 0.02, between half-siblings was 0.85 ± 0.01 and between parent-offspring pairs was 0.88 ± 0.01 (including one pair of full siblings $D_{st} = 0.90$).

2.3.2 RAD Sequencing

Sequencing resulted in 176,189,659 reads. From these reads 83,855 RAD loci were identified of which 38,304 had no adjacent polymorphisms and sufficient flanking sequence and are reported here. Of these loci 14,969 are polymorphic within bighorn sheep, and the remainder represents fixed differences between our samples and the domestic sheep reference. We have chosen to present these fixed differences as the small sample size may not have been sufficient to detect a rare minor allele at these positions. The total number of SNPs per chromosome are presented in Table 2-2, while the total number of SNPs per population are presented in Table 2-3. The total number of reads and genotypes per individual are presented in Table 2-4.

2.4 Discussion

We found that most (over 90%) of the domestic sheep loci on the OvineSNP50 BeadChip can be successfully called in two related wild counterparts. Many of these loci were fixed for one allele across both species, while only 868 (~2%) were polymorphic. Species specific call rates and level of polymorphism differed slightly, with 330 loci polymorphic in thinhorn sheep and 570 polymorphic in bighorn sheep. However, this difference is likely due to the large difference in the number of individuals genotyped between the two species (52 bighorn compared to 2 thinhorn sheep).

This level of conversion is on par with previous work using a panel of 1406 domestic sheep SNPs on five bighorn and four thinhorn sheep (Kijas *et al.*, 2009), and matches an empirical study (Sechi *et al.*, 2009) showing that call rate is dependant on sequence divergence between the organism for which the SNP array was developed and the organism to which it is being applied. Since the OvineSNP50 BeadChip was developed for detecting differences between domestic sheep breeds, many of which have only recently arisen, the selection of SNPs will be biased towards sites that have recent mutations. Thus, in a wild relative, many of the sites

likely exist in their ancestral monomorphic state. Pertoldi *et al*. (2010) found similar results (high genotyping success but low levels of polymorphism) when they genotyped three subspecies of bison on the BovineSNP50 BeadChip. However, they attributed the low level of polymorphism to a severe population bottleneck experienced by bison.

In the Ram Mountain population, we observed a decline in LD with increasing distance between the subset of 308 high frequency SNP markers. This is supports our assumption that SNP marker synteny and order is comparable between domestic and bighorn sheep. While it is possible that individual loci may deviate from this assumption we do not believe that such deviations are prevalent or impacted the results presented. Patterns of LD differed among chromosomes; similar interchromosomal differences in the level of LD have been seen in humans (Reich *et al.*, 2001) and other animals (Khatkar *et al.*, 2008; Slate & Pemberton, 2007). This includes a twenty-five fold difference in magnitude of LD between genic regions in humans (6 kb vs. 155 kb) which was attributed to stochastic processes such as gene history. The aberrant LD pattern observed on chromosomes 24, 25, and the X chromosome could be due to minor chromosomal rearrangements between bighorn sheep and domestic sheep but may simply be a chance artifact of the limited number of markers present on each chromosome.

The genome-wide half-lengths observed in this study indicate that there is extensive LD in the genome of bighorn sheep from Ram Mountain. The presence of known family groups in the dataset does not seem to account for this high level of LD as similar results are obtained even when only unrelated individuals are considered. The extent of LD observed in this study is longer than that reported in humans (De La Vega *et al.*, 2005; Reich *et al.*, 2001), mice (Laurie *et al.*, 2007), or other natural populations (Backström *et al.*, 2006; Gray *et al.*, 2009) where LD commonly extends for only tens to hundreds of kilobases, and usually less than one mega base. It is closer to the scale observed in domestic animals such as pigs (Harmegnies *et al.*, 2006), chickens (Heifetz *et al.*, 2005), cattle (Khatkar *et al.*, 2008), dogs (Gray *et al.*, 2009; Sutter *et al.*, 2004), thoroughbred horses (Tozaki *et al.*, 2005), and domestic sheep (McRae *et al.*, 2002;

Meadows *et al.*, 2008). Consequently, relatively fewer SNPs will be required to conduct genome-wide association studies than in many other species that have been characterized thus far. However, we still require considerably more SNPs than the 308 bighorn sheep SNPs for genome-wide coverage.

The similar levels of LD in bighorn sheep and domestic animals may seem contradictory, as one might expect a wild or outbred population to have lower levels of LD than a group that has been subject to the intense selective pressures and breeding regimes associated with domestication and commercial production. One contributing factor may be the properties of the SNP subset that was used to estimate LD. These markers were identified by virtue of their polymorphism in both wild and domestic sheep. Consequently, the SNP subset was enriched for loci with old mutations that likely predate the split between wild and domestic lineages and it is unclear what impact this is likely to have on observed levels of LD. However, despite this potential ascertainment bias it appears the marker set behaves as expected given that the distribution of allele frequency in the SNPs within Ram Mountain is balanced.

In addition to the SNPs used, several aspects of bighorn sheep biology may also lead to elevated LD. First, bighorns have a polygynous mating system where the majority of offspring are sired by a minority of males (Coltman *et al.*, 2002; Hogg & Forbes, 1997). Second, the Ram Mountain population is small and rarely receives migrants from other areas (Festa-Bianchet *et al.*, 2008). Together these factors are likely to lead to nonrandom mating and therefore extend the levels of LD throughout the genome. It is important to note that such factors are not limited to the population at Ram Mountain. Rather, they are characteristic of many mountain ungulates which tend to live in highly structured and effectively small populations (Crestanello *et al.* 2009; Mainguy *et al.*, 2008; Worley *et al.*, 2004).

Given that the distance between adjacent markers is sometimes longer that the expected distance that LD extends, the panel of high frequency SNPs currently available for bighorn sheep does not provide genome-wide coverage necessary for an exhaustive association study.

However, we were able to differentiate between populations as well as detect substructure within Ram Mountain. This substructure roughly corresponds to known relationships and family groups within the Ram Mountain population. The fact that the different relatedness categories were not unambiguously separated from one another could be due to the underlying population structure in the Ram Mountain population. For example, the apparent right skew in the distribution of unrelated pairs (Figure 2-2D) likely represents some genuine distant relatives rather than truly unrelated pairs.

We have shown that cross-species application of the OvineSNP50 BeadChip can provide a valuable source of genomic markers in taxa other than domestic sheep. This new genomic resource provides an excellent tool for future studies such as linkage mapping, candidate gene association studies, and population genomics. Moreover, knowledge of the level of LD in the bighorn sheep genome is an informative and useful baseline for future efforts to fine map QTLs in this species. In addition to taxa specific questions, broader application of the OvineSNP50 BeadChip could be used to study the genetic consequences of domestication and the relatedness between domestic sheep and their wild counterparts. Genomic position and sequence information for all SNPs on the OvineSNP50 BeadChip (including those presented in Supplementary Table 2-S1) are available via the Ovine Genome assembly (www.livestockgenomics.csiro.au/perl/gbrowse.cgi/).

Typing individuals on a platform such as the OvineSNP50 BeadChip can be more efficient and cost effective than other marker discovery efforts, such as cross-species amplification microsatellite loci (e.g. Poissant et al. 2009). However, the efficiency and efficacy of such an endeavor is highly dependent on the level of divergence between the study taxa and taxa for which the resource was developed (Miller et al. 2012). Nonetheless, cross-species use of SNP arrays developed in a domestic or model species to wild counterparts clearly appears to be a valuable approach to develop a set of genome-wide loci.

However, in species where array based SNP genotyping is not an option, genotype-by-sequencing methodologies such as RAD sequencing represent an alternative for discovery of large marker sets (Elshire et al. 2011; Puritz et al. 2014). Our RAD sequencing of 8 individual bighorn sheep from two populations returned nearly 15,000 loci polymorphic within these samples, and over 38,000 polymorphic compared to the domestic sheep reference. Sequencing additional individuals from the same populations or other populations would likely only increase this number.

Table 2 - 1 Summary statistics for high frequency chip derived SNPs in bighorn sheep from RM

Chr	No. SNPs	Mean MAF± SD	Avg. spacing between adjacent SNPs ± SD (kbp)	Mean r^2 ± SD	Median r^2
1	24	0.29 ± 0.13	12.19 ± 9.89	0.035 ± 0.058	0.016
2	36	0.29 ± 0.12	7.42 ± 7.78	0.037 ± 0.059	0.018
3	24	0.31 ± 0.12	9.44 ± 10.39	0.035 ± 0.048	0.018
4	14	0.25 ± 0.11	9.07 ± 5.81	0.039 ± 0.043	0.023
5	15	0.28 ± 0.10	7.95 ± 5.84	0.048 ± 0.082	0.022
6	13	0.26 ± 0.13	11.53 ± 11.42	0.036 ± 0.050	0.017
7	14	0.29 ± 0.12	6.72 ± 7.08	0.045 ± 0.070	0.019
8	17	0.33 ± 0.09	5.51 ± 5.19	0.051 ± 0.071	0.019
9	14	0.28 ± 0.13	7.31 ± 7.03	0.035 ± 0.047	0.015
10	11	0.27 ± 0.11	8.23 ± 3.99	0.042 ± 0.067	0.011
11	7	0.29 ± 0.10	9.24 ± 8.94	0.044 ± 0.066	0.029
12	8	0.25 ± 0.12	11.73 ± 11.86	0.043 ± 0.063	0.017
13	8	0.30 ± 0.15	8.23 ± 9.49	0.046 ± 0.068	0.028
14	4	0.32 ± 0.18	8.65 ± 4.26	0.035 ± 0.037	0.023
15	5	0.30 ± 0.14	17.82 ± 22.12	0.025 ± 0.032	0.011
16	9	0.28 ± 0.11	7.85 ± 10.85	0.071 ± 0.100	0.020
17	9	0.26 ± 0.10	8.36 ± 8.57	0.052 ± 0.058	0.030
18	12	0.33 ± 0.11	5.95 ± 5.64	0.060 ± 0.098	0.028
19	10	0.21 ± 0.09	6.64 ± 8.75	0.100 ± 0.142	0.051
20	11	0.28 ± 0.12	2.81 ± 3.36	0.085 ± 0.128	0.040
21	6	0.24 ± 0.15	9.56 ± 9.34	0.034 ± 0.027	0.032
22	3	0.18 ± 0.09	23.84 ± 16.25	0.006 ± 0.004	0.007
23	10	0.32 ± 0.14	6.54 ± 5.21	0.041 ± 0.042	0.023
24	5	0.28 ± 0.08	6.07 ± 2.18	0.026 ± 0.030	0.014
25	5	0.22 ± 0.11	9.11 ± 7.79	0.030 ± 0.033	0.022
26	8	0.24 ± 0.13	6.10 ± 8.96	0.041 ± 0.023	0.029
X	6	0.26 ± 0.12	11.90 ± 10.76	0.033 ± 0.032	0.021
All	308	0.27 ± 0.02	8.38 ± 8.53	0.042 ± 0.067	0.019

Table 2 - 2 Number of RAD loci per chromosome and average spacing of SNPs

Chr	All Loci		Bighorn Specific Loci	
	No. of Loci	Avg Intermarker Distance (bp)[a]	No. of Loci	Avg Intermarker Distance (bp)[a]
1	3489	79134	1353	204155
2	3231	77449	1215	205897
3	3537	63332	1384	161902
4	1519	77973	647	183198
5	1673	64726	595	182192
6	1100	106327	440	266182
7	1301	76698	494	202051
8	993	91450	369	246494
9	1193	79559	478	198814
10	1010	82785	387	215705
11	1635	38108	633	98367
12	1382	57383	534	148670
13	1820	45610	721	114470
14	1449	42938	565	109931
15	1203	67477	481	168974
16	900	79474	358	200132
17	1186	60626	491	142941
18	1225	55544	485	140010
19	1123	54053	429	140587
20	975	52161	397	128296
21	982	49283	379	127234
22	865	58483	321	156218
23	1046	59855	432	145124
24	1084	38435	378	110409
25	760	57993	318	137969
26	639	68974	231	191190
X	965	127082	454	270436
Grand Total	38285	67082	14969	171420

[a] Distances were taken from the domestic sheep genome version 2.0 (http://www.livestockgenomics.csiro.au/sheep/oar2.0.php)

Table 2 - 3 Number of RAD tags and polymorphic loci per population

Population	Average Number of Tags per Individual	SD of Number of Tags per Individual	Total Number of Polymorphic SNPs	Number of Population Specific SNPs
RM	37399	1461.8	10699	6786
NBR	38095	237.7	8183	4270

Table 2 - 4 Number of reads and RAD genotypes per individual

Population	Individual	No. of Reads	No. of Genotypes	No. of Hz Loci
RM				
	E12	19169111	38019	2617
	N11	11562855	38195	2875
	P16	28516030	38115	3000
	P6	23564854	34881	1944
NBR				
	02-04	20873793	38238	3112
	03-05	36964387	38186	2656
	89-11	7157717	37714	2782
	97-16	28380912	38203	3252

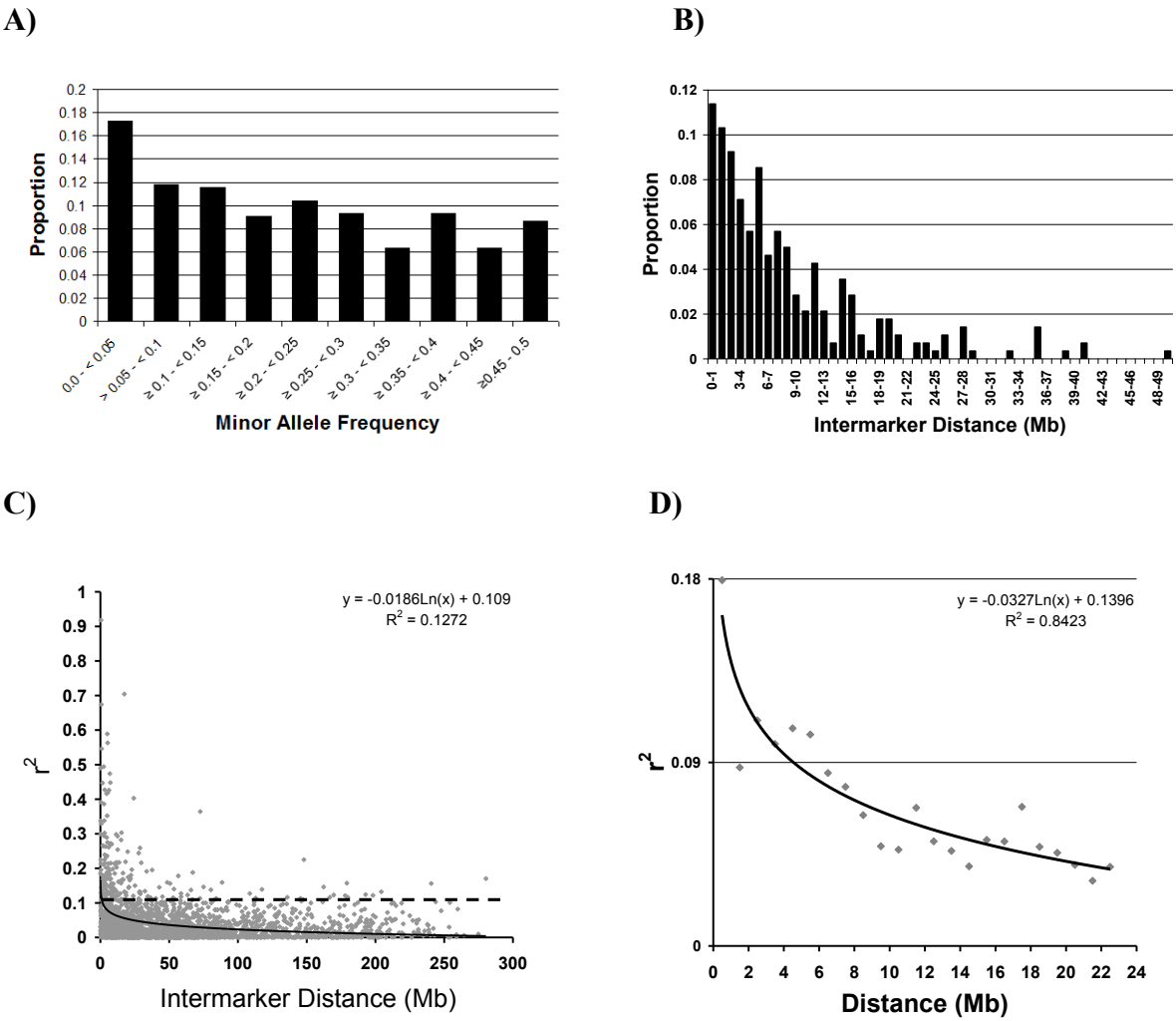


Figure 2 - 1 Allele frequency distribution for polymorphic SNPs within Ram Mountain (N = 441 loci). B: Frequency distribution of distance between adjacent marker pairs used in LD calculations (N = 308 loci). C: LD measured by r^2 plotted as a function of intermarker distance (Mbp). A logistic fitted line is shown (solid line); dashed line indicates empirically determined significance threshold ($r^2 = 0.107$). D: Genome wide half-length measured by r^2 plotted as a function of intermarker distance (Mbp). A logistic fitted line is shown.

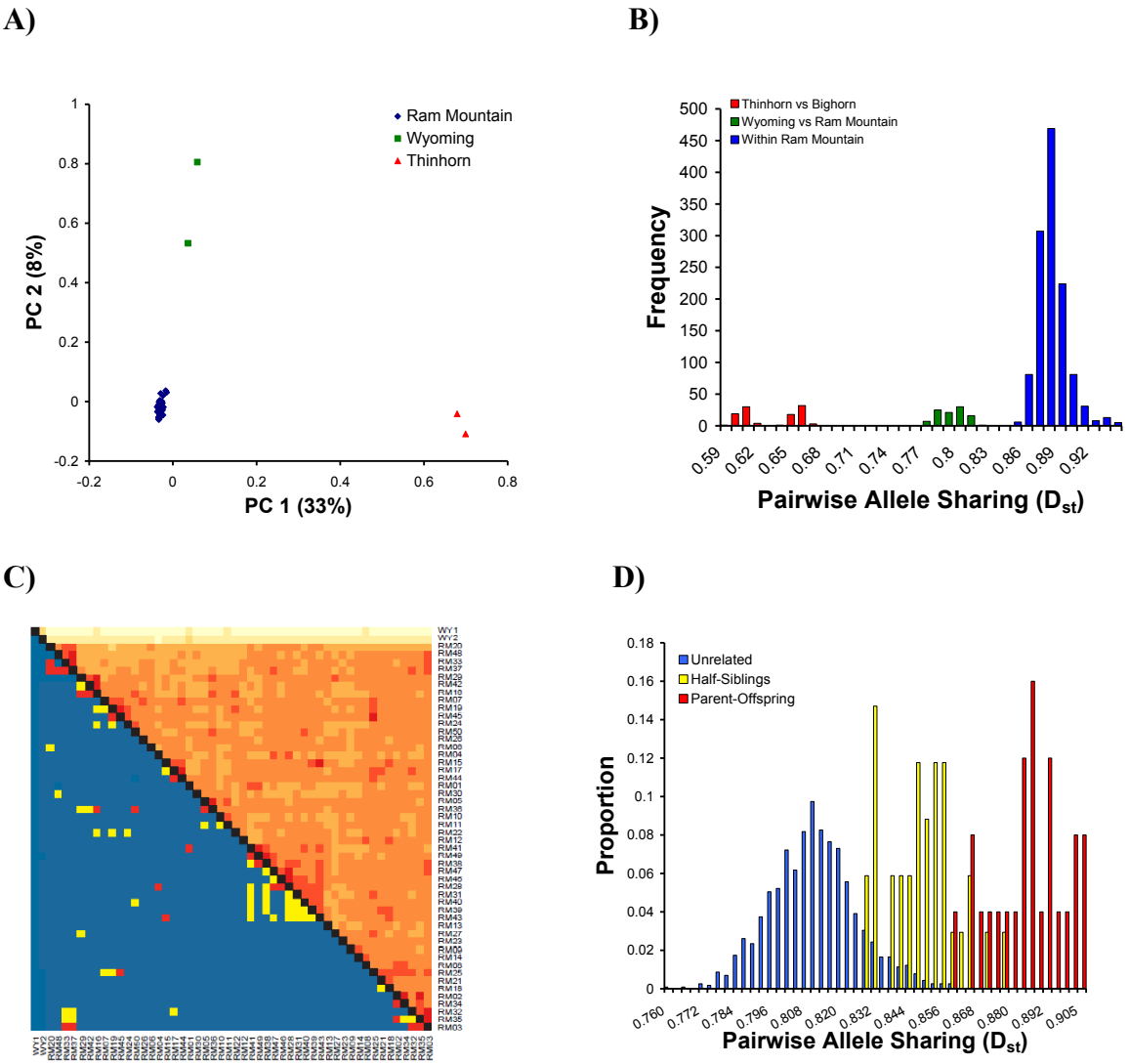


Figure 2 - 2 Clustering of individuals based on first two principal component axes. B: Distribution of D_{st} values between bighorn populations and thinhorn sheep. C: Heatmap of genetic similarity between individual bighorn sheep based on pedigree relationships (below diagonal) and allele sharing (above diagonal). Dark squares indicate high allele sharing between two individuals, light squares indicate low allele sharing. WY = Wyoming bighorn, RM = Ram Mountain bighorn. D: Distribution of D_{st} values within the Ram Mountain population.

2.5 Bibliography

Andersson L (2009) Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* **136**, 341-349.

Backström N, Qvarnström A, Gustafsson L, Ellegren H (2006) Levels of linkage disequilibrium in a wild bird population. *Biology Letters* **2**, 435-438.

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**, e3376.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265.

Bunch TD, Hoffmann RS, Nadler CF (1999) Cytogenetics and genetics. In: *Mountain Sheep of North America* (eds. Valdez R, Krausman PR). The University of Arizona Press, Tuscon.

Bunch TD, Wu C, Zhang YP, Wang S (2006) Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. *Journal of Heredity* **97**, 21-30.

Carlson CS, Eberle MA, Rieder MJ*, et al.* (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *The American Journal of Human Genetics* **74**, 106-120.

Coates BS, Sumerford DV, Miller NJ*, et al.* (2009) Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *Journal of Heredity* **100**, 556-564.

Coltman DW, Festa-Bianchet M, Jorgenson JT, Strobeck C (2002) Age-dependent sexual selection in bighorn rams. *Proceedings: Biological Sciences* **269**, 165-172.

Coltman DW, O'Donoghue P, Jorgenson JT*, et al.* (2003) Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655-658.

Crestanello B, Pecchioli E, Vernesi C*, et al.* (2009) The Genetic Impact of Translocations and Habitat Fragmentation in Chamois (Rupicapra) spp. *Journal of Heredity* **100**, 691-708.

Dalrymple B, Kirkness E, Nefedov M*, et al.* (2007) Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology* **8**, R152.

De La Vega FM, Isaac H, Collins A*, et al.* (2005) The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Research* **15**, 454-462.

Eberle MA, Rieder MJ, Kruglyak L, Nickerson DA (2006) Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome. *PLoS Genetics* **2**, e142.

Elshire R, Glaubitz J, Sun Q, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. Plos One 6.

Feltus FA, Wan J, Schulze SR*, et al.* (2004) An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. *Genome Research* **14**, 1812-1819.

Festa-Bianchet M, Coltman DW, Hogg JT, Jorgenson JT (2008) Age-related horn growth, mating tactics, and vulnerability to harvest: why horn curl limits may select for small horns in bighorn sheep. *Biennial Symposium of the Northern Wild Sheep and Goat Council* **15**, 42-49.

Frazer KA, Eskin E, Kang HM*, et al.* (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* **448**, 1050-1053.

Gray MM, Granka J, Bustamante CD*, et al.* (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* **181**, 1493-1505.

Gregory, TR (2010). Animal Genome Size Database. http://www.genomesize.com

Gregory TR, Nicol JA, Tamm H, *et al.* (2007) Eukaryotic genome size databases. *Nucleic Acids Research* **35**, D332-338.

Hengeveld PE, Festa-Bianchet M (2011) Harvest regulations and artificial selection on horn size in male bighorn sheep. *The Journal of Wildlife Management* **75**, 189-197.

Harmegnies N, Farnir F, Davin F*, et al.* (2006) Measuring the extent of linkage disequilibrium in commercial pig populations. *Animal Genetics* **37**, 225-231.

Heifetz EM, Fulton JE, O'Sullivan N*, et al.* (2005) Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* **171**, 1173-1181.

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95-108.

Hogg JT, Forbes SH (1997) Mating in bighorn sheep: frequent male reproduction via a high-risk "unconventional" tactic. *Behavioral Ecology and Sociobiology* **41**, 33-48.

Hogg JT, Forbes SH, Steele BM, Luikart G (2006) Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1491-1499.

Iannuzzi L, Meo GP (1995) Chromosomal evolution in bovids: a comparison of cattle, sheep and goat G- and R-banded chromosomes and cytogenetic divergences among cattle, goat and river buffalo sex chromosomes. *Chromosome Research* **3**, 291-299.

Karlsson EK, Baranowska I, Wade CM*, et al.* (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics* **39**, 1321-1328.

Khatkar M, Nicholas F, Collins A*, et al.* (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* **9**, 187.

Kijas JW, Townley D, Dalrymple BP*, et al.* (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* **4**, Article No.: e4668.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**.

Laurie CC, Nickerson DA, Anderson AD*, et al.* (2007) Linkage disequilibrium in wild mice. *PLoS Genetics* **3**, e144.

Li H, Handsaker B, Wysoker A*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Loehr J, Carey J, Hoefs M, Suhonen J, & Ylönen H. (2007). Horn growth rate and longevity: implications for natural and artificial selection in thinhorn sheep (Ovis dalli). *Journal of evolutionary biology*, **20**, 818-828.

Mainguy J, Côté SD, Cardinal E, Houle M (2008) Mating tactics and mate choice in relation to age and social rank in male mountain goats. *Journal of Mammalogy* **89**, 626-635.

McCarthy MI, Abecasis GR, Cardon LR*, et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* **9**, 356-369.

McRae AF, McEwan JC, Dodds KG*, et al.* (2002) Linkage disequilibrium in domestic sheep. *Genetics* **160**, 1113-1122.

Meadows J, Chan E, Kijas J (2008) Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* **9**, 61.

Miller, JM, Kijas, JW, Heaton, MP, McEwan, JC, & Coltman, DW (2012). Consistent divergence times and allele sharing measured from cross- species application of SNP chips developed for three domestic species. *Molecular ecology resources*, **12**, 1145-1150.

Morin PA, Luikart G, Wayne RK, SNP Workshop Grp (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19**, 208-216.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* **17**, 3599-3613.

Pertoldi C, Wójcik J, Tokarska M*, et al.* (2010) Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* **11**, 627-634 .

Poissant J, Hogg JT, Davis CS, *et al.* (2010). Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. *BMC genomics*, **11**.

Poissant J, Shafer ABA, Davis CS*, et al.* (2009) Genome-wide cross-amplification of domestic sheep microsatellites in bighorn sheep and mountain goats. *Molecular Ecology Resources* **9**, 1121-1126.

Poissant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW (2008) Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proceedings of the Royal Society B-Biological Sciences* **275**, 623-628.

Price AL, Patterson NJ, Plenge RM, *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904-909.

Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575.

Puritz JB, Matz MV, Toonen RJ, et al. (2014) Demystifying the RAD fad. Molecular Ecology 23, 5937-5942.

R Development Core Team (2009) R: A language and environment for statistical computing**.** R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2005 [http://www.Rproject.org].

Réale D, Martin J, Coltman DW, Poissant J, Festa-Bianchet M (2009) Male personality, life-history strategies and reproductive success in a promiscuous mammal. *Journal of Evolutionary Biology* **22**, 1599-1607.

Reich DE, Cargill M, Bolk S*, et al.* (2001) Linkage disequilibrium in the human genome. *Nature* **411**, 199-204.

Rezaei HR, Naderi S, Chintauan-Marquier IC*, et al.* (2009) Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution* **54**, 315-326.

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *Journal of Heredity* **98**, 692-704.

Sacks BN, Louie S (2008) Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. *Molecular Ecology Resources* **8**, 35-49.

Sechi T, Coltman DW, Kijas JW (2010) Evaluation of 16 loci to examine the cross-species utility of single nucleotide polymorphism arrays. *Animal Genetics* **41**, 199-202.

Slate J, Gratten J, Beraldi D*, et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97-107.

Slate J, Pemberton JM (2007) Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. *Journal of Evolutionary Biology* **20**, 1415-1427.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158-170.

Stram DO (2004) Tag SNP selection for association studies. *Genetic Epidemiology* **27**, 365-374.

Sutter NB, Eberle MA, Parker HG*, et al.* (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Research* **14**, 2388-2396.

The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861.

Tozaki T, Hirota K, Hasegawa T, Tomita M, Kurosawa M (2005) Prospects for whole genome linkage disequilibrium mapping in thoroughbreds. *Gene* **346**, 127-132.

Valdez R, Krausman PR (1999) Description, distribution, and abundance of mountain sheep in North America. In: *Mountain Sheep of North America* (eds. Valdez R, Krausman PR), pp. 3-22. The University of Arizona Press, Tuscon.

VanLiere JM, Rosenberg NA (2008) Mathematical properties of the r2 measure of linkage disequilibrium. *Theoretical Population Biology* **74**, 130-137.

Worley K, Strobeck C, Arthur S*, et al.* (2004) Population genetic structure of North American thinhorn sheep (*Ovis dalli*). *Molecular Ecology* **13**, 2545-2556.

Supplementary Information

Table 2-S1: Loci exhibiting fixed differences between bighorn and thinhorn sheep.

Table 2-S2: Polymorphic SNPs.

Table 2-S3: Summary statistics for high frequency SNPs in reduced set of bighorn sheep from Ram Mountain.

Tables S1-S3 can be found at http://onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2010.02918.x/suppinfo

RAD related sequence files (in fastq format) along with the results of the reference mapping and SNP calling were submitted to Dryad (doi:10.5061/dryad.4qk81). This included information on the position of each SNP, its allelic state, the number of reads covering the SNP, and consensus genotypes in each individual.

Chapter 3

ESTIMATING GENOME-WIDE HETEROZYGOSITY: EFFECTS OF DEMOGRAPHIC HISTORY AND MARKER TYPE

A version of this chapter has been published:

Miller, J. M., R. M. Malenfant, P. David, C. S. Davis, J. Poissant, J. T. Hogg, M. Festa-Bianchet, and D. W. Coltman. "Estimating genome-wide heterozygosity: effects of demographic history and marker type." *Heredity* 112, no. 3 (2013): 240-247.

3.1 Introduction

Individual heterozygosity can be easily measured using genetic markers and is often used as proxy for inbreeding (Balloux *et al.* 2004; Hansson & Westerberg 2002). Many studies have examined the relationship between individual genetic diversity and fitness using heterozygosity–fitness correlations (HFCs). Three processes are thought to potentially underlie HFCs (Hansson & Westerberg 2002). 1) The markers themselves can have functional consequences and be directly linked to differences in fitness (direct-effect HFCs). 2) The genotyped markers themselves may not have direct consequences on fitness, but rather are in linkage disequilibrium with variants that are (local-effect HFCs). 3) There may be an intrinsic benefit to being heterozygous and the heterozygosity of the typed markers is correlated to overall heterozygosity in the genome (general-effect HFCs).

HFCs appeal to wildlife and conservation biologists who cannot easily reconstruct pedigrees and directly measure inbreeding in natural populations, especially for endangered species (Balloux *et al.* 2004; Chapman *et al.* 2009; Grueber *et al.* 2008). Meta-analyses of HFCs, however, have revealed that effect sizes are often weak (Chapman *et al.* 2009; Coltman & Slate 2003; Szulkin *et al.* 2010). The modest numbers of genetic markers typically employed may provide inaccurate estimates of genome-wide heterozygosity. It is also unclear whether different types of markers provide similar information content.

Microsatellites have been the most commonly used markers to investigate HFCs. They are relatively abundant in the genome and highly polymorphic. However, their mutational mechanism is not well understood, and the high mutation rate likely leads to elevated levels of homoplasy which can underestimate true heterozygosity (Hansson & Westerberg 2002). In

addition, the process of isolating and characterizing novel loci often selects for the most polymorphic markers, resulting in ascertainment bias and an upwardly skewed estimate of genome-wide diversity (Brandstrom & Ellegren 2008).

Despite their growing use in molecular ecology and evolutionary biology, single nucleotide polymorphisms (SNPs) have been less widely used in HFC studies, perhaps because they are almost exclusively bi-allelic. However, they have some advantages over microsatellites: they are more abundant in the genome, have a well understood mutational mechanism with low levels of homoplasy, and are amenable to high throughput genotyping (Morin *et al.* 2004). Several authors contend that SNPs may be more suitable than microsatellites for HFCs. Tsitrone *et al.* (2001) used extensive simulation studies to examine the effect of different mutational patterns (corresponding to SNPs and microsatellites) and demographic history on the expected correlation between heterozygosity and fitness. Their results point to a complex interplay between these two factors. The high mutation rate of microsatellites should make them more suitable to detect HFCs that result from recent inbreeding due to crosses between relatives or small population size. The lower mutation rates typical of SNPs may make them better than microsatellites to detect HFCs resulting from ancient inbreeding, such as when two subpopulation accumulate genetic differentiation during a long period of isolation and then come back into contact (Tsitrone *et al.* 2001). Chakraborty (1981) and DeWoody & DeWoody (2005) argued that correlations between heterozygosity at a set of loci and genomic heterozygosity would be high only when the set of marker loci represents a high fraction of all polymorphisms in the genome. However, they modeled populations with no inbreeding and no correlations among loci, a condition under which HFC does not occur unless the marker loci themselves are coding for fitness traits (direct-effect HFCs), which is not the case for most recently published HFC studies. It remains unclear whether many markers with low genetic diversity (SNPs) or fewer markers with higher diversity (microsatellites) are more suitable to explore general-effect HFCs. This question becomes important as new technologies allow for the development of larger

genome-wide marker sets of both SNPs and microsatellites (Baird *et al.* 2008; Davey *et al.* 2011).

Studies of HFCs commonly use 10–30 loci (Chapman *et al.* 2009). But because the demographic history of a population will heavily influence correlations in marker heterozygosity within individuals (Ljungqvist *et al.* 2010; Szulkin *et al.* 2010) such modest numbers of markers may sometimes be insufficient (Balloux *et al.* 2004; Forstmeier *et al.* 2012; Ljungqvist *et al.* 2010; Väli *et al.* 2008). For example, Väli et al. (2008) looked at the correlation between heterozygosity at 10–17 microsatellites and allelic diversity in 10 introns across eight populations of carnivores. They found a positive correlation between average heterozygosity and allelic diversity among populations, but not between individual heterozygosity at SNPs and microsatellites. In general, HFCs are not expected within populations without measurable identity disequilibrium (ID), a correlation in identity by descent among markers (David *et al.* 2007; Slate *et al.* 2004; Szulkin *et al.* 2010). ID arises from a departure from random mating (e.g. inbreeding) or demographic events (e.g. a population bottleneck or admixture) that cause the heterozygosity of loci to become associated within individuals (Bierne *et al.* 2000; Szulkin *et al.* 2010). In the absence of ID, HFCs will be detected only if one or more markers are directly associated with a gene influencing fitness, so called local or direct effects (Hansson & Westerberg 2002). These direct effect correlations are difficult to detect because they depend on the specific marker set used in a study.

Here we examine the contrasting effects of the number of markers considered and marker type on the ability to detect general-effect HFCs. We first use existing models of HFC to derive broad theoretical predictions about how many loci are needed to adequately measure genomic heterozygosity assuming different levels of inbreeding and marker genetic diversity. We then use large sets of both microsatellites and SNPs genotyped in two populations of bighorn sheep (*Ovis canadensis*) to approach this question empirically. Our two study populations have very different demographic histories: one was founded in the 1920's with 12 individuals and experienced a

prolonged bottleneck post founding, then recent admixture following a 'genetic rescue' where 15 individuals were intentionally introduced into the population. The other is a native population with no genetic evidence of a comparable bottleneck. These contrasting histories should affect the magnitude of ID and hence our ability to detect HFCs. In the population subject to 'genetic rescue', ID is expected to be higher, arising both from historical inbreeding and admixture following the introductions. In the native population ID will likely track demography, arising if heterozygosity decreases due to inbreeding. Therefore power to detect HFCs should be greater in the bottlenecked population than the native one. To test these hypotheses we first sought to measure the strength of correlations between estimates of heterozygosity from microsatellites and SNPs within individuals. We then examined how many markers are needed to accurately reflect genome-wide heterozygosity and ID in these two populations.

3.2 Theory

General effect HFCs arise as the product of two correlations: the correlation between fitness (W) and inbreeding (f), and the correlation between f and heterozygosity (h) (Slate *et al.* 2004; Szulkin *et al.* 2010). Such that:

$$r(W,h) = r(W,f) \cdot r(f,h) \quad \text{Eq.1}$$

For the purposes of this paper we do not consider the correlation between W and f. Rather we focus on the power of different marker sets to detect HFCs.

Two sources of sampling variance may affect HFCs. One is the sampling of individuals: if a small sample is taken in a population that contains a small proportion of inbred individuals, the proportion of inbred individuals in the sample is subject to a large variance which directly affects HFC estimates. The estimated HFC will be stronger or weaker than true HFC in the population simply because the proportion of inbred individuals in the sample happens to be higher or lower than their frequency in the population. This source of error can be large but the only way to reduce it is to sample more individuals. The second source of variance arises from

the fact that heterozygosity measured at a set of marker loci is not perfectly correlated with genomic heterozygosity and/or with individual inbreeding level. This error depends on the characteristics of the marker loci (number and genetic diversity). We will mainly concentrate on this second type of error, assuming that all efforts have been made to reduce the first source of error.

The problem is now to estimate how well inbreeding is measured by (or correlated to) heterozygosity in a sample of markers. We consider standardized heterozygosity (sensu Coltman *et al.* (1999); and noted H** for consistency with notations in Szulkin *et al.* (2010)) at a set of loci (A) which comprises L_A loci; h_i is the observed heterozygosity at locus *i* and upper bar denotes expectations.

$$H_A^{**} = \frac{\sum_{i\in A} h_i}{\sum_{i\in A} \bar{h}_i} = \frac{\sum_{i\in A} h_i}{L_A \bar{h}_A} \qquad \text{Eq.2}$$

Based on Szulkin et al. 2010 the expected correlation between H** and inbreeding level (*f*) is

$$r^2(H_A^{**}, f) = \frac{g_2}{\sigma^2(H_A^{**})} \qquad \text{Eq.3}$$

Where g_2 is the covariance of heterozygosity between markers standardized by their average heterozygosity (David *et al.* 2007), and

$$\sigma^2\left(\mathrm{H}_\mathrm{A}^{**}\right) = \frac{\sum_{\mathrm{i}\in\mathrm{A}} \bar{\mathrm{h}}_\mathrm{i}(1-\bar{\mathrm{h}}_\mathrm{i}) + 2\mathrm{g}_2 \sum\sum_{(\mathrm{j}>\mathrm{i})\in\mathrm{A}} \bar{\mathrm{h}}_\mathrm{i}\bar{\mathrm{h}}_\mathrm{j}}{\left(\mathrm{L}_\mathrm{A}\bar{\mathrm{h}}_\mathrm{A}\right)^2} \qquad \text{Eq.4}$$

Assuming that all loci in set A have the same average heterozygosity h_A (for simplicity) this gives

$$r^2(H_A^{**}, f) = \frac{L_A g_2 \bar{h}_A}{1-\bar{h}_A + (L_A - 1) g_2 \bar{h}_A} \qquad \text{Eq. 5}$$

From this it can be seen that, after a certain number of loci are sampled, the correlation approaches unity as the number of markers increases, and depends mostly on the product of the number of loci by their average heterozygosity: 100 loci with h=0.1 are equivalent to 20 loci with h=0.5. The rate at which the correlation approaches 1 increases with the identity disequilibrium (represented by g_2). When g_2 is null, the correlation is necessarily zero because

inbreeding does not vary in the population. In such cases, trying to estimate genome-wide heterozygosity from a small set of markers is pointless, because all the variance comes from sampling error.

Often one does not have an independent measure of inbreeding (e.g. pedigrees) and therefore the above formula cannot be checked directly. Instead what can be done (and will be done below using real data) is (i) to check the consistency between two different subsets of marker loci (e.g. SNPs and microsatellites) and (ii) to check how fast estimates of heterozygosity based on increasing numbers of loci converge to the most precise estimate available (which uses all loci). Theoretical predictions can be obtained for (i) and (ii). The first is simply the correlation between heterozygosities at two non-overlapping sets of loci; it can be simply computed based on the assumption (underlying the general-effect model) that this correlation emerges only as a result of the common dependency of heterozygosities in both sets of markers on the extent of inbreeding. Therefore

$$r^2(H_A^{**}, H_B^{**}) = \frac{(g_2)^2}{\sigma^2(H_A^{**})\sigma^2(H_B^{**})} = r^2(H_A^{**}, f)r^2(H_B^{**}, f) \qquad \text{Eq.6}$$

The quantity relevant to point (ii) is the correlation between heterozygosity at L loci and heterozygosity at a subset (A) of these loci, which contains a fraction p_A of the loci. As correlations are insensitive to scaling by a constant, we can work here with raw heterozygosities H (not standardized heterozygosity H^{**}). Using raw heterozygosity, total heterozygosity H is the sum of heterozygosity at the A loci (H_A) and at the remaining loci (H_B). Thus

$$r(H_A^{**}, H^{**}) = r(H_A, H) = \frac{COV(H_A,H)}{\sqrt{\sigma^2(H_A)\sigma^2(H)}} = \frac{COV(H_A,H_A+H_B)}{\sqrt{\sigma^2(H_A)\sigma^2(H)}} = \sqrt{\frac{\sigma^2(H_A)}{\sigma^2(H)}} + \sqrt{\frac{\sigma^2(H_B)}{\sigma^2(H)}} r(H_A^{**}, H_B^{**})$$

Eq.7

where $\sigma^2(H_A)$ is the numerator of Eq.3 ($\sigma^2(H_B)$ and $\sigma^2(H)$ can be computed similarly, making the summations over the appropriate sets of loci). Assuming that all loci have the same heterozygosity h one obtains, after some algebra

$$r^2(H_A^{**}, H^{**}) = p_A\left(1 + (1 - p_A)\frac{L\bar{h}g_2}{1-\bar{h}+\bar{h}g_2(Lp_A-1)}\right) \quad \text{Eq.8}$$

In this formula one can distinguish two terms: the first is simply the proportion of loci included in the subset (p_A), and reflects the fact that subset A will always capture a proportion of the variance in total heterozygosity at all loci because they are part of the total. The ability of the A subset to inform about the other loci (hence about the genome in general) is reflected by the second term which relies on the existence of identity disequilibrium (g_2): through this disequilibrium, the loci in A inform about the state of other loci and hence capture a more than proportional share of total variance in heterozygosity.

3.3 Methods

3.3.1 Study Populations

We examined patterns of heterozygosity in bighorn sheep at National Bison Range (Montana, USA; NBR) and at Ram Mountain (Alberta, Canada; RM). In both populations long-term studies follow individuals throughout their lives. The National Bison Range population was founded in 1922 via translocation of 12 individuals from Banff National Park (Alberta, Canada). Individual monitoring started in 1979, with genetic sampling beginning in 1988. Beginning in 1985, NBR experienced a 'genetic rescue' via intentional translocation of 15 individuals from neighboring populations to prevent local extinction after years of isolation and inbreeding (Hogg *et al.* 2006; Miller *et al.* 2012). Prior to the introduction census size and growth rate had been steadily declining (average census size of 48 sheep between 1922 and 1985). Following the supplementation there has been an increase both in census size (142 sheep in late 2012) and genetic diversity (Hogg *et al.* 2006; Miller *et al.* 2012).

In contrast, Ram Mountain is a native population in which individual-based monitoring began in 1972 with genetic sampling starting in 1988 (Coltman *et al.* 2002; Jorgenson *et al.* 1997). Between 1988 and 2010 census size fluctuated between 38 and 210 sheep, declining

recently due to low recruitment (Jorgenson *et al.* 1997) and cougar (*Puma concolor*) predation (Festa-Bianchet *et al.* 2006).

3.3.2 Marker Genotyping and Selection

SNP genotypes used in this study were generated by typing 27 individuals from NBR and 50 from RM on the OvineSNP50 BeadChip, yielding 853 variable loci. For this study, we excluded loci that were genotyped in less than 90% of individuals in both populations (N = 38), had less than 5% minor allele frequency (MAF; N = 392), and did not conform to Hardy–Weinberg expectations following a Bonferroni correction (N = 2). This resulted in a final dataset of 412 SNPs (Supplementary Table 3-S1). We included loci polymorphic in one population but monomorphic in the other. Note that due to their discovery via cross-species application of the OvineSNP50 BeadChip the SNPs used in this study are widely distributed in the genome, and mostly intergenic as few are expected to be in or near genes based on annotation of the domestic sheep genome (Miller *et al.* 2011).

Microsatellite loci used in analyses were a subset (N = 200; Supplementary Table 3-S2) of those used to construct a bighorn sheep linkage map (Poissant *et al.* 2010). Primer information and PCR conditions for the markers can be found in Poissant et al. (2010; 2009) and references therein. All loci conformed to Hardy–Weinberg expectations following a Bonferroni correction. Loci were retained only if they were genotyped in both populations and had less than 25% missing genotypes in the samples from either population. For both marker sets loci on the X chromosome were excluded.

In total, 26 individuals from NBR and 48 from RM were genotyped at both sets of markers and included in subsequent analyses. The individuals from NBR were born between 1981 and 2004 and include descendants of the original founders of the population (N = 4), transplanted individuals (N = 2), and their progeny (N = 20).

3.3.3 Statistical Analyses

We calculated individual standardized multilocus heterozygosity (stMLH) following Coltman *et al.* (1999). The relationship between individual stMLH from each marker set was assessed using reduced major axis regression with 1000 jackknife iterations, as implemented in RMA version 1.21 (Bohonak & van der Linde 2004). Reduced major axis regression was chosen to account for the uncertainty associated with stMLH measures used as both dependent and independent variables. For resampling tests, 100 random subsets of markers were sampled without replacement from the full datasets using the "sample" function in R version 2.13.0 (R Development Core Team 2005). For microsatellites subsets of 5, 10, 20, 30, 50, 75, 100, 125, 150, and 175 loci were extracted, while for SNPs the subsets consisted of 20, 50, 75, 100, 150, 200, 250, 300, 350, and 400 loci. We then calculated stMLH for each subset using a custom Perl script. The coefficients of determination (r^2) were compared between the stMLH calculated for each subset and a total stMLH calculated from the concatenation of the SNP and microsatellite datasets (all 612 loci). In addition, we calculated r^2 between each subset and total stMLH for all loci of their respective marker type. We then compared these results to the theoretical predictions described in the previous section.

3.3.4 Estimates of Identity Disequilibrium & Expected Power to Detect HFCs

To measure ID we used the program RMES (David *et al.* 2007) to calculate the g_2 statistic. Significant covariance can be attributed to inbreeding, admixture, or a bottleneck (David *et al.* 2007; Szulkin *et al.* 2010). Assessment of significant levels of ID ($g_2 > 0$) utilized 1000 resampling iterations. Calculations were performed for both populations on the full microsatellite set, the full SNP set, the concatenated marker set, and all marker subsets used in the stMLH resampling calculations. We also examined the effect of the number of individuals sampled on the accuracy of g_2 estimates. For this analysis we bootstrapped both the full

microsatellite and full SNP datasets in each population, generating 100 replicates containing different numbers of individuals. Replicates contained 5, 10, 20, 30, 50, or 100 individuals.

To explore the power of different marker sets to detect HFCs we calculated the expected correlation between *f* and stMLH using Eq. 3 based on empirical estimation of variance in heterozygosity as well as its theoretical value based only on the number and average heterozygosity of the markers (Eq. 5). Eq. 5 has the advantage that it can be applied to assess the power of a study before actually performing it, as it requires only approximated parameter values. As with our estimates of g_2 we calculated $r^2(f, h)$ for the full SNP, microsatellite, and concatenated data sets, as well as all subsets. When the estimate of g_2 was negative we set $r^2(f, h)$ to 0.

3.4 Results

3.4.1 Summary Statistics of Markers

In NBR, the average (± SD) MAF for SNP loci was 0.197 (± 0.160), and average observed heterozygosity (Ho) was 0.279 (± 0.202). In RM, average MAF was 0.212 (± 0.151), and Ho 0.292 (± 0.178). For microsatellite loci, Ho was 0.643 (± 0.161), and number of alleles per locus ranged from 2 to 9 (average 4.39 ± 1.43) in NBR, while in RM Ho was 0.610 (± 0.157) and the number of alleles per locus ranged from 2 to 10 (average 4.21 ± 1.48).

3.4.2 Estimates of Identity Disequilibrium & Expected Power to Detect HFCs

All estimates of g_2 based on total marker sets were greater than zero for both populations ($p < 0.001$; Table 3-1). However, g_2 was much stronger in NBR than RM. Across both populations, the full SNP set produced higher estimates of g_2 than the full microsatellite set, and the combined datasets were intermediate. Our subsampling analyses showed that average values of g_2 on par with the genome-wide estimates of that same marker type were obtained even when

few markers were considered (Figure 3-1). However, there was considerable variation around these estimates. For example, in RM the SD estimates were larger than the average values of g_2 when fewer than 75 microsatellites or 150 SNPs were examined.

Bootstrapping the full datasets similarly showed that a stable average value of g_2 can be estimated with small sample sizes, however larger sample sizes increase the precision (Table 3-2). One might have expected the effects of increasing sample size to be more apparent in NBR, where there are a few highly inbred individuals and the chances of sampling them would therefore lead to large SD around the estimates of g_2. However, when scaled as a percentage of the average estimate, the effect of sampling is greater for RM than NBR. Even when 100 individuals were assumed SD values representing >14% of the average g_2 estimate in NBR and >23% in RM were seen.

Expected r^2 between *f* and stMLH for the various full marker sets were stronger in NBR than RM (Table 3-1). In both populations the strongest r^2 was seen from the combined marker dataset. The expected r^2 increased and the variation around estimates decreased as the number of markers increased (Figure 3-2). There was an apparent upward bias of the r^2 between *f* and stMLH when measured by the full marker sets compared to the subsets. This bias is likely an artifact given that there is only one estimate based on the full marker sets (rather than 100 permutations) and that the correlation is based on a g_2 statistic that still has error associated with it (Table 3-1).

3.4.3 Correlations Between Marker Types and Among Subsets

Individual stMLH was significantly positively correlated between marker types in both populations (Figure 3-3). However, the correlation was much stronger in NBR (NBR $r = 0.954$, $t_{24} = 15.557$, $p << 0.001$; RM $r = 0.370$, $t_{46} = 2.703$, $p = 0.001$). Based on Eq. 6 the expected correlations were $r = 1.14$ and $r = 0.465$ for NBR and RM respectively. Slight differences between the predicted and observed values (as well as the $r > 1$) likely arise because the

combined g_2 value is measured with error (Table 3-1). To ensure that the larger sample size in RM was not the main driver of the difference in correlation between RM and NBR we jackknifed our data from RM, resampling 100 sets of 26 individuals which yielded an average correlation of 0.379 ± 0.107 (SD) indicating that sample size was not driving the observed patterns.

In our resampling analyses, r^2 values were stronger in NBR than RM regardless of the number or type of markers examined (Figure 3-4). In both the SNP and microsatellite datasets, the correlation with the total measure of stMLH strengthened with increasing number of markers (Figure 3-4A & 3-4C). In NBR, correlations between microsatellite subsets and total stMLH were higher than those for an equal number of SNPs (Figure 3-4C). However, an asymptote of strong correlation ($r^2 = 0.9$) was reached with as few as 75 microsatellites or 200 SNPs. These differences between marker types disappeared when marker subsets were compared to stMLH of only that same marker type and scaled as a proportion of the total number of either SNPs or microsatellites considered (Figure 3-4D). For RM, when equal numbers of markers were compared, microsatellites produced a marginally higher average correlation to total stMLH than SNPs (Figure 3-4A) though these differences were not significant. The two marker types gave near-identical correlations when marker subsets were compared to stMLH values from only the same marker type (Figure 3-4B). Average correlations to marker specific estimates of stMLH were stronger than when subsets were compared to total stMLH (average increase of 0.10 for SNPs and 0.19 for microsatellites). For both populations, empirical correlations exceeded the null expectations and closely paralleled predicted correlations (Figure 3-4). The exception was in RM where microsatellites were less correlated to genome-wide stMLH than was expected.

3.5 Discussion

3.5.1 Influence of Population History

As expected, across all analyses the strength of association between marker heterozygosities and the expected ability to detect HFCs was highly dependent on the demographic history of the population (Ljungqvist *et al.* 2010; Szulkin *et al.* 2010). Bighorn sheep tend to be philopatric and have a highly polygynous mating system, where a few dominant males sire the majority of offspring (Coltman *et al.* 2002; Hogg & Forbes 1997). Thus, even in a native population such as RM a certain level of identity disequilibrium is to be expected. In addition, RM is relatively isolated and rarely receives immigrants (Rioux-Paquette *et al.* 2010), furthering the likelihood of non-random association of alleles due to inbreeding. Disequilibrium is even more likely in NBR given its population history. Descendants of NBR founders are expected to have low overall genetic diversity after years of inbreeding, translocated individuals from neighboring herds will have relatively higher levels of diversity, and their progeny are expected to have the highest heterozygosity as a result of the admixture between the founder and translocated individuals (Hogg *et al.* 2006; Miller *et al.* 2012).

Theory predicts that population history as well as mating system (i.e. partial inbreeding or selfing), as summarized through the g_2 parameter, determines how well heterozygosity at a set of markers reflects heterozygosity at other loci; and by extension, genomic heterozygosity and inbreeding (Szulkin et al. 2010, our equations 3, 5 and 6). Theoretical predictions correctly match the observed correlations between heterozygosity at SNPs and heterozygosity at microsatellites in our data. Though significant correlations were seen in both populations, it was much tighter in NBR (Figure 3-3). In contrast, Väli *et al*. (2008) found no significant correlation between individual heterozygosity at SNPs and microsatellites at the level of the individual in four populations of wolves (*Canis lupus*) and one of coyotes (*C. latrans*). However, both of these species have high dispersal rates and large effective population sizes (Pilot *et al.* 2006) which

may not allow for such correlations to develop. In addition, Väli *et al*. (2008) used only 10–17 microsatellites and 25–51 SNPs in 10 introns, which our results suggest may not have had the power to detect an association in a population with low g_2 (e.g. 0.001-0.005). The contrasting effect of demographic history is equally apparent when trying to estimate genome-wide heterozygosity from a subset of markers (Figure 3-4).

3.5.2 The Number of Markers, Not Marker Type, Influences Correlations in stMLH

In NBR microsatellites were more highly correlated to total stMLH than SNPs when equal numbers of markers were compared: 20 microsatellites predicted inbreeding as well as 75 SNPs – as expected given their higher average heterozygosity (0.626 compared to 0.275) and our theoretical equations (Eq. 8). However, even a small number of either type of marker was highly correlated with inbreeding level, and with total heterozygosity at all markers, because of the high g_2. The situation was slightly different in RM. Here SNPs and microsatellites gave essentially the same correlations to our total measure of stMLH when equal numbers of markers were compared, though the microsatellite subsets are less correlated than predicted by Eq. 8. Given the low g_2 a much larger number of loci (microsatellites or SNPs) is needed to adequately measure inbreeding in RM. However, we still observe, as in NBR, that a lower number of microsatellites is required compared to SNPs to reach a given accuracy, consistent with theoretical expectations (Figure 3-4).

In short, microsatellites are more informative than SNPs because they have higher genetic diversity per locus, however, to find the most efficient strategy one must consider that it is now becoming technically easier to develop and type a large number of SNPs than an equivalent number of microsatellites (Baird *et al.* 2008; Davey *et al.* 2011; Guichoux *et al.* 2011). These results agree with previous theoretical and empirical studies which suggested that highly heterozygous multi-allelic markers will have higher correlation between MLH and genome-wide

heterozygosity than bi-allelic ones (Forstmeier *et al.* 2012; Ljungqvist *et al.* 2010; Slate *et al.* 2004; Online Appendix 2 in Szulkin *et al.* 2010) except in special cases (Tsitrone *et al.* 2001).

One factor that could seem to limit the robustness of our conclusions is that correlations to total heterozygosity could be biased because the dataset contains a larger number of SNPs than microsatellites (421 vs. 207 loci). We do not feel that this is a problem for several reasons. First, individual heterozygosity between the two marker types was correlated (Figure 3-3) and should therefore show the same patterns no matter the ratio of loci examined. Also, if the relative proportion of markers biased our estimates one would expect correlations to the measure of total stMLH to be constrained by the marker's abundance in the total dataset; for example, $r^2 < 0.33$ for microsatellites. Figure 3-4 shows that this is not the case: both sets of loci quickly rose to essentially perfect correlations in NBR, and increased to levels well above their relative proportions in RM. Finally, in general SNPs are more abundant in a genome than microsatellites, and though the ratio is not 2:1, our dataset reflects this difference in abundance. Together these points suggest that there should not be any substantial bias based on the relative composition of the markers.

3.5.3 Identity Disequilibrium and Expected Correlations Between f and stMLH

On average, modest numbers of markers seemed to accurately estimate the levels of ID, but variability was high when only a few markers were considered. Several recent studies have estimated ID for both wild and captive populations (Borrell *et al.* 2011; Grueber *et al.* 2011; Jourdan-Pineau *et al.* 2012; Küpper *et al.* 2010; Olano-Marin *et al.* 2011; Wetzel *et al.* 2012). In all but one case (Olano-Marin *et al.* 2011) these estimates were non-significant, even in the highly endangered takahe (*Porphyrio hochstetteri*) which had experienced a bottleneck reducing the population to 17 individuals (Grueber *et al.* 2011). However, all of the studies showing non-significant results used between 7 and 24 microsatellite loci (average 18.2), which we have shown can give an inaccurate picture of diversity depending on the specific loci examined and

the population history. In contrast, Olano-Marin et al. (2011) used 80 microsatellites to study a wild population of blue tits (*Cyanistes caeruleus*).

3.5.4 Time to Move Towards SNPs for Use in HFCs?

Our results suggest that SNPs are more suited for HFCs than previously thought. First, significant correlations between individual stMLH at SNPs and microsatellites indicate that there is no loss of information when using a bi-allelic rather than a multi-allelic marker to estimate heterozygosity. Second, SNPs may be more suited for examining the various hypotheses that underlie HFCs, such as direct effects and local effects (Hansson & Westerberg 2002), given that they are more abundant in the genome than microsatellites, can be in coding regions, and can be more readily genotyped at ultra high density. However, to perform the same job as microsatellites, SNP's need to be more numerous as they are on average less heterozygous. The exact number of markers needed to obtain high correlations depended heavily on the demographic history of the population. For populations such as NBR that have experienced a severe bottleneck or admixture, fewer markers will be needed to obtain significant g_2 estimates and detect HFCs. In this situation it will be more beneficial for researchers to type additional individuals, getting an accurate estimate of the variance in inbreeding and fitness, rather than typing more markers in fewer individuals. For populations with no history of a bottleneck or severe inbreeding, such as RM, significantly more markers will be needed to accurately estimate genome-wide heterozygosity. It is then questionable whether lots of effort should be invested into typing the required number of markers, whatever their type, given that the signal (HFCs and inbreeding) is necessarily very weak in such situations. Our equations can be directly used to assess the required number of loci needed to achieve a given accuracy in the measure of inbreeding (or genomic heterozygosity) provided a value of g_2 is available (or can be estimated from preliminary data with fewer loci). For example, in RM using the combined g_2 estimate from all markers (the most precise value available), Eq. 5 can be used to predict that no less than 1732

microsatellites or 6469 SNPs would be needed for stMLH to be highly correlated (r^2=0.9) to inbreeding.

Although SNPs are still moderately expensive to develop for wild species, new methods allow rapid discovery of large marker panels (Baird *et al.* 2008; Davey *et al.* 2011) at diminishing costs. Once discovered, new technologies, such as array-based genotyping assays (Shen *et al.* 2005) and genotype-by-sequencing approaches (Baird *et al.* 2008) will allow for SNP datasets to be rapidly genotyped in many individuals. Comparable methods for scaling up the genotyping of microsatellites are not currently available.

While we were unable to directly compare SNPs and microsatellites in terms of their ability to detect HFCs for specific traits due to the small number of individuals genotyped, we now have an indication of the number of markers that would be needed for future efforts. More broadly, our results highlight that accurate calculations of stMLH, assessment of ID, and thereby detection of HFCs will likely require a large number of markers, be they SNPs or microsatellites. However, the exact number is highly dependent on the demographic history or mating system of the population being examined, the key parameter being the identity disequilibrium (which can be estimated with g_2). Efforts should be directed towards precisely estimating this parameter in natural populations. To this end, assuming that the number of loci available in population genetic studies will continue to increase, the main limitation will become the sample size in terms of numbers of individuals.

Table 3 - 1 Estimate of identity disequilibrium (g_2) and expected r^2 between inbreeding (f) and stMLH (H_A) for the different full marker sets in each population of sheep**

		Average stMLH	SD of stMLH	g_2	SD of g_2	Expected $r^2(H_A**, f)$
NBR						
	Microsatellites	1.002549	0.225138	0.055933	0.027856	0.95532
	SNPs	0.998982	0.239685	0.066346	0.031269	0.91574
	Combined	0.999913	0.251398	0.061671	0.029483	0.96283
RM						
	Microsatellites	1.000044	0.073378	0.002375	0.005465	0.42736
	SNPs	1.000017	0.100141	0.005071	0.003672	0.46512
	Combined	1.001711	0.072890	0.003418	0.004342	0.58123

Table 3 - 2 Average estimates of g_2 for different sample sizes in each population. Averages are based on 100 bootstrap replicates

		No. Individuals	Average g_2	SD of g_2
NBR	Microsatellite			
		5	0.0882	0.0694
		10	0.0616	0.0343
		20	0.0549	0.0240
		30	0.0581	0.0200
		50	0.0564	0.0150
		75	0.0573	0.0122
		100	0.0535	0.0092
	SNP			
		5	0.0840	0.0563
		10	0.0648	0.0378
		20	0.0623	0.0192
		30	0.0645	0.0186
		50	0.0671	0.0154
		75	0.0645	0.0117
		100	0.0660	0.0097
RM	Microsatellite			
		5	0.0026	0.0044
		10	0.0016	0.0021
		20	0.0016	0.0016
		30	0.0017	0.0011
		50	0.0017	0.0009
		75	0.0016	0.0008
		100	0.0016	0.0006
	SNP			
		5	0.0052	0.0066
		10	0.0046	0.0040
		20	0.0036	0.0024
		30	0.0040	0.0018
		50	0.0043	0.0015
		75	0.0043	0.0013
		100	0.0041	0.0012

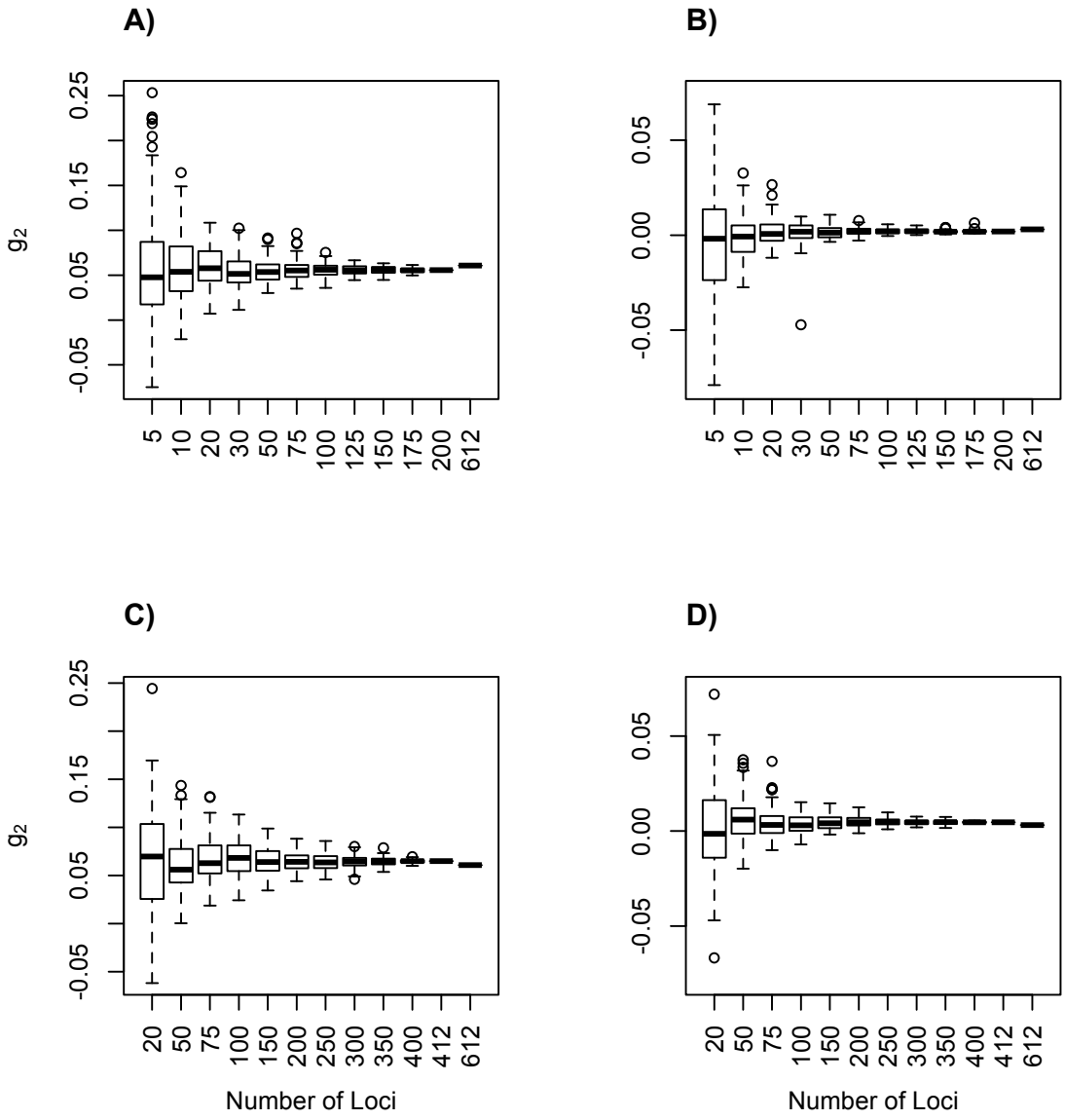


Figure 3 - 1 Box plots showing the average level of identity disequilibrium for the different marker subsets

Each subset was generated by sampling markers from the total dataset 100 times. Plots A and C show correlation in NBR at microsatellites (A) and SNPs (C), while plots B and D show correlations in RM at microsatellites (B) and SNPs (D)

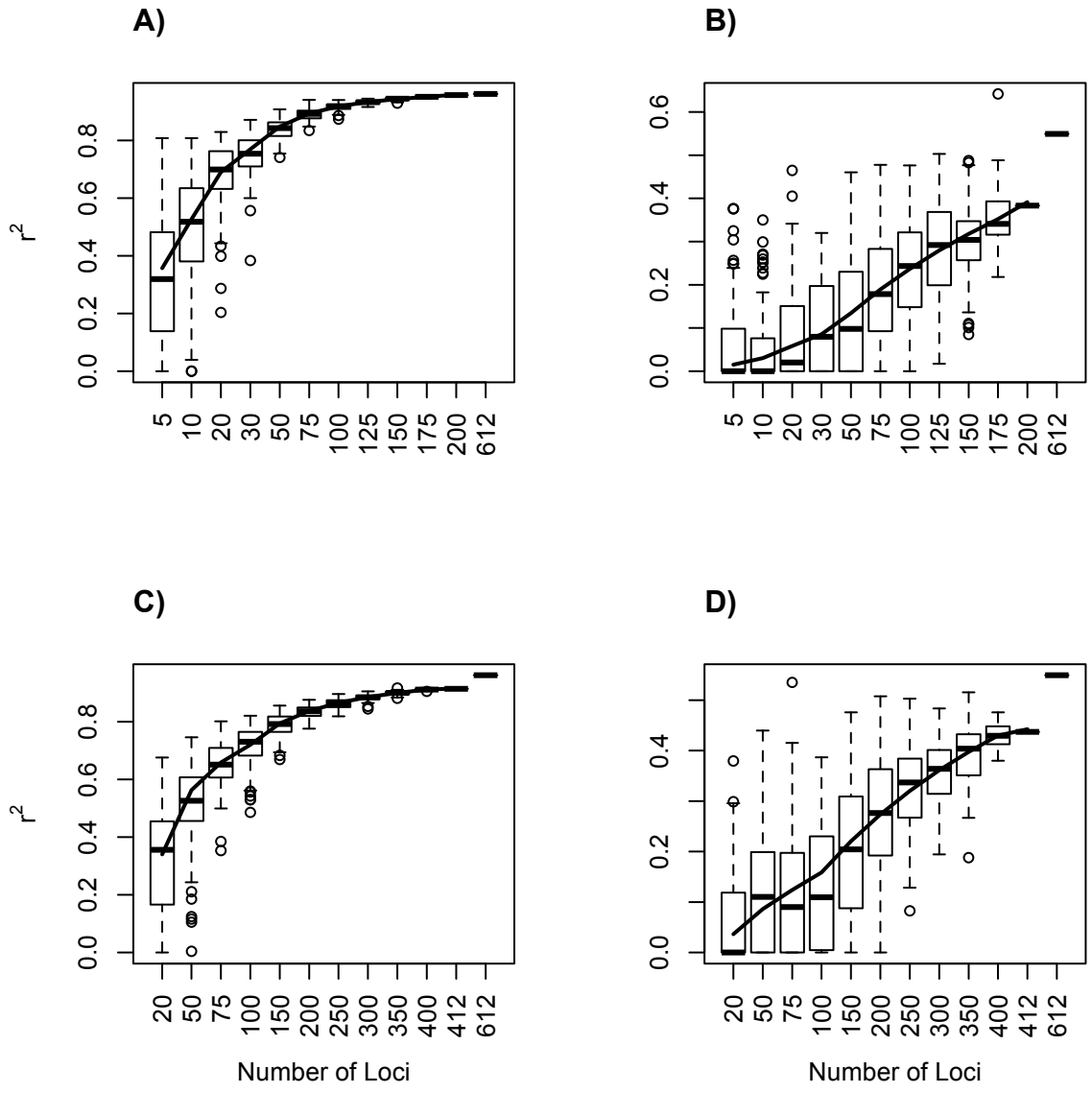


Figure 3 - 2 Box plots showing the average expected r^2 between inbreeding and heterozygosity for the different marker subsets

Plots A and C show correlation in NBR at microsatellites (A) and SNPs (C), while plots B and D show correlations in RM at microsatellites (B) and SNPs (D). Solid lines show predicted correlations based on Equation 5

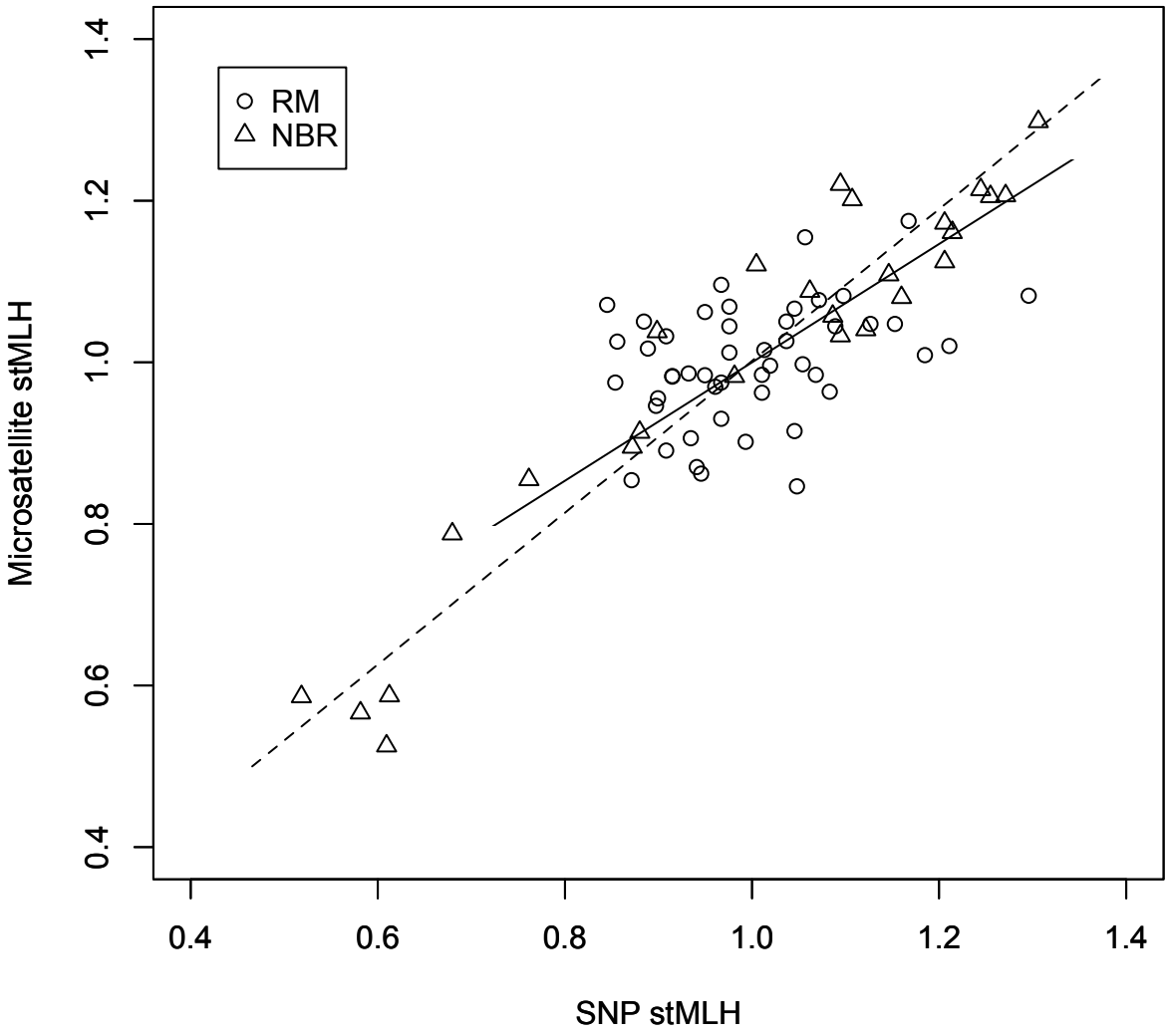


Figure 3 - 3 Correlation between individual heterozygosity at SNPs and microsatellites in RM and NBR

Reduced major axis regression lines are shown: y = 0.7327x + 0.2673 for RM (solid line); y = 0.9393x + 0.0642 for NBR (dashed line)

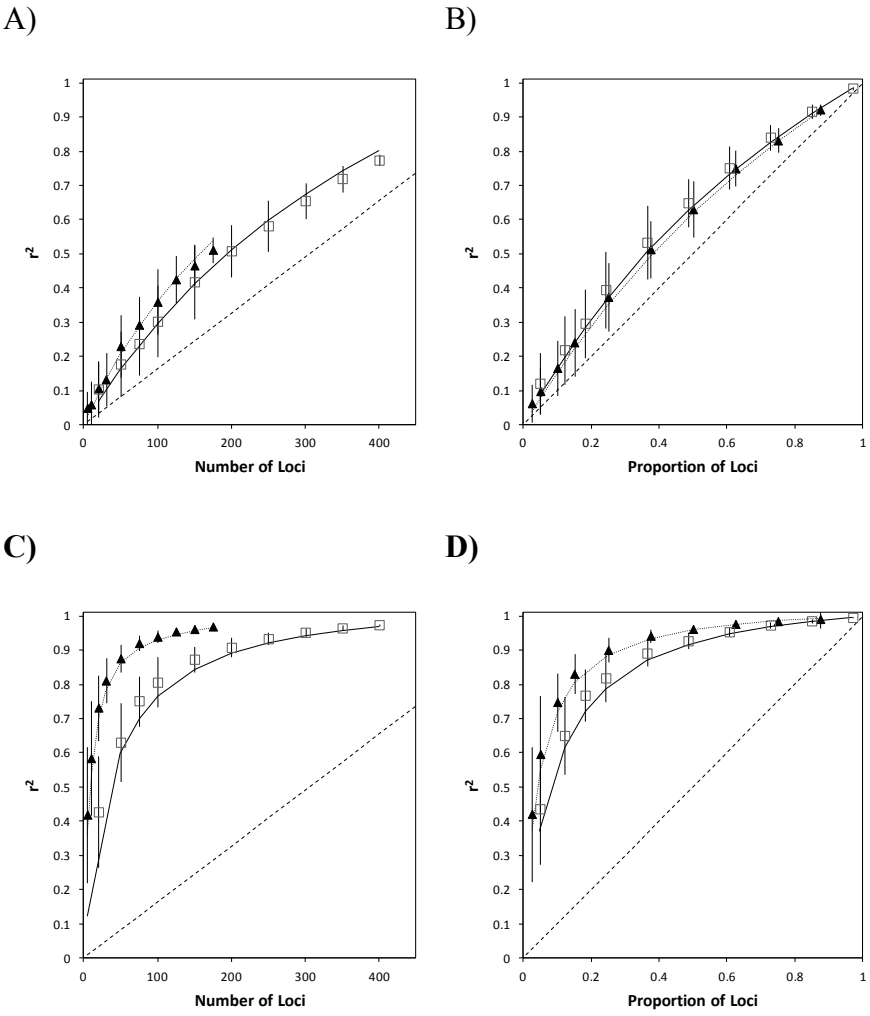


Figure 3 - 4 Average r^2 between marker subset stMLH and genome-wide stMLH
Each subset was generated by sampling markers from the total dataset 100 times; error bars show standard deviations. Plots A and C show correlations for SNPs (open squares) and microsatellites (filled triangles) when all 612 loci are considered in RM and NBR respectively. Plots B and D show correlations when subsets are compared to stMLH exclusively from the same marker type in RM (B) and NBR (D). Note that the x-axis is now scaled as a proportion of the total number of either SNPs or microsatellites. Predicted correlations based on Equation 8 are shown for SNPs (solid lines) and microsatellites (dotted lines), dashed lines show predicted correlations among subsets in the absence of identity disequilibrium.

3.6 Bibliography

Baird NA, Etter PD, Atwood TS*, et al.* (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **3**, e3376.

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* **13**, 3021-3031.

Bierne N, Tsitrone A, David P (2000) An inbreeding model of associative overdominance during a population bottleneck. *Genetics* **155**, 1981-1990.

Bohonak AJ, van der Linde K (2004) RMA: Software for Reduced Major Axis regression, Java version. *Website: http://www.kimvdlinde.com/professional/rma.html*.

Borrell YJ, Carleos CE, Sánchez JA*, et al.* (2011) Heterozygosity–fitness correlations in the gilthead sea bream *Sparus aurata* using microsatellite loci from unknown and gene-rich genomic locations. *Journal of Fish Biology* **79**, 1111–1129.

Brandstrom M, Ellegren H (2008) Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome Research* **18**, 881-887.

Chakraborty R (1981) The distribution of the number of heterozygous loci in an individual in natural-populations. *Genetics* **98**, 461-466.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology* **18**, 2746-2765.

Coltman D, Pilkington J, Smith J, Pemberton J (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53**, 1259-1267.

Coltman D, Slate J (2003) Microsatellite measures of inbreeding: A meta-analysis. *Evolution* **57**, 971-983.

Coltman DW, Festa-Bianchet M, Jorgenson JT, Strobeck C (2002) Age-dependent sexual selection in bighorn rams. *Proceedings of the Royal Society B-Biological Sciences* **269**, 165-172.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* **12**, 499-510.

David P, Pujol B, Viard F, Castella V, Goudet J (2007) Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology* **16**, 2474-2487.

DeWoody Y, DeWoody J (2005) On the estimation of genome-wide heterozygosity using molecular markers. *Journal of Heredity* **96**, 85-88.

Festa-Bianchet M, Coulson T, Gaillard JM, Hogg JT, Pelletier F (2006) Stochastic predation events and population persistence in bighorn sheep. *Proceedings of the Royal Society B-Biological Sciences* **273**, 1537-1543.

Forstmeier W, Schielzeth H, Mueller JC, Ellegren H, Kempenaers B (2012) Heterozygosity–fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Molecular Ecology* **21**, 3237–3249.

Grueber C, Waters J, Jamieson I (2011) The imprecision of heterozygosity-fitness correlations hinders the detection of inbreeding and inbreeding depression in a threatened species. *Molecular Ecology* **20**, 67-79.

Grueber CE, Wallis GP, Jamieson IG (2008) Heterozygosity–fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology* **17**, 3978-3984.

Guichoux E, Lagache L, Wagner S*, et al.* (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**, 591-611.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* **11**, 2467-2474.

Hogg JT, Forbes SH (1997) Mating in bighorn sheep: frequent male reproduction via a high-risk "unconventional" tactic. *Behavioral Ecology and Sociobiology* **41**, 33-48.

Hogg JT, Forbes SH, Steele BM, Luikart G (2006) Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B: Biological Sciences* **273**, 1491-1499.

Jorgenson JT, Festa-Bianchet M, Gaillard J-M, Wishart WD (1997) Effects of age, sex, disease, and density on survival of bighorn sheep. *Ecology* **78**, 1019-1032.

Jourdan-Pineau H, David P, Crochet P-A (2012) Phenotypic plasticity allows the Mediterranean parsley frog *Pelodytes punctatus* to exploit two temporal niches under continuous gene flow. *Molecular Ecology* **21**, 876–886.

Küpper C, Kosztolányi A, Augustin J*, et al.* (2010) Heterozygosity-fitness correlations of conserved microsatellite markers in Kentish plovers *Charadrius alexandrinus*. *Molecular Ecology* **19**, 5172–5185.

Ljungqvist M, ÅKesson M, Hansson B (2010) Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al. (2008). *Molecular Ecology* **19**, 851-855.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012) Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Molecular Ecology* **21**, 1583–1596.

Miller JM, Poissant J, Kijas J, Coltman DW, Consortium TISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Morin PA, Luikart G, Wayne RK, Grp SNPW (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* **19**, 208-216.

Olano-Marin J, Mueller JC, Kempenaers B (2011) Correlations between heterozygosity and reproductive success in the blue tit (*Cyanistes caeruleus*): an analysis of inbreeding and single locus effects. *Evolution* **65**, 3175–3194.

Pilot M, Jedrzejewski W, Branicki W*, et al.* (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology* **15**, 4533-4553.

Poissant J, Hogg JT, Davis CS*, et al.* (2010) Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. *BMC Genomics* **11**, doi:10.1186/1471-2164-1111-1524.

Poissant J, Shafer ABA, Davis CS*, et al.* (2009) Genome-wide cross-amplification of domestic sheep microsatellites in bighorn sheep and mountain goats. *Molecular Ecology Resources* **9**, 1121-1126.

R Development Core Team (2005) R: A language and environment for statistical computing, reference index version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria.

Rioux-Paquette E, Festa-Bianchet M, Coltman D (2010) No inbreeding avoidance in an isolated population of bighorn sheep. *Animal Behaviour* **80**, 865-871.

Shen R, Fan J-B, Campbell D*, et al.* (2005) High-throughput SNP genotyping on universal bead arrays. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **573**, 70-82.

Slate J, David P, Dodds KG*, et al.* (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **93**, 255-265.

Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: A time for reapprasial. *Evolution* **64**, 1202-1217.

Tsitrone A, Rousset F, David P (2001) Heterosis, Marker Mutational Processes and Population Inbreeding History. *Genetics* **159**, 1845-1859.

Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* **17**, 3808-3817.

Wetzel DP, Stewart IRK, Westneat DF (2012) Heterozygosity predicts clutch and egg size but not plasticity in a house sparrow population with no. *Molecular Ecology* **21**, 406-420.

Supplementary Information

Supplementary Table 3-S1: List of SNP loci used in this study.

Supplementary Table 3-S2: List of microsatellite loci used in this study.

All supplementary tables can be accessed at:

http://www.nature.com/hdy/journal/v112/n3/suppinfo/hdy201399s1.html

Chapter 4

ASSESSMENT OF IDENTITY DISEQUILIBRIUM AND ITS RELATION TO EMPIRICAL HETEROZYGOSITY FITNESS CORRELATIONS: A META- ANALYSIS

A version of this chapter has been published:

Joshua M. Miller and David W. Coltman. "Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: a meta- analysis." *Molecular Ecology* 23, no. 8 (2014): 1899-1909.

4.1 Introduction

Heterozygosity fitness correlations (HFCs) have become a prevalent tool in conservation genetics and evolutionary biology (Balloux *et al.* 2004; Chapman *et al.* 2009; Coltman & Slate 2003). In these analyses individual heterozygosity (as averaged over a number of loci) is often used as a proxy for inbreeding (Santure *et al.* 2010; Townsend & Jamieson 2013) and when associated with measures of fitness (such as survival or reproductive success) may reveal evidence of inbreeding depression. HFCs can be especially useful in situations where a direct measure of inbreeding (such as from a pedigree) is not available, as is often the case for wild and endangered species (Grueber *et al.* 2008; Klauke *et al.* 2013; Ruiz-Lopez *et al.* 2012).

Three processes are thought to potentially underlie HFCs (Hansson & Westerberg 2002). "Direct-effects" result when markers themselves have functional consequences and are directly linked to differences in fitness. "Local-effects" occur when the genotyped markers themselves may not have direct consequences on fitness, but rather are in linkage disequilibrium with variants that are. Finally, "general-effect" HFCs are attributed to an intrinsic benefit to being heterozygous and the heterozygosity of the typed markers is correlated to overall heterozygosity in the genome. The chances of detecting direct- or local-effect HFCs are relatively small, unless a substantial number of loci are considered, and thus most HFC studies to date have focused on examining general-effect HFCs.

Previous reviews of HFC studies, however, have highlighted that many either do not find correlations, or the effects they find are small (Britten 1996; Chapman *et al.* 2009; Coltman & Slate 2003; Szulkin *et al.* 2010). Some authors have argued that the power of HFC studies may be limited either by the demographic history of the population, or the inability of the markers used to reflect that history (Grueber *et al.* 2011b; Ljungqvist *et al.* 2010; Väli *et al.* 2008). The latter problem may be solved by considering large sets of markers (DeWoody & DeWoody 2005; Miller *et al.* 2014), a feat that is becoming increasingly feasible with simultaneous marker

discovery and genotyping (e.g. restriction site associated DNA sequencing or genotype-by-sequencing; Davey *et al.* 2011; Elshire *et al.* 2011) using high-throughput sequencing technology.

The issue of demographic history is potentially more complicated. One does not expect to detect general-effect HFCs unless there is variance in the level of inbreeding within a population (Slate *et al.* 2004; Szulkin *et al.* 2010). Such variance can be caused by demographic events (e.g. admixture, bottlenecks) or due to a non-random mating system (e.g. partial selfing) which cause similar changes in the nature of heterozygosity in a genome (David *et al.* 2007; Szulkin *et al.* 2010). Recent work has highlighted identity disequilibrium (ID) as one measure that may capture these differences in heterozygosity (Bierne *et al.* 2000; David 1998; David *et al.* 2007). ID is the covariance in heterozygosity among markers within individuals, which should reflect identity by descent (IBD) of those markers (Bierne *et al.* 2000; David *et al.* 2007; Szulkin *et al.* 2010). The magnitude of the covariance is affected by demographic history and mating systems. In the absence of ID the set of markers used in an HFC analysis are expected to only reflect their local genomic environment, thereby limiting a study to detecting only local-effect or direct-effect HFCs.

Two metrics have been developed to test for the presence of ID: heterozygosity-heterozygosity correlations (Balloux *et al.* 2004) and the g_2 statistic (David *et al.* 2007). Heterozygosity-heterozygosity correlations gauge ID by dividing a given set of loci into two even sets and then assessing the correlation in heterozygosity between them. This process is repeated hundreds of times to yield an average correlation and test its significance. In contrast, the g_2 statistic assesses the covariance of heterozygosity between markers standardized by their average heterozygosity, and its significance can be tested by permuting genotypes. Thus g_2 is a population parameter that summarizes the variance in inbreeding, rather than individual realized IBD (Ruiz-Lopez *et al.* 2012; Szulkin *et al.* 2010). Recent work has suggested the use of g_2 rather than heterozygosity-heterozygosity correlations is more appropriate in assessing ID as

calculation of the latter involves non-independent datasets and may be influenced by the properties of the exact marker set used. In contrast, calculation of g_2 uses all markers simultaneously, is expected to only be effected by demographic history, and is more central to HFC theory (Szulkin *et al.* 2010).

While a significant g_2 estimate may be a good indication of the ability to identify a general-effect HFC, Szulkin et al. (2010) noted that non-significant values of g_2 do not preclude presence or detection of an HFC. They assert that the phenotypic effects of inbreeding are more readily detected than correlations among marker heterozygosity (i.e. ID). Indeed, the paper by Kardos et al. (2014) highlighted this same point. Here the authors used simulations to assess the ability of g_2 to detect HFCs due to inbreeding over a range of demographic scenarios and with variable numbers of markers. They found that g_2 does give an indication of the strength of HFCs due to inbreeding (i.e. general-effects), however, often either the magnitude of a g_2 estimate or HFC calculation will not reach the level of statistical significance, even if an association was present.

Empirical studies are now starting to assess ID in wild populations (e.g. Borrell *et al.* 2011; Olano-Marin *et al.* 2011a; Wetzel *et al.* 2012). However, no thorough review of the metric, or its relation to observed effect sizes of HFCs, has been conducted. In this work we perform a meta-analysis to examine the effect of ID in empirical HFC studies. We first assess the magnitude of ID (as measured by g_2) in previously published HFC studies. We then look to see if the magnitude of g_2 is a predictor of the observed effect sizes in these studies. Finally, we use recently derived equations (Miller *et al.* 2014) to examine how much power the studies had to detect general-effect HFCs, and the number of markers that would have been needed to generate an expected high correlation ($r^2 = 0.9$) between marker genotypes and inbreeding.

4.2 Methods

4.2.1 Data Acquisition

For this meta-analysis we only considered studies of outbreeding vertebrates that conducted individual based HFC analyses. Studies that pooled individuals to create population averages were not considered. Studies also needed to report a summary statistic that we could convert into an effect size (see below). Our literature search began by considering all papers citing David et al. (2007) that reported a g_2 statistic. We then expanded our criterion by searching the Dryad Digital Repository (http://datadryad.org, last accessed November 2013) for HFC studies with publicly available datasets from which we could calculate a g_2 statistic. For this we used the key words "heterozygosity correlation", "fitness", and "inbreeding". In addition, we conducted a Google Scholar search (www.scholar.google.com; last accessed November 2013) for "Heterozygosity Fitness Correlations", limiting dates of the papers returned to those published since 2009. This filter was applied to increase the chance that genotypes would be archived online. However, these measures resulted in only a small number of studies (n = 18) with either online marker data or g_2 estimates. Thus, to increase our sample size we directly solicited data from the authors of studies cited in Chapman et al. (2009), as well as those citing Chapman et al. (2009) but that did not have marker data available online. In these cases authors were asked to provide either raw genotype data, tables of individual homozygosity at each locus, or to calculate g_2 themselves.

For each study we recorded the following covariates: number of individuals, number of loci, heterozygosity metric (i.e. multilocus heterozygosity, standardized multilocus heterozygosity (Coltman *et al.* 1999), homozygosity by locus (Aparicio *et al.* 2008), internal relatedness (Amos *et al.* 2001), or mean d^2 (Coulson *et al.* 1998)), average observed heterozygosity, and trait category. Internal relatedness is a measure of heterozygosity that takes into account the frequency of alleles in the population such that rare alleles have higher weight

(Amos *et al.* 2001) while mean d^2 measures the square difference in allele sizes at each locus and then averages those values over all loci (Coulson *et al.* 1998). Trait categories were chosen to match Coltman and Slate (2003) and Chapman et al. (2009): life history (e.g. survival, breeding success), morphological (e.g. size, symmetry), and physiological (e.g. parasite burden, hormone levels).

A common feature of many HFC studies is for the authors to examine multiple traits within a single population, e.g. egg size, clutch size, hatching success, and fledging success (Wetzel *et al.* 2012). It is also common for studies to examine more than one heterozygosity measure due to different rationales underlying their calculations (Amos *et al.* 2001; Aparicio *et al.* 2008; Coulson *et al.* 1998), though recently this practice has been discouraged (Chapman *et al.* 2009). If a study used multiple measures of heterozygosity, or reported multiple HFCs for different traits in the same population we recorded these as independent data points. Whenever possible, we updated the associated covariates to reflect the specific subset of individuals used for each HFC.

If reported, g_2 was recorded from the manuscript, otherwise it was calculated from available marker data via RMES (David *et al.* 2007) using 1000 permutations to test the significance of the measure. g_2 was recorded or calculated for each population and where possible, each sub-set of individuals identified by the authors. For example, if HFCs for male and female breeding success were reported independently we would calculate g_2 estimates for males and females separately.

4.2.2 Effect Size Calculations

We recorded the correlation coefficient (r) between heterozygosity and fitness measures. If r was not reported we recorded other summary statistics then transformed them to r following Coltman and Slate (2003). If t values were reported we used

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

where df is the degrees of freedom on which the test was based. If an F statistic was reported we used

$$r = \sqrt{\frac{F}{F + df_{error}}}$$

If a χ^2 value was reported we used

$$r = \sqrt{\frac{\chi^2}{n}}$$

where n is the sample size. For R^2 values the transformation was via

$$r = \sqrt{R^2 - \frac{p(1 - R^2)}{n - p - 1}}$$

where p is the number of parameters in the model. Finally, if only p values were reported we used

$$r = \sqrt{\frac{Z^2}{n}}$$

where Z^2 is the standard normal deviate of the p value. In cases where exact p-values were not given, but rather stated as >0.05 or "non-significant" we set p equal to 0.5. Directions for effect sizes were assigned *a posteriori* depending on the observed correlation in the study; for studies using internal relatedness or homozygosity by locus the sign of the correlation was reversed to match the other estimators.

Finally, r values were transformed into effect sizes for use in all subsequent analyses using the equation

$$Z_r = \frac{1}{2} log_e \left(\frac{1+r}{1-r} \right)$$

4.2.3 Univariate Analysis

As noted above, it is common for HFC studies to examine multiple fitness measures, multiple measures of heterozygosity, or combinations of both using the same sets of individuals and markers. Thus, inherent in many HFC studies is a level of pseudoreplication or non-independence. Previous meta-analyses of HFCs have dealt with this problem by running several models that each treat the data in a different way and then comparing the results (Chapman *et al.* 2009; Coltman & Slate 2003). These included ignoring the issue of pseudoreplication and treating each point as independent, averaging metrics within studies, or running mixed effect models including study, species, and population as random effects. In our case the issue of multiple fitness or heterozygosity measures within a study is coupled with the fact that often only one measure of g_2 was available, even if different subsets of individuals were used to calculate several HFCs. Thus, to avoid pseudoreplication we performed a univariate analysis considering a single effect size and g_2 estimate for each trait type within a population. A univariate analysis is a conservative approach and avoids the additional pseudoreplication that would come from having one g_2 estimate associated with multiple points in a multivariate mixed model. We chose to keep trait categories separate as one may expect HFCs to be more apparent in life history traits than morphological ones due to different selection pressures (directional versus stabilizing) that act on each (Chapman *et al.* 2009; Szulkin *et al.* 2010).

Weighted average effect sizes were calculated using the formula:

$$\bar{z} = \frac{\sum z_i x_i}{\sum x_i}$$

Where z_i is the i'th effect size and x_i is n-3 samples that went into calculating z_i. For example, Luquet et al. (2013) looked at six traits (male body size, female body size, male body condition,

female body condition, clutch mass, and egg size) in four population of European tree frog (*Hyla arborea*). We considered four of the traits to be morphological and two to be life history related. To obtain average effect sizes we grouped the individual effect sizes within each trait type and calculated a weighted average using the formula above. This was done for each population separately resulting in a total of eight effect size estimates. In cases where authors presented analysis of the same traits using multiple measures of heterozygosity we considered all measures of heterozygosity together (similar to MLH_{inc} of Chapman et al. (2009)). Grand mean effect sizes and their confidence intervals were back transformed to r for presentation of summary statistics. For populations where multiple g_2 values were available we calculated an average g_2.

Average effect size values were used in a linear model and assigned weights based on the variance of the values (as estimated by $1/(n_{avg}-3)$, where n_{avg} is the average number of individuals per population). All linear models were fit in R version 2.15.2 (R Development Core Team 2005).We included g_2, trait type, average number of loci, and average heterozygosity as covariates in the model. Model simplification proceeded using an information theoretic approach (Grueber *et al.* 2011a) as implemented in the package MuMIn version 1.9.5 (Bartoń 2009). We fit a maximal model containing all covariates and then assessed model differences with AICc (AIC values corrected for small sample sizes) values using the dredge function. In cases where ΔAICc scores did not differ by more than 2 we retained the simpler model.

We tested for evidence of publication bias using Egger's regression (Egger *et al.* 1997). Specifically, we regressed normalized study average effect size (i.e. average effect size divided by the standard error of the measurement) against average sample size. Publication bias is indicated by an intercept that is significantly different than zero (Sterne *et al.* 2005a; Sterne *et al.* 2005b).

4.2.4 Power of Studies to Detect Inbreeding

We investigated the predicted correlation between observed heterozygosity and inbreeding with equation 5 of Miller et al. (2014). Here, the correlation between inbreeding and heterozygosity is a function of the number of loci considered, their average heterozygosity (as reported in the manuscript or calculated from available marker data), and the magnitude of ID as measured by g_2. In cases where the estimate of g_2 was negative we set the correlation to 0. Finally, we calculated the number of markers that would have been needed for these populations to have a large correlation ($r^2 = 0.9$) between marker heterozygosity and genome-wide heterozygosity. To do this we modified equation 5 of Miller et al. (2014) solving for the number of loci (L_a)

$$L_a = \frac{-r^2 g_2 + r^2 - (r^2 g_2 h)}{h g_2 - r^2 h g_2}$$

where h is average observed heterozygosity of the markers.

4.3 Results

4.3.1 Data acquisition and Summary Statistics

The literature search and survey of researchers resulted in data from 50 studies (49 papers and 1 PhD thesis). Collectively these represent 45 species and 105 populations or subsets of individuals, for a total of 585 individual effect size estimates (Table 4-1). Study averaging within each trait type resulted in 129 effect size estimates which are summarized in Table 4-2. The average r value for each trait type was low (range 0.025 – 0.064) but none of the 95% confidence intervals cross zero (Table 4-2). Egger's regression indicated that the intercept was not significantly different than zero (intercept = 0.255, 95% CI = -0.066 – 0.575; Supplementary Figure 4-1).

We found a wide range of g_2 estimates (average ± SD = 0.007 ± 0.022; range -0.058 – 0.159) for the 129 effect sizes, only 26 of which were significantly different than zero (Figure 4-1). These 26 estimates had an average value of 0.025 (± 0.031). To better understand what may be driving this wide range of g_2 values we examined the relationship of g_2 to other covariates. However, we found that there was no relation to the average number of loci (Pearson's r = -0.006, t_{127} = -0.066, p = 0.948), average sample size (Pearson's r = -0.031, t_{127} = -0.349, p = 0.728), or year of publication (Pearson's r = -0.118, t_{127} = -1.342, p = 0.182).

4.3.2 Univariate Analysis

Our model selection criterion retained 4 models within 2 AICc values of the top ranked model. Three contained g_2 and one of the other covariates (either average heterozygosity, average number of markers, or the trait category), while the second highest ranked (ΔAICc = 0.24 from the top model) contained only g_2. This model containing only g_2 was retained for further inspection. Here, g_2 was positively correlated with average effect size (estimate ± SE = 2.66 ± 0.73; adjusted R^2 = 0.09) (Figure 4-2A). Although four outliers are visible in the graph, removing these points did not change the pattern observed (g_2 estimate ± SE = 1.54 ± 0.67; adjusted R^2 = 0.03) and so were retained for all further analyses.

Given that the majority of g_2 values were not significantly different than zero, we conducted two additional analyses. In the first we set all g_2 values that did not differ from zero to zero. This resulted in the same 4 models being selected with ΔAICc < 2, but now the top ranked model was the one containing only g_2 (ΔAICc = 1.36 from the next model). This new model still showed a positive relationship with average effect size, but the magnitude of the relationship and associated R^2 value were increased (estimate ± SE = 5.85 ± 1.02; adjusted R^2 = 0.20) (Figure 4-2B). In the second analysis we considered only the g_2 values that differed from zero. This returned only 2 models with ΔAICc < 2. The top model contained only g_2, and the second contained g_2 and average heterozygosity (ΔAICc = 1.41 from the top model). As with the two

previous models g_2 had a positive relationship with average effect size, but the magnitude of the estimate and the R^2 were increased yet again (estimate ± SE = 6.05 ± 1.63; adjusted R^2 = 0.34) (Figure 4-2C).

4.3.3 Power of Studies to Detect Inbreeding

After zeroing negative correlations (n = 32) average expected correlation between marker heterozygosity and inbreeding was 0.13, but a wide range of values were observed (0 – 0.82, Figure 4-3). Estimates of the number of loci needed to achieve an r^2 = 0.9 between heterozygosity and inbreeding were then based only on those values with g_2 estimates greater than zero (n = 95). Corresponding to the low average g_2 values in these populations a large number of loci (average ± SD = 5611 ± 8996; range 126 - 19642) would be needed to have correlations of 0.9. This distribution of values is shown in Figure 4-4.

4.4 Discussion

As with previous meta-analyses of HFC studies (Britten 1996; Chapman *et al.* 2009; Coltman & Slate 2003) we found that average effect sizes were very low. For life history and morphological traits we observed average effect sizes lower than the study average values reported by both Coltman and Slate (2003) and Chapman et al. (2009). However, unlike those two previous studies, the 95% CI for physiological traits did not overlap zero (Table 4-2). We found no evidence of funnel plot asymmetry (intercept was not different from zero) using Egger's regression, suggesting that in our sample, studies with small sample sizes are not overestimating effect sizes nor is there a publication bias against studies with negative results. This contrasts previous evidences for publication bias that were found by Coltman and Slate (2003) and Chapman et al. (2009).

Our assessment of g_2 from empirical HFC studies found a wide range of values, the majority of which were not significantly different than zero (Figure 4-2). We found no evidence that the magnitude of g_2 was associated with the average number of markers or average sample size. In addition, there was no trend in observed g_2 values over time, despite the fact that both the number of individuals and the number of loci considered in each HFC study was observed to have increased over time (data not shown).

The univariate analysis of study average effect sizes highlighted g_2 as the only variable consistently associated with average effect size. As predicted (Szulkin *et al.* 2010) the relationship between g_2 and effect size was positive, where studies that had large estimates of g_2 were able to explain more of the variance between heterozygosity and fitness. This pattern holds even with inclusion of both significant and non-significant g_2 estimates (Figure 4-2), but the relationship explains a small amount of variation. When we reduced the dataset to only estimates of g_2 that were significantly different than zero the association increased greatly (adjusted R^2 = 0.34 vs. 0.09). Thus, even when non-significant, g_2 is still an indicator of the presence of general effect HFCs, but there is a lot of noise around the estimate. These findings are in line with previous simulation studies (Kardos *et al.* 2014) as well as theoretical predictions on the influence of g_2 on the correlation between heterozygosity and fitness (Miller *et al.* 2014).

Part of the reason previous studies have reported small effect sizes could be that they were simply underpowered. We tested this hypothesis by looking at the expected correlation between observed heterozygosity and inbreeding (Miller *et al.* 2014) for each population. We found that many of the previous HFC studies had low expected correlation between heterozygosity and fitness (Figure 4-3). Coupled with the low g_2 estimates, lack of power may have been due to the small number of loci used. On average 40 loci were used in a study (though this number drops to 19 if we exclude the work of Forstmeier et al (2012) who genotyped 1359 SNPs in a population of zebra finches, *Taeniopygia guttata*). A much larger number of loci

(average = 5611) would have been needed to confidently explore general-effect HFCs in these populations (Figure 4-4).

Interestingly, the observation that all studies seem to be underpowered contradicts the thinking that the first study to publish a significant result will set an effect size threshold against which future studies have to match in order to be published, a so called "winner's curse" (Forstmeier & Schielzeth 2011; Nakaoka & Inoue 2009; Zollner & Pritchard 2007). This curse leads to inflation of effect size estimates and publication bias against any study that reports a null result or lower correlation. In contrast, our findings hold that average effect sizes were small, all the studies were underpowered, and there was no evidence for a publication bias. A supplementary analysis also showed that there was no trend in average effect size over time (estimate ± SE = -0.01 ± 0.00, p = 0.09), which stands in contrast to other studies (Jennions & Moller 2002). Taken together this indicates that the pool of HFC studies has managed to avoid the "winner's curse" and allows for robust inferences to be made in this study.

The observation that most HFC studies will need a much larger number of markers has been suggested by others (Balloux *et al.* 2004; Forstmeier *et al.* 2012; Kardos *et al.* 2014; Ljungqvist *et al.* 2010; Miller *et al.* 2014; Väli *et al.* 2008), and it is now becoming possible to generate such large marker sets by capitalizing on genomic technology (Angeloni *et al.* 2012; Ekblom & Galindo 2011). One point to consider is that all but one of the studies we included (Forstmeier *et al.* 2012) were based on microsatellite data. Moving forward we imagine that most new large-scale datasets will consist of SNP loci rather than microsatellites as SNP genotyping can be automated (Shen *et al.* 2005), while scaling up microsatellites genotyping is not currently possible. Thus, the estimates we present of the number of loci required for a robust HFC study likely represent a lower bound as, on average, microsatellites have higher heterozygosity than SNPs. Higher heterozygosity translates to higher expected correlation to genome-wide heterozygosity if the same number of loci are considered (Miller *et al.* 2014). We should also note that setting the desired correlation to genome-wide heterozygosity at $r^2 = 0.9$ necessitates

the need for more makers than smaller values would. It will be up to individual researcher to determine their desired level of correlation when performing similar calculations.

More broadly, use of large sets of loci will be a great boon to researchers investigating HFCs. Not only will they allow for confident exploration of general-effect HFCs, but also for detailed assessments of local-effects or direct-effects. Efforts to assess local-effects or direct-effects will be aided if the loci are anchored in the genome (via linkage mapping or alignment to a reference sequence) so that specific gene regions of interest can be identified (Olano-Marin *et al.* 2011b; Slate *et al.* 2009; Voegeli *et al.* 2013).

4.5 Conclusion

In this meta-analysis we assessed the magnitude of identity disequilibrium (as measured by the g_2 statistic) in 109 populations or analysis subsets from 50 previously published HFC studies. Across the majority of studies, g_2 values were not significantly different than zero. However, we found that the magnitude of g_2 was associated the average effect sizes observed in a population, even when non-significant g_2 estimates were considered. These low values of g_2 also translated into low expected correlations between heterozygosity and inbreeding, and suggested that many more markers would have been needed for robust HFC calculations.

However, we would argue that before researchers concern themselves with getting a large number of markers they should consider the demographic history of the population, and if it will be possible to detect general-effect HFCs. Such an assessment can be done with a small preliminary dataset to gauge ID at the outset. Though point estimates may not be precise (especially if the estimate is not different than 0) it can give a sense of the effect size that could be observed and the number of markers that will be needed for robust HFC calculations. We imagine that HFC analysis will remain a key toolset used by both researchers and wildlife managers; and with genomic techniques new avenues of research into local- or direct-effects may be on the horizon.

Table 4 - 1 Taxa, studies, number of populations, and effect for each trait type included in the meta-analysis. Trait types are either life-history (LH), morphological (M) or physiological (P)

Organism	Number of Populations or Subsets	Trait type (effect sizes reported)	Citations
Agamid lizard (*Ctenophorus ornatus*)	1	LH (6)	(LeBas 2002)
Antarctic fur seals (*Arctocephalus gazella*)	3	LH (1) M (5)	(Hoffman *et al.* 2010a; Hoffman *et al.* 2010b)
Arabian oryx (*Oryx leucoryx*)	1	LH (2)	(Marshall & Spalton 2000)
Attwater's Prairie-chicken (Tympanuchus cupido attwateri)	1	LH (1)	(Hammerly *et al.* 2013)
Black grouse (*Tetrao tetrix*)	1	LH (2)	(Hoglund *et al.* 2002)
Blue tits (*Cyanistes caeruleus*)	3	LH (24)	(Olano-Marin *et al.* 2011a)
Bluegill sunfish (*Lepomis macrochirus*)	1	LH (1) M (1) P (2)	(Neff 2004)
Bluethroat (*Luscinia svecica*)	1	LH (7) P (2)	(Fossøy *et al.* 2008)
Brent goose (*Branta bernicla hrota*)	1	LH (3)	(Harrison *et al.* 2011)
Cactus finch (*Geospiza scandens*)	1	LH (4)	(Markert *et al.* 2004)
Chinook salmon (*Oncorhynchus tshawytscha*)	1	LH (8)	(Heath *et al.* 2002)
Common shrew (*Sorex araneus*)	19	M (76)	(White & Searle 2008)
Crested newt (*Triturus cristatus*)	1	M (1)	(Herdegen *et al.* 2013)

El Oro parakeet (*Pyrrhura orcesi*)	1	LH (10)	(Klauke *et al.* 2013)
Elk (*Cervus canadensis*)	1	M (2)	(Hicks & Rachlow 2006)
European rabbit (*Oryctolagus cuniculus*)	1	LH (4) M (1)	(Gage *et al.* 2006)
European tree frog (*Hyla arborea*)	4	LH (8) M (16)	(Luquet *et al.* 2013)
Fire salamander (*Salamandra salamandra*)	1	LH (3)	(Caspers *et al.* 2014)
Great tit (*Parus major*)	3	LH (5) M (7) P (3)	(Otter *et al.* 2001; Voegeli *et al.* 2013)
Grey wolf (*Canis lupus*)	1	LH (1)	(Bensch *et al.* 2006)
Harbor seal (*Phoca vitulina*)	1	LH (2) M (2)	(Coltman *et al.* 1998)
House sparrow (*Passer domesticus*)	2	LH (14)	(Stewart & Westneat 2013; Wetzel *et al.* 2012)
Iberian lynx (*Lynx pardinus*)	2	LH (2)	(Ruiz-Lopez *et al.* 2012)
Kentish plovers (*Charadrius alexandrinus*)	2	LH (18)	(Küpper *et al.* 2010)
Lemon shark (*Negaprion brevirostris*)	1	LH (9) M (9)	(DiBattista *et al.* 2008)
Lipizzan horse (*Equus caballus*)	1	M (54)	(Curik *et al.* 2003)
Mandrill (*Mandrillus sphinx*)	1	LH (14)	(Charpentier *et al.* 2005)
Marmots (*Marmota marmota*)	1	LH (2)	(Da Silva *et al.* 2006)
Medium ground finch (*Geospiza fortis*)	1	LH (4)	(Markert *et al.* 2004)

Mohor gazelle (*Gazella dama mhorr*)	1	LH (1)	(Ruiz-Lopez *et al.* 2012)
New Zealand sea lion (*Phocarctos hookeri*)	1	LH (6) M (3)	(Osborne 2011)
Nine-spined stickleback (*Pungitius pungitius*)	16	LH (48)	(Laine *et al.* 2012)
Parsley frog (*Pelodytes punctatus*)	6	M (6)	(Jourdan-Pineau *et al.* 2012)
Red deer (*Cervus elaphus*)	3	LH (12) M (4)	(Coulson *et al.* 1999; Slate *et al.* 2000; Slate & Pemberton 2002)
Reindeer (*Rangifer tarandus*)	1	P (4)	(Cote *et al.* 2005)
Ring-tailed lemur (*Lemur catta*)	2	LH (2) M (2) P (35)	(Charpentier *et al.* 2008a; Charpentier *et al.* 2008c)
Roe deer (*Capreolus capreolus*)	5	M (10)	(Zachos *et al.* 2007)
Sea bream (*Sparus aurata*)	1	LH (2) M (1)	(Borrell *et al.* 2011)
Song sparrow (*Melospiza melodia*)	1	LH (1)	(Taylor *et al.* 2010)
Superb starling (*Lamprotornis superbus*)	1	LH (2)	(Rubenstein 2007)
Taita thrushes (*Turdus helleri*)	3	M (9)	(Lens *et al.* 2000)
Threespine stickleback (*Gasterosteus aculeatus*)	1	LH (5) M (2)	(Lieutenant-Gosselin & Bernatchez 2006)
Yellow baboons (*Papio cynocephalus*)	1	LH (6)	(Charpentier *et al.* 2008b)
Zebra finch (*Taeniopygia guttata*)	2	LH (56) M (32)	(Forstmeier *et al.* 2012)
Zenaida doves (*Zenaida aurita*)	1	LH (1) M (5)	(Monceau *et al.* 2013)

Table 4 - 2 Number of estimates (k) the average effect sizes (Z_r) and their confidence intervals for each trait category

Trait Category	k	Average Z_r	SD r	95% CI
Life History	66	0.064	0.153	0.045 - 0.084
Morphology	56	0.025	0.218	0.002 - 0.048
Physiological	7	0.041	0.110	0.002 - 0.079
All	129	0.046	0.183	0.032 - 0.060

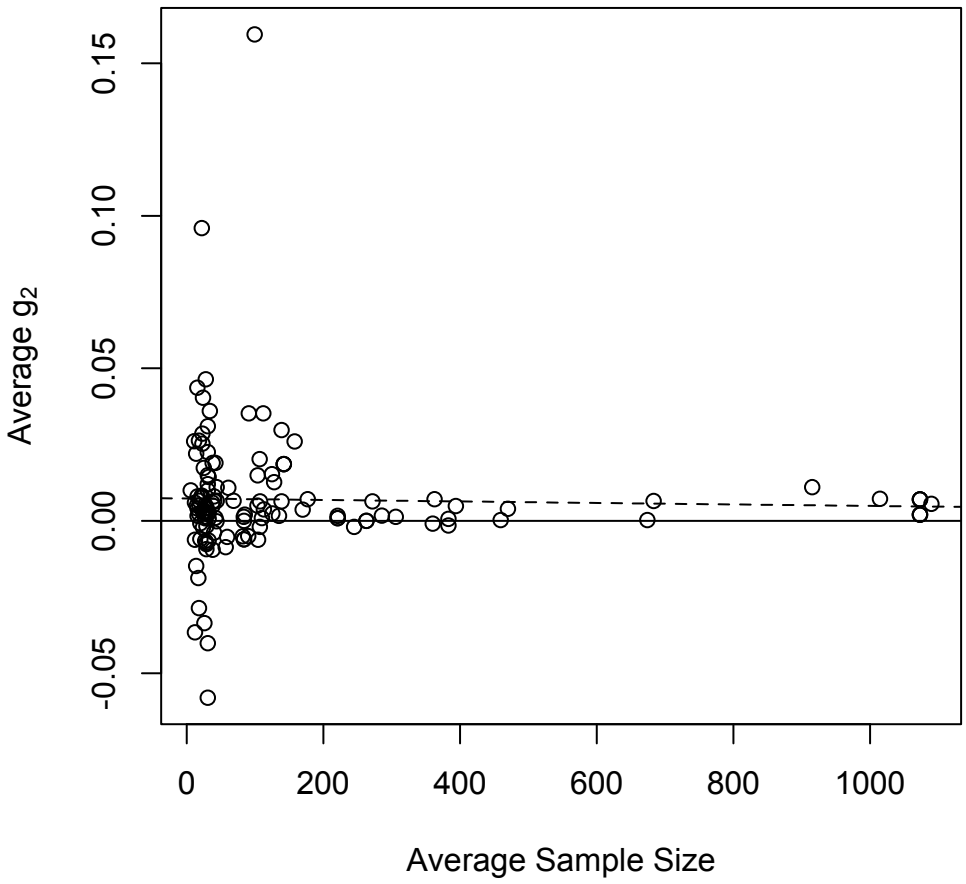
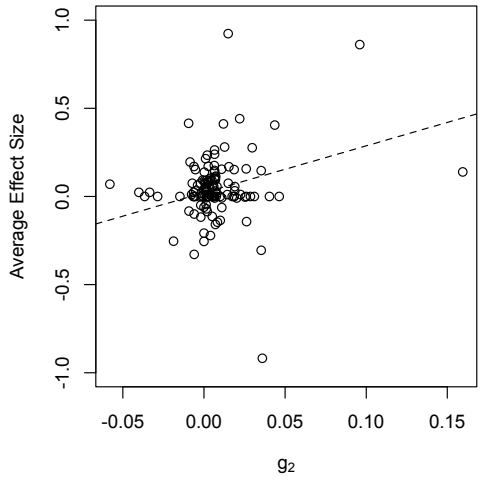
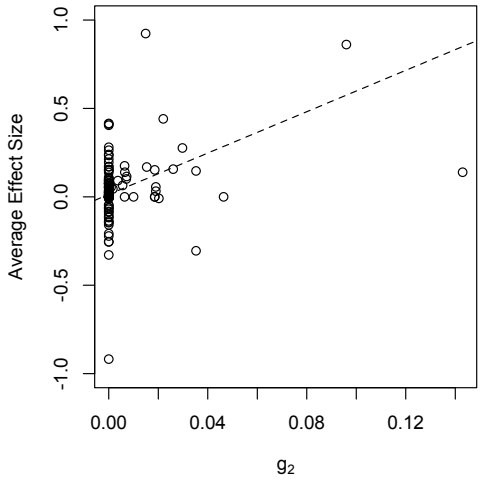


Figure 4 - 1 Funnel plot showing normalized weighted average effect sizes against average sample size for the 129 data points used in our meta-analysis
Dotted line shows an Egger's regression line.

A)



B)



C)

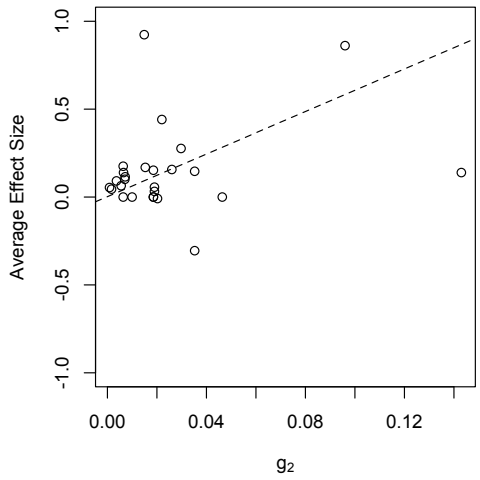


Figure 4 - 2 Scatter plots of study average effect sizes against all g_2 estimates (A), non-significant g_2 estimates reduced to zero (B), and only significant g_2 estimates (C). Dotted lines represent weighted regressions.

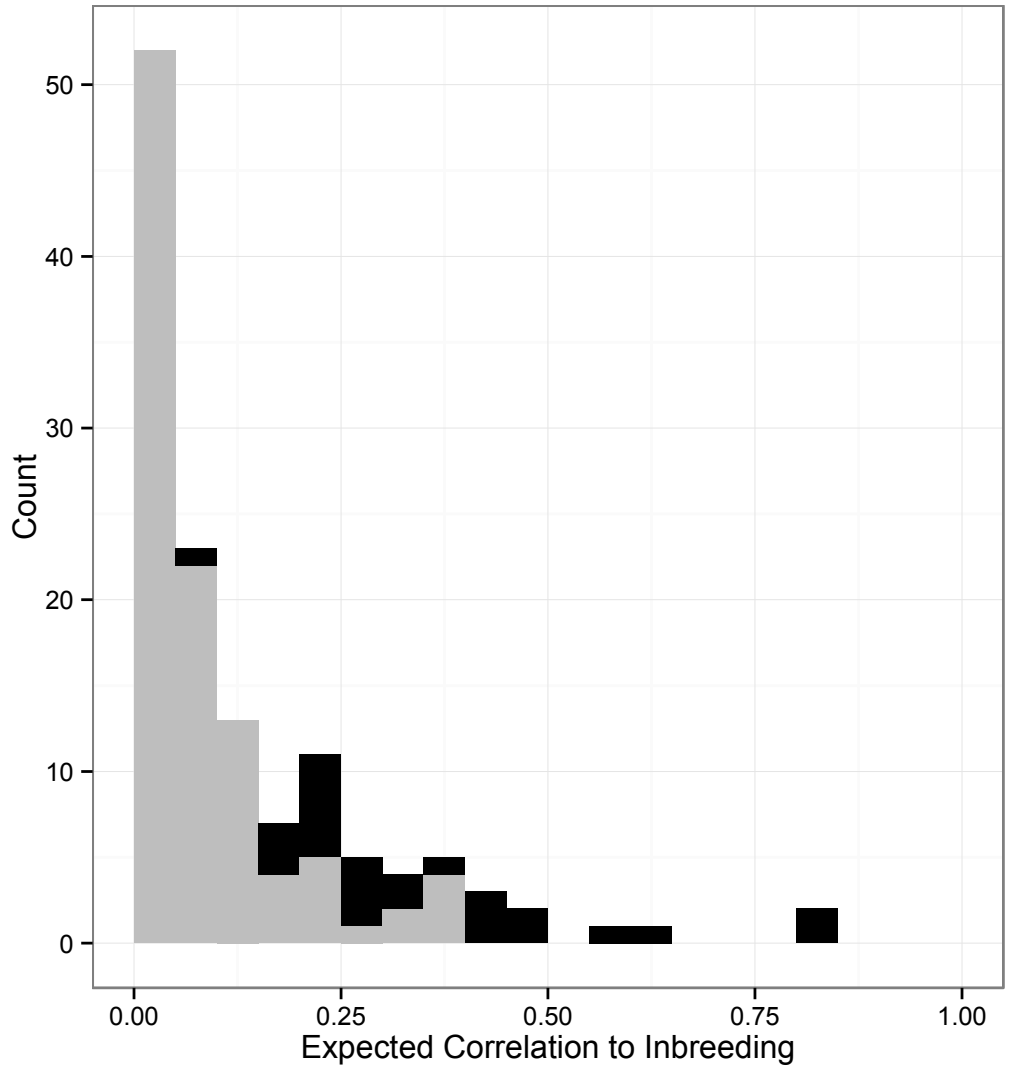


Figure 4 - 3 Histogram of expected correlations between marker heterozygosity and inbreeding
Grey shading represents populations where g_2 did not differ from zero, black shading represents populations with significant g_2.

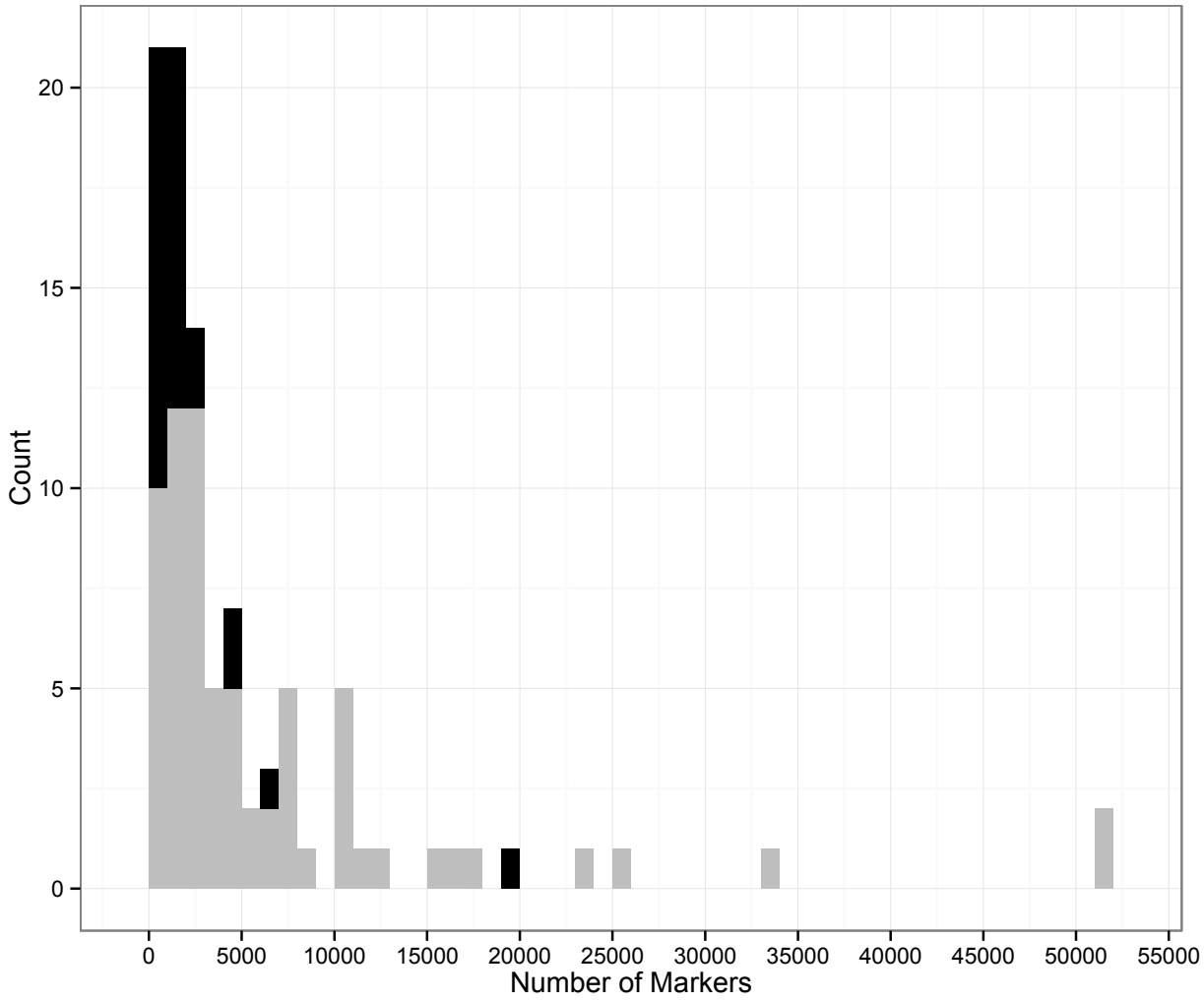


Figure 4 - 4 Histogram showing the number of markers that would have been required for the populations considered to have a 0.9 correlation between marker heterozygosity and inbreeding

Grey shading represents populations where g_2 did not differ from zero, black shading represents populations with significant g_2.

4.6 Bibliography

Amos W, Wilmer J, Fullard K*, et al.* (2001) The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society B-Biological Sciences* **268**, 2021-2027.

Angeloni F, Wagemaker N, Vergeer P, Ouborg J (2012) Genomic toolboxes for conservation biologists. *Evolutionary Applications* **5**, 130-143.

Aparicio JM, Ortego J, Cordero PJ (2008) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology* **15**, 4659-4665.

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* **13**, 3021-3031.

Bartoń K (2009) MuMIn: multi-model inference, Available at: http://r-forge.r-project.org/projects/mumin/.

Bensch S, Andrén H, Hansson B*, et al.* (2006) Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLoS ONE* **1**, e72.

Bierne N, Tsitrone A, David P (2000) An inbreeding model of associative overdominance during a population bottleneck. *Genetics* **155**, 1981-1990.

Borrell Y, Carleos C, Sanchez J*, et al.* (2011) Heterozygosity-fitness correlations in the gilthead sea bream Sparus aurata using microsatellite loci from unknown and gene-rich genomic locations. *Journal of Fish Biology* **79**, 1111-1129.

Britten H (1996) Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* **50**, 2158-2164.

Caspers BA, Krause ET, Hendrix R*, et al.* (2014) The more the better – polyandry and genetic similarity are positively linked to reproductive success in a natural population of terrestrial salamanders (*Salamandra salamandra*). *Molecular Ecology* **23**, 239–250.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology* **18**, 2746-2765.

Charpentier M, Boulet M, Drea C (2008a) Smelling right: the scent of male lemurs advertises genetic quality and relatedness. *Molecular Ecology* **17**, 3225-3233.

Charpentier M, Setchell J, Prugnolle F*, et al.* (2005) Genetic diversity and reproductive success in mandrills (Mandrillus sphinx). *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16723-16728.

Charpentier M, Tung J, Altmann J, Alberts S (2008b) Age at maturity in wild baboons: genetic, environmental and demographic influences. *Molecular Ecology* **17**, 2026-2040.

Charpentier M, Williams C, Drea C (2008c) Inbreeding depression in ring-tailed lemurs (Lemur catta): genetic diversity predicts parasitism, immunocompetence, and survivorship. *Conservation Genetics* **9**, 1605-1615.

Coltman D, Bowen W, Wright J (1998) Birth weight and neonatal survival of harbour seal pups ape positively correlated with genetic variation measured by microsatellites. *Proceedings of the Royal Society B-Biological Sciences* **265**, 803-809.

Coltman D, Pilkington J, Smith J, Pemberton J (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* **53**, 1259-1267.

Coltman D, Slate J (2003) Microsatellite measures of inbreeding: A meta-analysis. *Evolution* **57**, 971-983.

Cote S, Stien A, Irvine R*, et al.* (2005) Resistance to abomasal nematodes and individual genetic variability in reindeer. *Molecular Ecology* **14**, 4159-4168.

Coulson T, Albon S, Slate J, Pemberton J (1999) Microsatellite loci reveal sex-dependent responses to inbreeding and outbreeding in red deer calves. *Evolution* **53**, 1951-1960.

Coulson T, Pemberton J, Albon S*, et al.* (1998) Microsatellites reveal heterosis in red deer. *Proceedings of the Royal Society B-Biological Sciences* **265**, 489-495.

Curik I, Zechner P, Solkner J*, et al.* (2003) Inbreeding, microsatellite heterozygosity, and morphological traits in Lipizzan horses. *Journal of Heredity* **94**, 125-132.

Da Silva A, Luikart G, Yoccoz N, Cohas A, Allaine D (2006) Genetic diversity-fitness correlation revealed by microsatellite analyses in European alpine marmots (Marmota marmota). *Conservation Genetics* **7**, 371-382.

Davey JW, Hohenlohe PA, Etter PD*, et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* **12**, 499-510.

David P (1998) Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* **80**, 531-537.

David P, Pujol B, Viard F, Castella V, Goudet J (2007) Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology* **16**, 2474-2487.

DeWoody Y, DeWoody J (2005) On the estimation of genome-wide heterozygosity using molecular markers. *Journal of Heredity* **96**, 85-88.

DiBattista JD, Feldheim KA, Gruber SH, Hendry AP (2008) Are indirect genetic benefits associated with polyandry? Testing predictions in a natural population of lemon sharks. *Molecular Ecology* **17**, 783-795.

Egger M, Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* **315**, 629-634.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1-15.

Elshire R, Glaubitz J, Sun Q*, et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One* **6**.

Forstmeier W, Schielzeth H (2011) Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology* **65**, 47-55.

Forstmeier W, Schielzeth H, Mueller J, Ellegren H, Kempenaers B (2012) Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Molecular Ecology* **21**, 3237-3249.

Fossøy F, Johnsen A, Lifjeld J (2008) Multiple genetic benefits of female promiscuity in a socially monogamous passerine. *Evolution* **62**, 145-156.

Gage M, Surridge A, Tomkins J*, et al.* (2006) Reduced heterozygosity depresses sperm quality in wild rabbits, Oryctolagus cuniculus. *Current Biology* **16**, 612-617.

Grueber C, Nakagawa S, Laws R, Jamieson I (2011a) Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology* **24**, 699-711.

Grueber C, Waters J, Jamieson I (2011b) The imprecision of heterozygosity-fitness correlations hinders the detection of inbreeding and inbreeding depression in a threatened species. *Molecular Ecology* **20**, 67-79.

Grueber CE, Wallis GP, Jamieson IG (2008) Heterozygosity–fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology* **17**, 3978-3984.

Hammerly SC, Morrow ME, Johnson JA (2013) A comparison of pedigree- and DNA-based measures for identifying inbreeding depression in the critically endangered Attwater's Prairie-chicken. *Molecular Ecology* **22**, 5313–5328.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* **11**, 2467-2474.

Harrison X, Bearhop S, Inger R*, et al.* (2011) Heterozygosity-fitness correlations in a migratory bird: an analysis of inbreeding and single-locus effects. *Molecular Ecology* **20**, 4786-4795.

Heath D, Bryden C, Shrimpton J*, et al.* (2002) Relationships between heterozygosity, allelic distance (d(2)), and reproductive traits in chinook salmon, Oncorhynchus tshawytscha. *Canadian Journal of Fisheries and Aquatic Sciences* **59**, 77-84.

Herdegen M, Nadachowska-Brzyska K, Konowalik A, Babik W, Radwan J (2013) Heterozygosity, sexual ornament and body size in the crested newt. *Journal of Zoology* **291**, 146-153.

Hicks J, Rachlow J (2006) Is there a genetic basis for antler and pedicle malformations in reintroduced elk in Northern Arizona? *Southwestern Naturalist* **51**, 276-282.

Hoffman J, Forcada J, Amos W (2010a) Exploring the Mechanisms Underlying a Heterozygosity-Fitness Correlation for Canine Size in the Antarctic Fur Seal Arctocephalus gazella. *Journal of Heredity* **101**, 539-552.

Hoffman J, Hanson N, Forcada J, Trathan P, Amos W (2010b) Getting Long in the Tooth: A Strong Positive Correlation between Canine Size and Heterozygosity in Antarctic Fur Seals Arctocephalus gazella. *Journal of Heredity* **101**, 527-538.

Hoglund J, Piertney S, Alatalo R*, et al.* (2002) Inbreeding depression and male fitness in black grouse. *Proceedings of the Royal Society B-Biological Sciences* **269**, 711-715.

Jennions M, Moller A (2002) Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society B-Biological Sciences* **269**, 43-48.

Jourdan-Pineau H, Folly J, Crochet P-A, David P (2012) Testing the influence of family structure and outbreeding depression on heterozygosity-fitness correlations in small populations. *Evolution* **66**, 3624–3631.

Kardos M, Allendorf FW, Luikart G (2014) Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium? *Molecular Ecology Resources* **14**, 519–530.

Klauke N, Segelbacher G, Schaefer H (2013) Reproductive success depends on the quality of helpers in the endangered, cooperative El Oro parakeet (Pyrrhura orcesi). *Molecular Ecology* **22**, 2011-2027.

Küpper C, Kosztolányi A, Augustin J*, et al.* (2010) Heterozygosity-fitness correlations of conserved microsatellite markers in Kentish plovers *Charadrius alexandrinus*. *Molecular Ecology* **19**, 5172–5185.

Laine V, Herczeg G, Shikano T, Primmer C (2012) Heterozygosity-behaviour correlations in nine-spined stickleback (Pungitius pungitius) populations: contrasting effects at random and functional loci. *Molecular Ecology* **21**, 4872-4884.

LeBas N (2002) Mate choice, genetic incompatibility, and outbreeding in the ornate dragon lizard, Ctenophorus ornatus. *Evolution* **56**, 371-377.

Lens L, Van Dongen S, Galbusera P*, et al.* (2000) Developmental instability and inbreeding in natural bird populations exposed to different levels of habitat disturbance. *Journal of Evolutionary Biology* **13**, 889-896.

Lieutenant-Gosselin M, Bernatchez L (2006) Local heterozygosity-fitness correlations with global positive effects on fitness in threespine stickleback. *Evolution* **60**, 1658-1668.

Ljungqvist M, ÅKesson M, Hansson B (2010) Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al. (2008). *Molecular Ecology* **19**, 851-855.

Luquet E, Lena J, David P*, et al.* (2013) Within- and among-population impact of genetic erosion on adult fitness-related traits in the European tree frog Hyla arborea. *Heredity* **110**, 347-354.

Markert J, Grant P, Grant B*, et al.* (2004) Neutral locus heterozygosity, inbreeding, and survival in Darwin's ground finches (Geospiza fortis and G-scandens). *Heredity* **92**, 306-315.

Marshall TC, Spalton JA (2000) Simultaneous inbreeding and outbreeding depression in reintroduced Arabian oryx. *Animal Conservation* **3**, 241-248.

Miller JM, Malenfant RM, David P*, et al.* (2014) Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity* **112**, 240–247.

Monceau K, Wattier R, Dechaume-Moncharmont F, Dubreuil C, Cezilly F (2013) Heterozygosity-Fitness Correlations in Adult and Juvenile Zenaida Dove, Zenaida aurita. *Journal of Heredity* **104**, 47-56.

Nakaoka H, Inoue I (2009) Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *Journal of Human Genetics* **54**, 615-623.

Neff B (2004) Stabilizing selection on genomic divergence in a wild fish population. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 2381-2385.

Olano-Marin J, Mueller JC, Kempenaers B (2011a) Correlations between heterozygosity and reproductive success in the blue tit (*Cyanistes caeruleus*): an analysis of inbreeding and single locus effects. *Evolution* **65**, 3175–3194.

Olano-Marin J, Mueller JC, Kempenaers B (2011b) Heterozygosity and survival in blue tits (*Cyanistes caeruleus*): contrasting effects of presumably functional and neutral loci. *Molecular Ecology* **20**, 4028–4041.

Osborne A (2011) *Assessment of genetic variation in the threatened New Zealand sea lion, Phocarctos hookeri, and its association with fitness*, University of Otago.

Otter K, Stewart I, McGregor P*, et al.* (2001) Extra-pair paternity among Great Tits Parus major following manipulation of male signals. *Journal of Avian Biology* **32**, 338-344.

R Development Core Team (2005) R: A language and environment for statistical computing, reference index version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria.

Rubenstein D (2007) Female extrapair mate choice in a cooperative breeder: trading sex for help and increasing offspring heterozygosity. *Proceedings of the Royal Society B-Biological Sciences* **274**, 1895-1903.

Ruiz-Lopez M, Ganan N, Godoy J*, et al.* (2012) Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *Conservation Biology* **26**, 1121-1129.

Santure AW, Stapley J, Ball AD*, et al.* (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology* **19**, 1439-1451.

Shen R, Fan J-B, Campbell D*, et al.* (2005) High-throughput SNP genotyping on universal bead arrays. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **573**, 70-82.

Slate J, David P, Dodds KG*, et al.* (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **93**, 255-265.

Slate J, Gratten J, Beraldi D*, et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97-107.

Slate J, Kruuk LEB, Marshall TC, Pemberton JM, Clutton-Brock TH (2000) Inbreeding depression influences lifetime breeding success in a wild population of red deer (Cervus elaphus). *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**, 1657-1662.

Slate J, Pemberton J (2002) Comparing molecular measures for detecting inbreeding depression. *Journal of Evolutionary Biology* **15**, 20-31.

Sterne J, Becker B, Egger M*, et al.* (2005a) The Funnel Plot. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 75-98.

Sterne J, Egger M, Rothstein H, Sutton A, Borenstein M (2005b) Regression Methods to Detect Publication and Other Bias in Meta-Analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 99-110.

Stewart I, Westneat D (2013) Patterns of hatching failure in the house sparrow Passer domesticus. *Journal of Avian Biology* **44**, 69-79.

Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: A time for reapprasial. *Evolution* **64**, 1202-1217.

Taylor S, Sardell R, Reid J*, et al.* (2010) Inbreeding coefficient and heterozygosity-fitness correlations in unhatched and hatched song sparrow nestmates. *Molecular Ecology* **19**, 4454-4461.

Townsend SM, Jamieson IG (2013) Molecular and pedigree measures of relatedness provide similar estimates of inbreeding depression in a bottlenecked population. *Journal of Evolutionary Biology* **26**, 889–899.

Voegeli B, Saladin V, Wegmann M, Richner H (2013) Heterozygosity is linked to the costs of immunity in nestling great tits (*Parus major*). *Ecology and Evolution* **3**, 4815–4827.

Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* **17**, 3808-3817.

Wetzel D, Stewart I, Westneat D (2012) Heterozygosity predicts clutch and egg size but not plasticity in a house sparrow population with no evidence of inbreeding. *Molecular Ecology* **21**, 406-420.

White T, Searle J (2008) Mandible asymmetry and genetic diversity in island populations of the common shrew, Sorex araneus. *Journal of Evolutionary Biology* **21**, 636-641.

Zachos F, Hartl G, Suchentrunk F (2007) Fluctuating asymmetry and genetic variability in the roe deer (Capreolus capreolus): a test of the developmental stability hypothesis in mammals using neutral molecular markers. *Heredity* **98**, 392-400.

Zollner S, Pritchard J (2007) Overcoming the winner's curse: Estimating penetrance parameters from case-control data. *American Journal of Human Genetics* **80**, 605-615.

Chapter 5

EXPLORING THE GENOMIC BASIS FOR FITNESS RELATED TRAITS IN BIGHORN SHEEP

5.1 Introduction

One of the goals of molecular ecology is to identify the genomic regions influencing traits that have ecological relevance (Ellegren & Sheldon 2008; Slate *et al.* 2009). There is a particular interest in finding those genes associated with differences in fitness as such traits are expected to be subject to strong selection. Under strong directional selection, the genetic variability underlying fitness related traits should rapidly go to fixation, and yet a large amount of phenotypic variation in such traits is observed in the wild (Chenoweth & McGuigan 2010; Kruuk *et al.* 2008). Elucidating the genetic basis of fitness related traits might help to clarify the mechanism(s) by which the phenotypic variation is maintained.

A classic example of a secondary sexual characteristic often associated with fitness is horn size in ruminants. Horns are one of four types of ruminant headgear, characterized by a keratin sheath around a bony projection from the skull that grow continuously throughout an animal's life (Davis *et al.* 2011). Across a number of species horn size in males determines social status and mating access to females (Bro-Jørgensen 2007).

The genetic basis of horn development in the genus *Ovis* has begun to be investigated in a number of studies on domestic sheep (*Ovis aries*). Within this species some breeds have horns while others are polled (lacking horns entirely), and from an agronomic production standpoint there is interest in removing horns from certain breeds (Kijas *et al.* 2012). In addition, Soay sheep, a primitive breed now living feral on the islands of St. Kilda Scotland, have additional "morphs" of horns within and between the sexes. In females, there are three morphs: normal horned, scurred (deformed horns composed only of keratin sheaths not attached to the skull), and polled. While in males there are only two morphs: normal horns and scurs (Johnston *et al.* 2011; Johnston *et al.* 2010).

Researchers have found a single genomic region on chromosome 10 that is associated with the presence and absence of horns in domestic breeds (Kijas *et al.* 2012) and Soay sheep

(Johnston *et al.* 2011; Johnston *et al.* 2010), as well as being linked to quantitative differences in horn length of normal horned male Soay sheep (Johnston *et al.* 2011). This region contains a single gene relaxin/insulin-like family peptide receptor 2 (RXFP2). RXFP2 has previously been shown to affect osteoporosis and testicular decent in mice and humans (Feng *et al.* 2009; Ferlin *et al.* 2008; Yuan 2010). Thus, its association with both bone development and secondary sexual characteristics make it an interesting candidate for influencing horn morphology. Furthermore, different genotypes at this locus in male Soay sheep have been shown to be associated with trade-offs between reproductive success and longevity, which is thought to maintain the presence of the different horn morphs though heterozygote overdominance (Johnston *et al.* 2013). Finally, though not the major QTL underlying horn phenotype in cattle (*Bos taurus*), RXFP2 has been implicated in horn development in several association studies (Allais-Bonnet *et al.* 2013; Gautier & Naves 2011; Wiedemar *et al.* 2014), indicating that it may have similar function across species.

Another species of the genus *Ovis* known for large horns are bighorn sheep (*Ovis canadensis*). In this species all individuals have normal horns, though there is substantial sexual dimorphism with males having larger horns than females. Previous research on bighorn sheep has shown that horn size and body mass are important to intrasexual competition among males for reproductive access to females (Coltman *et al.* 2002; Martin *et al.* 2013). However, for female bighorn sheep horn length was found to be unrelated to social rank or other life history characteristics, which were more determined by body mass and age (Favre *et al.* 2008). In addition, horn size determines the trophy status of an individual and, in some jurisdictions, the age at which it can be legally harvested (as individuals with fast growing horns will be removed at a younger age) which directly influences longevity and survival (Bonenfant *et al.* 2009; Festa-Bianchet *et al.* 2008; Festa-Bianchet *et al.* 2014; Hengeveld & Festa-Bianchet 2011). Previous studies of bighorn sheep have shown that both horn size and body mass are heritable (Coltman 2005; Coltman *et al.* 2005; Poissant *et al.* 2012; Poissant *et al.* 2008) and quantitative trait locus

(QTL) mapping with microsatellite loci highlighted several suggestive regions for different aspects of horn morphology (e.g. horn volume and base circumference) as well as body mass (Poissant *et al.* 2012). These regions appear on several chromosomes, but notably included suggestive QTL for horn volume and base circumference in male bighorn sheep on chromosome 10 that spans the region predicted to contain RXFP2.

In this study, we build on these results using genomic methodologies to look for association between single nucleotide polymorphism (SNP) loci and fitness related characteristics in bighorn sheep including two aspects of horn morphology, body mass, and a variety of life history traits. Genomic methods increase a researcher's ability to detect associations between phenotypes and genotypes by rapidly generating large numbers of genotypes in many individuals, even in non-model species (Ellegren & Sheldon 2008; Mackay *et al.* 2009; Slate *et al.* 2009; Stinchcombe & Hoekstra 2008). This includes various genotype-by-sequencing methodologies (Elshire *et al.* 2011; Hohenlohe *et al.* 2010), as well as harnessing previously developed genomic resources and using them across 'genome-enabled' taxa (Kohn *et al.* 2006; Miller *et al.* 2012).

This work is based on phenotypic data from a long-term study of individually marked sheep followed throughout their lives at Ram Mountain, Alberta Canada. We also capitalize on the close evolutionary relationship between domestic and bighorn sheep (Bunch *et al.* 2006) to efficiently genotype many SNPs in a large number of individual bighorn sheep using a genomic technology developed for domestic sheep. We then follow-up suggestive results by typing additional individuals from the same population at candidate loci, and assess if allele frequencies at these loci have changed over time.

By investigating this suite of characters, we add to the literature on the genetic basis of complex quantitative traits (such as fitness), and provide a comparison to domestic sheep to see if the genetic architecture underlying horn morphology is similar in a wild relative. In addition, knowledge of the genetic variants associated with horn and body size in bighorn sheep can

inform management strategies aimed at preserving the genetic ability to grow horns (Coltman 2008).

5.2 Materials and Methods

5.2.1 Population History and Phenotypic Data Collection

The bighorn sheep at Ram Mountain (RM) are a native population that is highly isolated and philopatric. Individual-based monitoring of the population began in 1972 (Jorgenson *et al.* 1997; Jorgenson *et al.* 1993), where individuals are marked with unique tags as lambs, so most sheep are of known age. Individuals are followed their lives and every spring and summer sheep are drawn into a corral trap baited with salt where several phenotypic measurements taken, including body mass and horn morphology measures (Jorgenson 1993). Genetic sampling of the population began in 1988, from which a genetic pedigree has been maintained (Coltman *et al.* 2002; Poissant *et al.* 2008). In some cases full or half siblings were inferred from unsampled males using the program COLONY (Wang 2013). By 2013, the pedigree contained 864 maternal and 528 paternal links involving 1134 sheep.

5.2.2 Phenotypic Measures

We considered three morphological characteristics: average horn length, average horn base circumference, and body mass. Sheep were weighed to the nearest 250 g with a Detecto spring scale, while horn length (measured along the outside curvature) and base circumference were measured to the nearest millimeter with tape. Each trait was standardized to a sex and age specific standard deviation of one (value divided by the SD for that sex in that age class). We only considered individuals aged 1 or greater to avoid maternal effects (Poissant *et al.* 2012; Wilson *et al.* 2005), and pooled males aged ≥9 and females aged ≥14 to increase sample sizes in those age classes. We do not believe that pooling will have a major consequences for our

analyses as even though horns continue to grow throughout an animal's entire life most horn growth is completed by age four in male bighorn sheep (Jorgenson *et al.* 1998), age six in females (Favre *et al.* 2008), and for both sexes mass gain asymptotes by age seven (Festa-Bianchet *et al.* 1996).

We also considered five life history measures: survival to age 1 (binary trait), longevity, number of offspring, fecundity (number of offspring divided by longevity), and age of primiparity. Survival and longevity were recorded from field observations and should be robust to tag loss given that the population is highly isolated and resighting probabilities are over 95% for both males and females (Jorgenson *et al.* 1997). Number of offspring and was derived directly from the pedigree, while age at primipatry was calculated as the difference between the first year an individual had an offspring recorded in the pedigree and that individual's year of birth. For each trait we included only individuals born before 2010 so when considering survival and longevity all individuals had the opportunity to reach 3 years old.

5.2.3 Quantitative Genetic Analyses for Morphological Characteristics

Quantitative genetic variation in our morphological characteristics was estimated using a series of 'animal models'. Animal models are linear mixed effects models that incorporate pedigree information along with phenotypic measures to partition phenotypic variation (V_p) into that due to additive genetic variation (V_a), permanent environmental effects (V_{pe}), and residual variation (V_r) (Kruuk 2004; Wilson *et al.* 2010). For our analyses fixed effects included sex, age (fit as a factor), date on which the measurement was taken (fit as a continuous, second-order polynomial), as well as all interactions between the three variables. Random effects were individual identity to account for permanent environmental effects associated with having repeated measures of individuals (V_{pe}), as well as year of birth (V_{yb}) and year of measurement (V_y) to account for environmental effects. Thus phenotypic variation was broken into five components $V_p = V_a + V_{pe} + V_y + V_{yb} + V_r$.

The three morphological traits were modeled independently using univariate animal models run in ASReml version 3.0 (Gilmour *et al.* 2009). Models were based on datasets of measurements taken between 1972 and 2012. To maximize statistical power we considered both sexes simultaneously. The effect size of each random effect was calculated as the proportion of V_P explained by the random effect, and its significance tested by comparing a model with the term removed to the full model using a likelihood ratio test with 1 degree of freedom. From these models, we calculated heritability (h^2) of each trait as the ratio of V_a/V_p. We also estimated individual breeding values using best linear unbiased predictors (BLUPs) for use in association analyses (see below).

5.2.4 SNP Genotyping

Individuals were chosen for genotyping based on their BLUP value for horn length (n = 95). Specifically we attempted to maximize our chances of detecting an association by choosing an approximately equal number of those individuals with the highest and lowest BLUP values with respect to horn length. The selected individuals were typed on the Ovine Infinium® HD SNP BeadChip, a newly developed SNP array for domestic sheep that contains 606,006 loci (Kijas *et al.* 2014). Initial assessment of genotype quality was performed using Genome Studio version 2011.1 (Genotyping Module 1.9; Illumina). We used cluster information based on 288 domestic sheep samples (provided by the International Sheep Genomics Consortium) and discarded all loci with GenTrain scores less than 0.8 and GenCall scores less than 0.6. Genotypes were then exported to PLINK version 1.07 (Purcell *et al.* 2007) for additional filtering. Specifically, we considered only those loci which mapped to the autosomes in domestic sheep, had a minor allele frequency >5%, and were in Hardy-Weinberg Equilibrium (adjusted $p > 1.28 \times 10^{-5}$) in our sample set.

5.2.5 GWAS Analyses

Association tests were done using GCTA version 1.24 (Yang *et al.* 2011a). When analyzing association statistics, GCTA accounts for (cryptic) population structure by calculating a matrix of pairwise relatedness values among samples that is then used as a covariate in the association analysis. Furthermore, when testing for an association all the markers on the chromosome on which the candidate SNP is located are removed from calculation of the relatedness matrix. Together these procedures are thought to reduce the number of false positive associations (Yang *et al.* 2014).

For the three morphological characters, the phenotypic measure used in the association tests were BLUP estimates from ASReml. While the use of breeding values rather than direct phenotypes in analyses can lead to biased inferences (Hadfield *et al.* 2010) in this case it was necessary, as GCTA cannot use repeated measures data. For life history variables we used the phenotypic data, and the model included covariates of sex and if the individual was still alive as of 2013.

To correct for multiple testing we used Keff (Moskvina & Schmidt 2008) to determine significance thresholds genome-wide, and for each chromosome individually assuming an alpha value of 0.05. Association results were then visualized with the ggplot2 package version 1.0.0 (Wickham 2009) in R version 2.13 (R Core Team 2012).

5.2.6 Candidate Loci Validation

Loci that were found in association reaching chromosome-wide significance level (n = 11; see results) were genotyped in a second panel of individuals. This panel consisted of 136 additional individuals (prioritized to contain those with the most phenotypic measurements in our long term dataset) along with seven sheep that had been previously genotyped on the SNP chip to test concordance of genotypes across the two methods. Primers were designed for individual

loci based on the genomic sequence used to build the 700k SNP chip. Interrogation reactions were done using Type-it (Qiagen) followed by SNaPshot (Life Technologies) adhering the manufacturer's protocol except that Type-it reactions were conducted in 10uL volumes. Genotypes were resolved on a 3730 DNA Analyser (Applied Biosystems International) using GeneScan LIZ 120 size standard (Applied Biosystems International) and were called with GeneMapper version 4.0 (Applied Biosystems). Detailed description of the methods, including primer sequences and PCR conditions, is provided in Appendix 5-1.

Following Johnston et al. (2011), we tested for significant effects of marker genotypes on morphological variables by fitting additional animal models in ASReml that included genotype at a locus as a random effect. Loci were fit individually and compared to a null model containing all covariates from the initial ASReml analysis. In these analyses, we ran all typed individuals simultaneously rather than considering only the newly typed individuals. We chose this procedure because while having a completely separate discovery and test sample sets would be more conservative as to whether or not the observed associations are true, removing the chip typed individuals, which represent the high and low ends of the range of breeding values, would have led to a severe reduction in power. Note that datasets vary slightly between tests as individuals with missing genotypes at a locus were also removed from the null model to preserve equal degrees of freedom. Model comparisons were then done using likelihood ratio tests. This procedure was repeated for all locus-trait combinations. For tests showing significant genotype effects, we further examined the proportion of variation explained by the locus genotypes.

As ASReml cannot consider traits with non-normal distributions we assessed the effect of locus genotypes on life-history variables using generalized linear mixed effect models (GLMM) with a Poisson distribution, via the lme4 package version 1.1-7 (Bates *et al.* 2014) in R version 3.1.2 (R Core Team 2014). For all traits except fecundity we first constructed null models with fixed effects of the individual's sex as well as a binary factor indicating if the individual was still alive as of 2013, and a random effect of year of birth. For fecundity, the response variable was

the number of offspring, but we added longevity as an additional fixed effect. These null models were then compared to models with locus genotype added as a fixed effect. Comparisons were done using AICc values via the MuMIn package version 1.13.4 (Bartoń 2009). Models with a change of >2 AICc were considered significantly different (Burnham & Anderson 2002). As with morphological characters, we examined all pairwise combinations of life history traits and loci.

5.2.7 Temporal Analyses

For each candidate locus, allele frequency was calculated for cohorts between 1981-2013 (excluding 2012 as no individuals born in that year were genotyped). Average sample size per cohort was 6.5 individuals (SD = 6.8, range 1-27 individuals). Note that these values are correlated to the total cohort size in the population (r = 0.62), and grouping into two year or three year bins produced the same patterns (results not shown). To test for changes over time we regressed allele frequency against year using a linear model with weights according to the number of samples genotyped in each cohort. We implemented a weighted regression so years with small sample sizes would not bias temporal trend estimates.

To test if the observed changes in allele frequency were different from what is expected under mutation-drift equilibrium, we used gene-dropping analyses (Gratten *et al.* 2012; Johnston *et al.* 2013). In this method, alleles are passed through the RM pedigree multiple times (n = 1000) to get a distribution of changes in allele frequency assuming no selection. At the beginning of each iteration alleles were assigned randomly in the starting cohort based on their observed frequency in the individuals present in the population in 1988 (the first year of genetic sampling, n = 93 individuals). For individuals with one or two unknown parents in subsequent generations alleles were randomly assigned based on their frequencies in the previous cohort. At the end of each simulation, the rate of allele frequency change was assessed using a linear model weighted by the sample size of each cohort. For consistency with the empirical analysis, we calculated

allele frequencies and sample sizes based only on those individuals genotyped at the locus. We then compared the observed change in allele frequency to the locus specific null distribution using a one-tailed test.

5.3 Results

5.3.1 GWAS for Fitness Related Traits

All three morphological traits examined exhibited significant additive genetic variation, on par with what was seen in other studies of this population (Table 5-1). Datasets used in the GWAS analyses differed slightly depending on what variable was examined: for morphological variables the dataset consisted of 76 individuals with measures of all three characteristics; for survival, longevity and number of offspring the number of individuals included was 94, while age at primiparity had 47 individuals, and fecundity had 79.

In total, 95 individuals were genotyped on the SNP chip and used to quality filter loci. One individual was subsequently removed from further analyses after significant (>5%) pedigree inconsistencies were found using VIPER (Paterson *et al.* 2012). The final dataset contained 3777 loci, and there were at least 60 markers on each autosome (average ± SD = 145.3 ± 88.6; Appendix 5-2).

No loci were found to be associated at the genome-wide significance level for any of the morphological traits examined. Five loci associated significantly at the level of individual chromosomes (Figure 5-1). Base circumference was associated with a single locus on chromosome 14, horn length with two loci, one on chromosome 3 and the other on 24, and body mass was also associated with two loci, one on chromosome 2 and the other on 10.

As with morphological traits, no loci were found to be significantly associated with the life-history traits examined at on the genome-wide level. However, all traits except survival were found to have at least one locus associated at the per-chromosome significance level (Figure 5-

2). Longevity had two on chromosome 12 and another on chromosome 15; age at primiparity had one on chromosome 19; number of offspring had two on chromosome 12 (the same that were involved with longevity); and fecundity had one on chromosome 10, and another on chromosome 24.

5.3.2 Candidate Loci Validation

Genotypes were obtained for 10 of 11 loci in 140 samples (133 of 136 additional samples and 7 used for concordance testing), with the second locus for length on chromosome 24 failing completely. Excluding the three samples that failed to be typed at all loci, genotyping success was 99.9%, and concordance was 100% for the seven individuals genotyped by both platforms.

For morphological variables, combining the new samples with those originally typed on the SNP chip gave a sample size of up to 209 individuals depending on the trait. We found that in all but one case adding genotype to our base model improved fit when the locus was added to the phenotype model it was originally associated with (Table 5-2). In no cases did model fit improve when considering genotypes for loci not originally associated with that phenotype. The one case where fit was not improved was for the locus on chromosome 10 originally associated with average body mass ($p = 0.20$). The amount of variation explained by the genotypes at a locus ranged from 4% - 14% (Table 5-2), but in all cases, the standard errors were larger than the estimates themselves.

For life history measures, the full data set consisted of up to 223 individuals depending on the trait. We found that 6 of 24 models (25%) incorporating locus genotype resulted in a better fit compared to the null model (Appendix 5-3). None of the loci originally showing association with age of primiparity or fecundity improved fit for models considering those traits. The two loci originally associated with longevity remained significant, and two additional loci also improved fit for this trait. Including genotypes of two loci improved model fit for models considering number of offspring, but neither of these loci were originally associated with the

trait. Given the inconsistency of these results we restricted subsequent analyses to the morphological characteristics.

5.3.3 Temporal Analyses

One of the four loci associated with morphological characteristics (OAR14_45166076; horn base circumference) showed significant changes in allele frequency over time based on linear modeling (Figure 5-3). Gene-dropping simulations of this locus were nominally significant ($P < 0.081$), suggesting that the observed changes may be different than expected from mutation-drift equilibrium. For this locus, we fit an additional animal model with genotype as a fixed effect to examine how each genotype corresponds to changes in morphology. We found that relative to heterozygotes, GG homozygotes have larger horn base circumference (0.376 ± 0.110 standard deviations), and AA homozygotes have smaller horn bases (-0.489 ± 0.263 standard deviations).

5.4 Discussion

In this study we examined the genetic bases of several fitness related characteristics in bighorn sheep. To do so we utilized a new genomic technology originally designed for domestic sheep to rapidly genotype markers in a wild species, and then combined this data with phenotypic measures from a long-term individual based study. Altogether we found 11 loci with suggestive associations to eight fitness related characteristics (Figures 5-1 & 5-2). We then attempted to confirm the initial associations by genotyping 10 of these loci in an expanded set of individuals. We found that for the majority of morphological characteristics the associations were maintained, but the associations did not hold up for life history measures.

5.4.1 Associations with Morphological Traits

Our initial GWAS analysis of three morphological traits found suggestive associations at five loci. Previous QTL mapping with microsatellite loci for these same traits in the RM population highlighted several candidate regions (Poissant *et al.* 2012). Interestingly, the body mass locus observed on chromosome 2 is relatively close, 7.4 megabases (Mbp) upstream, to a QTL for that same trait (assuming 1 centimorgan = 1 Mbp; Dumont & Payseur 2008). None of the remaining loci were near any of the other QTL described in Poissant *et al.* (2012). In addition, we found no overlap in location between the loci found here and morphological traits in the sheep QTL database (Hu et al. 2013; Hu et al. 2007).

Given the general lack of association with known QTL we examined gene annotations in the domestic sheep genome near the associated loci. In order to determine the genomic window within which to search we estimated 'half-length' of linkage disequilibrium (LD) for our marker set, i.e. the inter-marker distance at which LD decreased to half its maximal value (Reich *et al.* 2001). This value is thought to reflect the extent to which an association between genotypes at a given locus and a QTL can be detected. For this analysis we used PLINK version 1.90b2l (Chang *et al.* 2015) to calculate pairwise values of r^2 between syntenic markers on all chromosomes (n = 370,568 pairwise comparisons). These estimates were then compared to inter-marker physical distance based on map positions from the domestic sheep genome, and half-length was calculated using a custom script which calculated LD decay rate as in Appendix 2 of Hill and Weir (1988).

As expected, we found that there was a general decrease in the magnitude of LD with increasing inter-marker distance, and half-length was estimated to be 412,834 bp (Figure 5-4). It interesting to note that between a previous assessment of LD in bighorn sheep from RM (Miller et al. 2011) and this one the number of markers increased by an order of magnitude (308 vs. 3777 loci) and similarly the extent of LD dropped by an order of magnitude (~4,000,000 vs.

~400,000 bp). Analogous decreases in LD with the addition of markers have been seen in other species including cows (McKay et al. 2007; Porto-Neto et al. 2014), sheep (García-Gámez et al. 2012; Kijas et al. 2014), and flycatchers (Backström et al. 2006; Kawakami et al. 2014).

Based on the new half-length estimate we extracted gene names from the *Ovis aries* gene set (Oar v3.1, genebuild last updated Dec 2013) within a 413,000 bp window on either side of the candidate markers using bedtools version 2.23.0 (Quinlan & Hall 2010). This analysis returned 48 gene names (Appendix 5-4). These genes were not evenly distributed among loci with the locus on chromosome 14 (associated with horn base circumference) having the majority of annotations at 30, while the remaining loci were adjacent to between 4 and 7 genes. Where possible gene ontology (GO) terms were added to the genes in these lists using BioMart (Kinsella *et al.* 2011) and Ensembl version 77 (Flicek *et al.* 2014). Examination of GO terms did not reveal obvious candidates for association with our morphological characteristics (e.g. growth, muscle properties, or bone development).

It is somewhat surprising that we did not see even a suggestive association between horn morphology and the region surrounding RXFP2 on chromosome 10 given the very strong links seen in previous studies of both domestic sheep and cattle (Gautier & Naves 2011; Johnston et al. 2011; Johnston et al. 2013; Kijas et al. 2012; Wiedemar et al. 2014) as well as the suggestive QTL for horn volume in bighorn sheep in this same region (Poissant et al. 2012). However, based on the estimate of half-length is appears as if we did not have sufficient marker coverage to adequately test for associations in the horns region. Within our set of loci the closest marker to RXFP2 was 698,861bp away.

We wanted to more formally quantify the expected power of a marker to detect a hypothetical causal QTL given the average MAF and genome wide critical p-value for the loci in this study. To do so we used an R script developed by (Minikel 2012) which implements the QTL association feature of the Genetic Power Calculator (Purcell *et al.* 2003; Sham *et al.* 2000). Specifically we explored a variety of sample sizes, effect sizes, and LD estimates. This

exploration showed that even at extreme effect sizes for the QTL and levels of LD (well above what was seen at the half-length estimate) the number of samples used in the original GWAS analyses was likely not enough to have the power to detect associations (Figure 5-5). Note that these simulations assume that unrelated individuals were used in the GWAS, so the presence of related individuals in our test set will boost power slightly. In general though, the simulations indicate that our marker coverage likely increased the chance of Type II errors. However, we do not believe this diminishes the associations that were observed.

5.4.2 Temporal Patterns for Morphological Traits

One of four loci genotyped in our expanded panel of individuals was found to have a nominally significant change in allele frequency over time (Figure 5-3). There has been an active debate about whether human activities at RM, in particular trophy hunting, have led to phenotypic changes. A recent study (Traill *et al.* 2014) suggested that changes in morphology in this population are due mainly to environmental stochasticity (though see Hedrick *et al.* 2014). However, many previous studies have argued that the various hunting regimes RM has experienced have led to both phenotypic changes in horn and body size (Bonenfant *et al.* 2009; Festa-Bianchet *et al.* 2014; Pelletier *et al.* 2012) as well as associated changes in the genetic variation underlying these traits (Coltman *et al.* 2005; Coltman *et al.* 2003). In this work, we show for the first time a change in allele frequency at a locus that was associated with horn morphology. Moreover, at this locus the allele found to have increased over time was associated smaller horn base circumference. While horn base circumference is likely not the direct subject of hunting pressure, it is correlated to horn length and volume (Poissant *et al.* 2012; Poissant *et al.* 2008).

5.4.3 Lack of Consistent Associations for Life History Traits

In our initial GWAS analysis, we found six loci in association with four different life history characteristics. However, unlike the morphological traits, genotyping these loci in an expanded set of individuals found no consistent association. One possible explanation for this is that our selection criterion for the expanded set of individuals prevented an association from being observed. Specifically, the fact that we attempted to choose individuals with the most phenotypic measures for morphological characteristics might have removed individuals that died young from being included in the analysis. However, we do not think this lead to bias as each trait showed considerable variation, and the final analyses also included the SNP Chip typed individuals that were not selected based on number of phenotypic measures.

Alternatively, the lack of association could be due to a lack of additive genetic variation underlying these traits given that there should be strong directional selection for such traits. However, this seems unlikely, as previous studies of life history characteristics both in RM (Coltman 2005; Réale & Festa-Bianchet 2000) and other species (see table 1 in McFarlane *et al.* 2014; McFarlane *et al.* 2015) have found such traits to be heritable, albeit with a substantial amount of residual variation. Finally, a more plausible explanation is that such complex phenotypes are not single locus traits. Rather there may be several loci of small effect that jointly contribute to the phenotype, similar to the "missing heritability" phenomenon seen in many quantitative traits (Manolio *et al.* 2009; Yang *et al.* 2010). New methods, such as chromosome partitioning, can start to investigate this possibility (Robinson *et al.* 2013; Santure *et al.* 2013; Yang *et al.* 2011b). Unfortunately, we are unable to utilize chromosome partitioning at this time due to the small number of individuals typed on the 700k SNP chip. Attempts to use this method with our data produced unstable estimates of per-chromosome heritability (results not shown).

5.4.4 Potential Biases

One concern could be ascertainment bias caused by using a SNP chip across species. Specifically, that all the loci remaining polymorphic in bighorn sheep would have to be shared variants in the common ancestor of bighorn and domestic sheep, as a parallel mutation at a single site is highly unlikely. As such, variants that do segregate with our phenotypes may simply not be present on the chip, creating Type II errors. This source of error may be particularly true for loci underlying production related characteristics in domestic sheep (including horn morphology), where such variants could have been swept to fixation between the two species. An indication of this comes from Johnston et al. (2013) who examined haplotype sharing in the "horns region" among breeds of domestic sheep and 50 bighorn sheep that were typed on a 50k domestic sheep SNP chip. They showed that the in this horns region (a 300kb area on chromosome 10 containing 6 SNPs) all the bighorn sheep typed were fixed for a single haplotype, but the extent of haplotype sharing outside this region was small. Future studies could address this bias by testing for associations using loci specifically discovered in bighorn sheep (e.g. Genomics Resources Consortium *et al.*2013) and predicted to be in the region near RXFP2 and the other loci presented here.

5.4.5 Significance for Management

Recently there have been increasing calls for integration of genomic methodologies and insights into conservation programs (Shafer et al. 2015). In particular, to start to characterize the genetic variation underlying ecologically important traits so that such knowledge can be factored into management plans and help ensure long-term survival of species (Ashley et al. 2003; Harrisson et al. 2014; Hoffmann et al. 2015). In this paper, we showed three loci to be associated with fitness related morphological characters. Though we do not recommend making management decisions based on these handful of loci discovered in a single population, our

results add to calls for continued monitoring and expanded genetic testing (Coltman 2008). However, if the association is confirmed, and additional significant loci can be found using larger sample sizes and validated in additional populations, management actions such as translocations or 'genetic rescues' can be conducted with the express purpose of ensuring continued production of large horned sheep (Weeks *et al.* 2011).

5.4.6 Future Directions

The associations found here represent a step forward for finding the genes underlying fitness related traits in bighorn sheep. Future studies could build on these findings by expanding genotyping, both in terms of the number of loci uses as well as including additional individuals and populations. Considering more loci, possibly directly discovered in bighorn sheep (e.g. Genomic Resources Development Consortium *et al.* 2013), would allow for fine mapping of the observed associations, as well as detection of unobserved ones. While consideration of additional populations will allow for assessment of the consistency of associations observed. In addition, use of analyses besides GWAS could highlight novel associations or if a different genetic architecture better explains variation at these traits. For example, haplotype based analyses may have increased power to detect associations (Browning & Browning 2011), while chromosome partitioning methods can highlight if the traits fit a polygenic framework (Robinson *et al.* 2013; Santure *et al.* 2013; Yang *et al.* 2011b).

Table 5 - 1 Proportion of phenotypic variance after having accounted for fixed effects in the full datasets; standard errors generated by ASReml are shown in parentheses unless otherwise noted

Trait	Ind[1]	Obs[2]	Mean (s.d.)	Transformed data mean (s.d)	V_p	h^2	V_y	V_{yb}	V_{pe}
Horn Length	652	8011	27.40 (16.98)	6.62 (2.46)	0.85 (0.04)	0.15 (0.05)*	0.07 (0.02)*	0.10 (0.03)*	0.42 (0.05)*
Horn Base Circumference	637	7994	17.33 (8.33)	12.00 (4.49)	0.84 (0.04)	0.23 (0.05)*	0.08 (0.02)*	0.11 (0.03)*	0.27 (0.04)*
Body Mass	677	9552	58.69 (15.85)	7.39 (2.00)	0.58 (0.03)	0.20 (0.04)*	0.16 (0.03)*	0.07 (0.02)*	0.24 (0.04)*

[1]Number of individuals
[2]Number of phenotypic measurements
*P<0.00001

Table 5 - 2 Estimated random effect sizes for morphology associated suggestive loci. All models simultaneously considered individuals genotyped on the 700k SNPChip and by SNaPShot reactions. Standard errors generated by ASReml are shown in parentheses

Trait	Model	V_p	h^2	Locus[1]	V_{pe}	V_y	V_{yb}
Horn Length	Polygenic	0.83 (0.06)	0.18 (0.09)	NF	0.41 (0.09)	0.10 (0.03)	0.05 (0.04)
	Polygenic + OAR3_138991772	0.85 (0.08)	0.13 (0.08)	0.04 (0.06)*	0.43 (0.08)	0.10 (0.03)	0.05 (0.04)
Horn Base Circumference	Polygenic	0.90 (0.08)	0.39 (0.10)	NF	0.15 (0.08)	0.06 (0.02)	0.15 (0.06)
	Polygenic + OAR14_45166067	0.97 (0.18)	0.25 (0.10)	0.14 (0.15)*	0.18 (0.08)	0.06 (0.02)	0.13 (0.06)
Body Mass	Polygenic	0.55 (0.04)	0.27 (0.09)	NF	0.21 (0.07)	0.12 (0.03)	0.06 (0.04)
	Polygenic + OAR2_148529592	0.57 (0.07)	0.23 (0.08)	0.08 (0.09)*	0.18 (0.07)	0.12 (0.03)	0.06 (0.04)
Body Mass	Polygenic	0.55 (0.04)	0.27 (0.09)	NF	0.21 (0.08)	0.12 (0.03)	0.06 (0.04)
	Polygenic + OAR10_85023560	0.56 (0.06)	0.24 (0.09)	0.04 (0.07)	0.22 (0.07)	0.12 (0.03)	0.06 (0.04)

[1]NF = variable not fit
*P<0.05

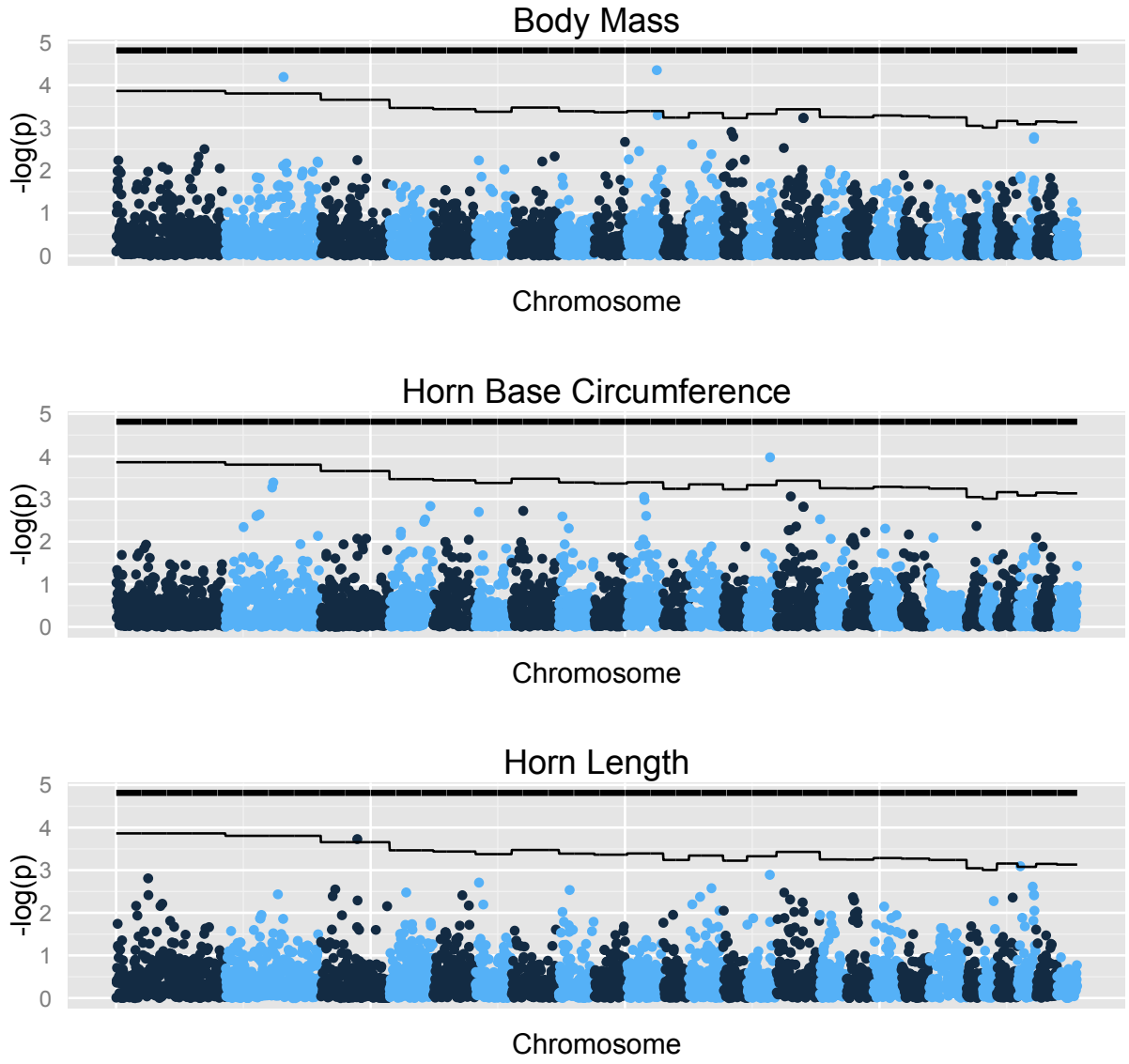


Figure 5 - 1 Manhattan plots for morphological characteristics
The thick line represents the genome-wide significance threshold, the thin lines represent per-chromosome significance thresholds. Positions are relative to the domestic sheep genome assembly (version 3.1).

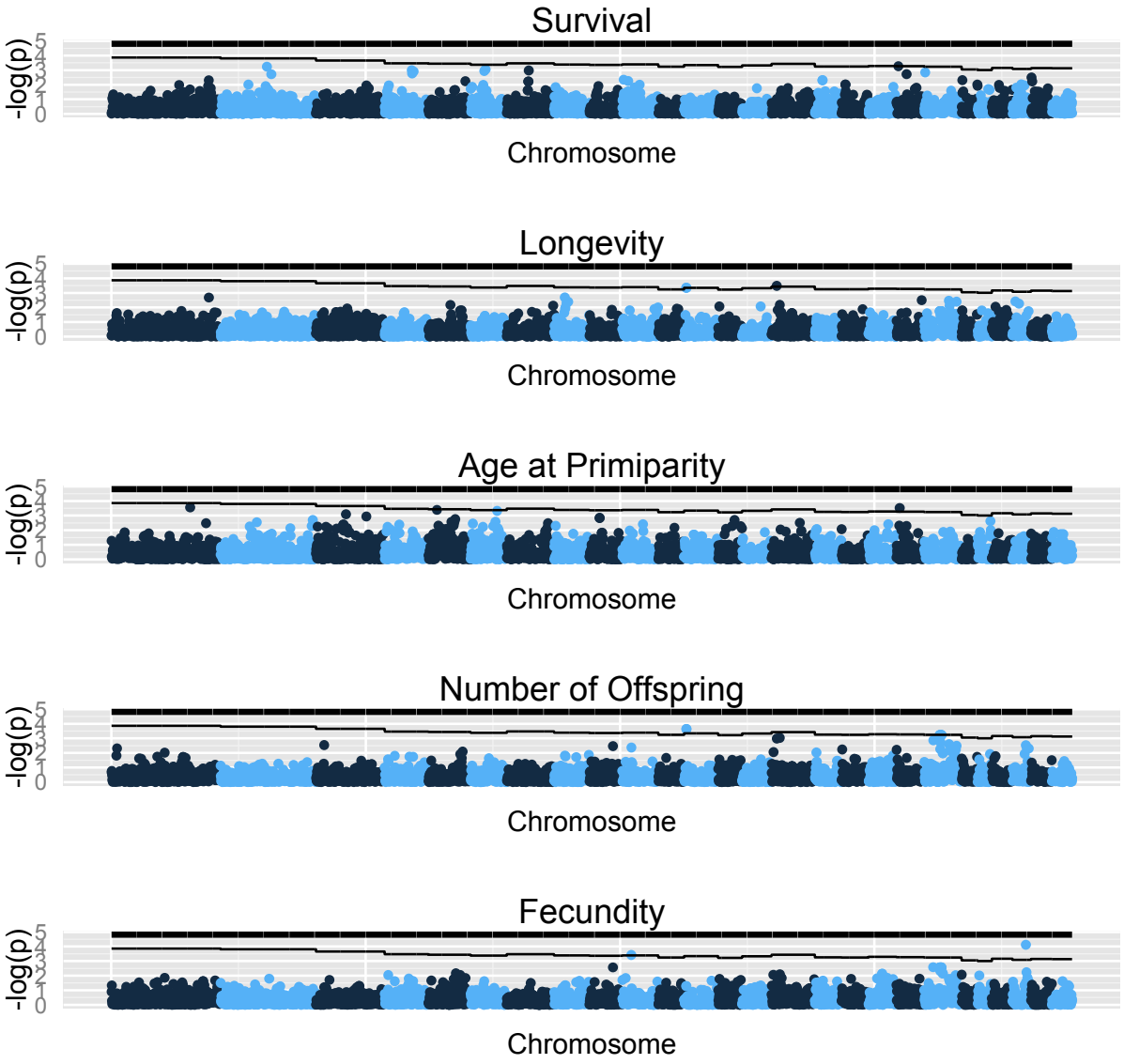


Figure 5 - 2 Manhattan plots for life-history characteristics
The dashed line represents the genome-wide significance threshold, the solid lines represent per-chromosome significance thresholds. Positions are relative to the domestic sheep genome assembly (version 3.1).

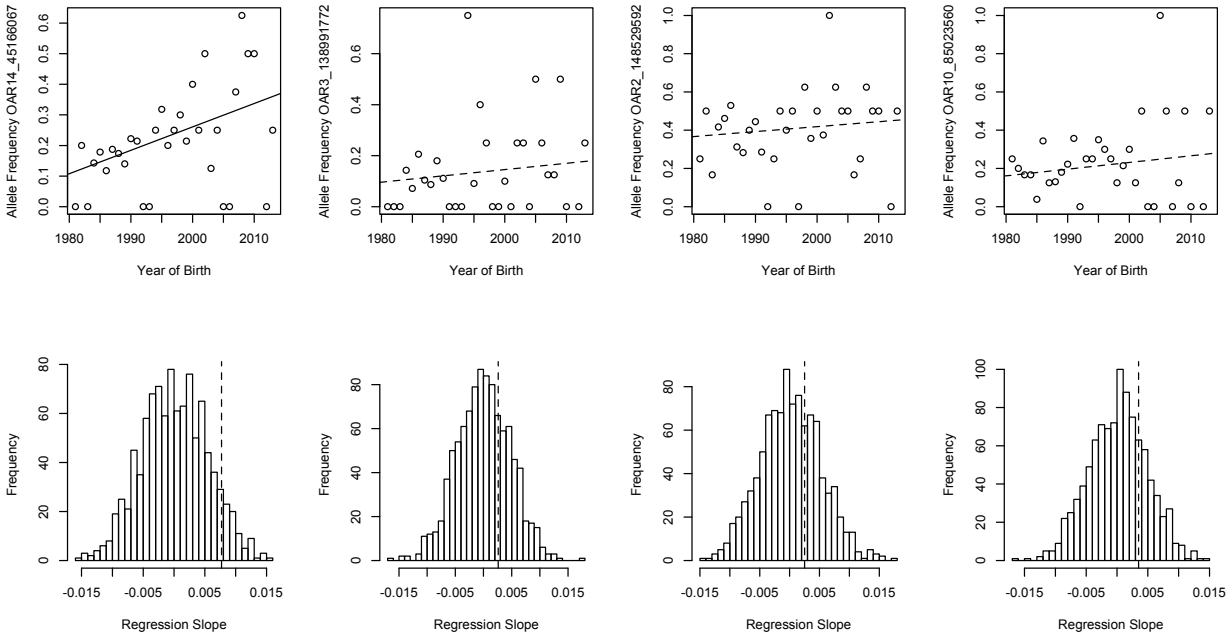


Figure 5 - 3 Changes in allele frequencies over time and results of gene dropping simulations

Top row shows plots of allele frequency in each cohort. Linear regression lines are shown where dashed lines are those showing non-significant changes over time. Bottom row shows histograms of the distribution of regression slopes from 1000 gene dropping simulations with vertical lines denoting the observed slopes.

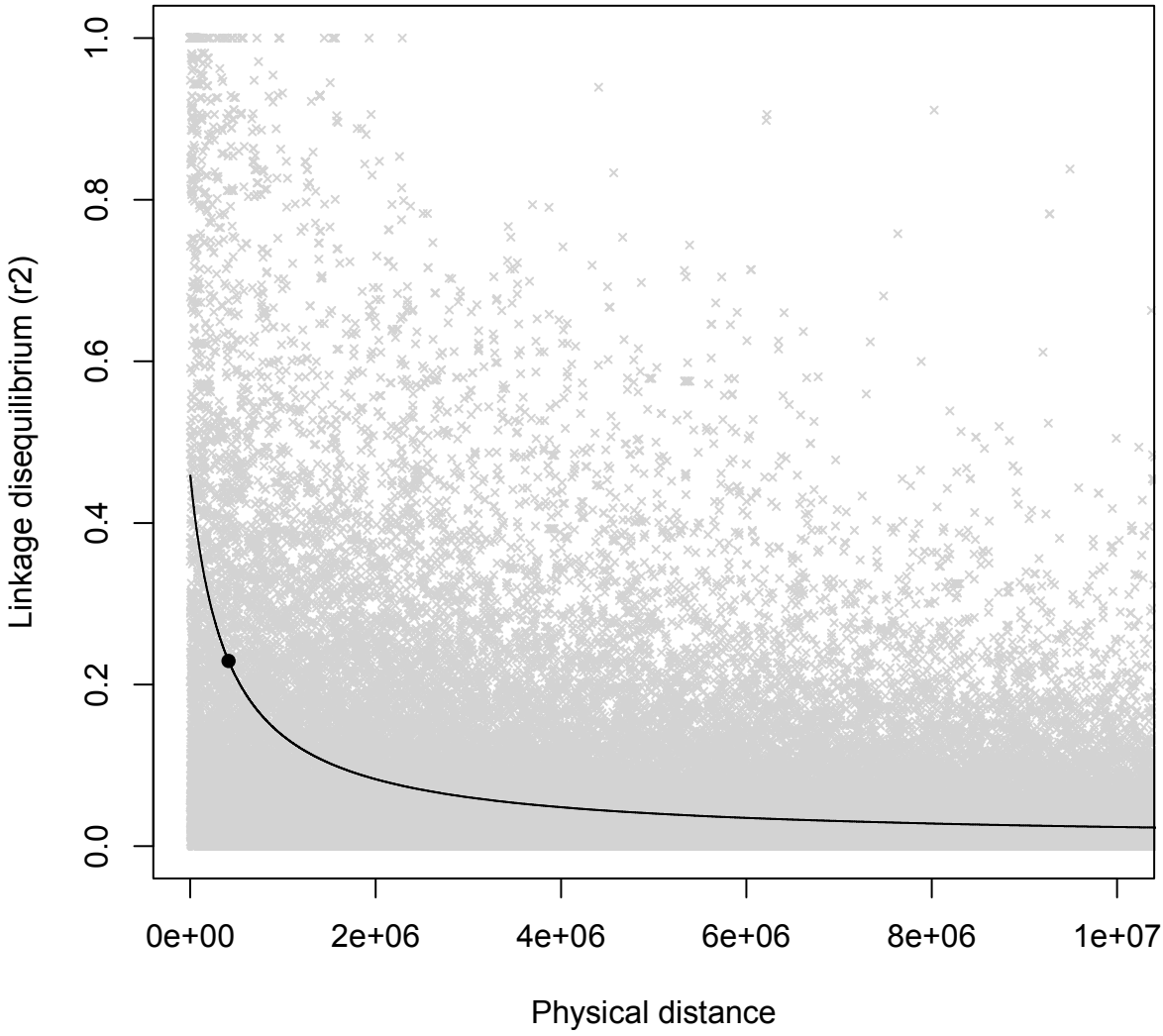


Figure 5 - 4 Scatterplot of LD estimates versus inter-markers distance
A non-linear least squares regression line is shown, with the round point indicating the half-length estimate.

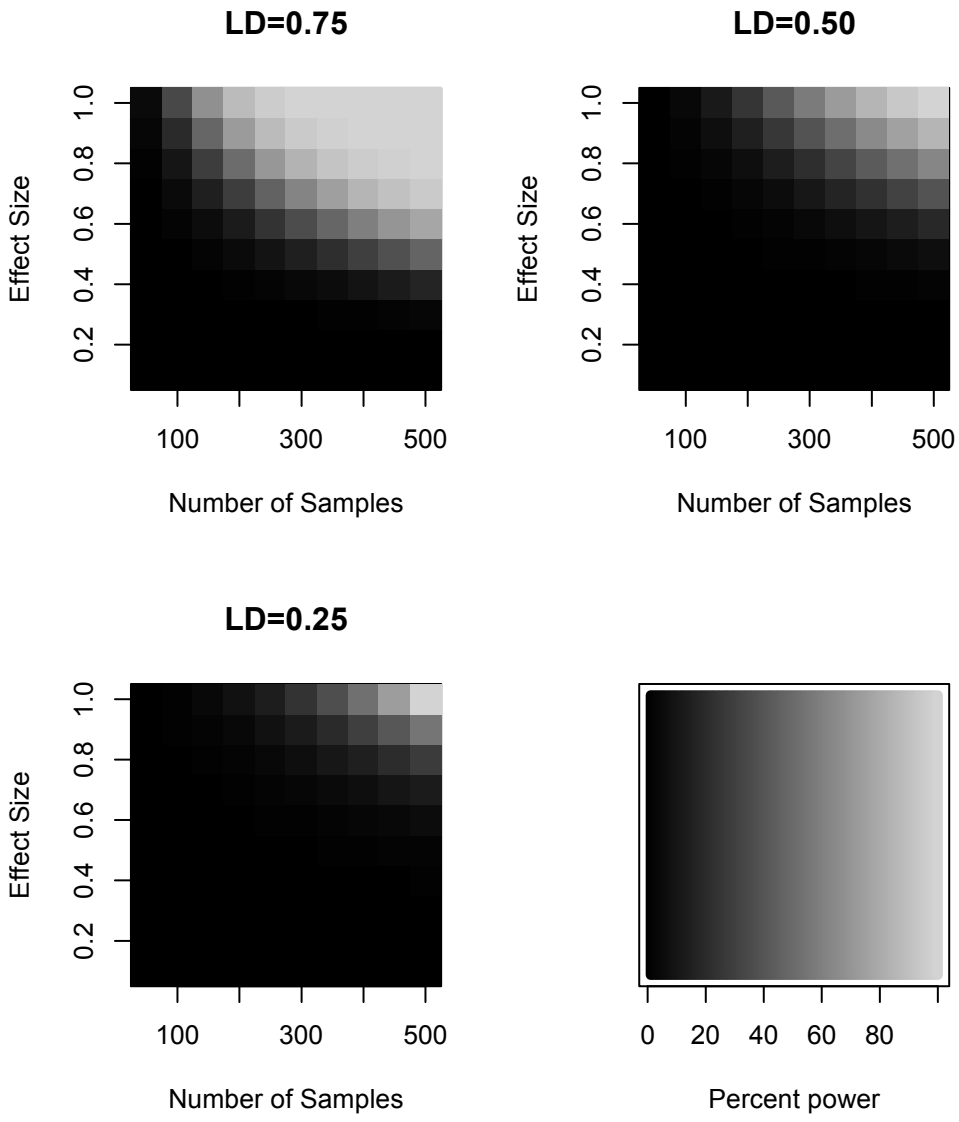


Figure 5 - 5 Heat maps of expected percent power of a GWAS as a function of sample size and effect size for a variety of linkage disequilibrium (LD) estimates

5.5 Bibliography

Allais-Bonnet A, Grohs C, Medugorac I*, et al.* (2013) Novel Insights into the Bovine Polled Phenotype and Horn Ontogenesis in Bovidae. *Plos One* **8**.

Ashley MV, Willson MF, Pergams ORW*, et al.* (2003) Evolutionarily enlightened management. *Biological Conservation* **111**, 115-123.

Backström N, Qvarnström A, Gustafsson L, Ellegren H (2006) Levels of linkage disequilibrium in a wild bird population. *Biology Letters* **2**, 435-438.

Bartoń K (2009) MuMIn: multi-model inference, Available at: http://r-forge.r-project.org/projects/mumin/.

Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6. http://CRAN.R-project.org/package=lme4.

Bonenfant C, Pelletier F, Garel M, Bergeron P (2009) Age-dependent relationship between horn growth and survival in wild sheep. *Journal of Animal Ecology* **78**, 161-171.

Bro-Jørgensen J (2007) The intensity of sexual selection predicts weapon size in male bovids. *Evolution* **61**.

Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. *Nat Rev Genet* **12**, 703-714.

Bunch T, Wu C, Zhang Y, Wang S (2006) Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. *Journal of Heredity* **97**, 21-30.

Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach* Springer Science & Business Media.

Chang C, Chow C, Tellier L*, et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**.

Chenoweth SF, McGuigan K (2010) The genetic basis of sexually selected variation. *Annual Review of Ecology, Evolution, and Systematics* **41**, 81-101.

Coltman DW (2005) Testing marker-based estimates of heritability in the wild. *Molecular Ecology* **14**, 2593-2599.

Coltman DW (2008) Molecular ecological approaches to studying the evolutionary impact of selective harvesting in wildlife. *Molecular Ecology* **17**, 221-235.

Coltman DW, Festa-Bianchet M, Jorgenson JT, Strobeck C (2002) Age-dependent sexual selection in bighorn rams. *Proceedings of the Royal Society B-Biological Sciences* **269**, 165-172.

Coltman DW, O'Donoghue P, Hogg JT, Festa-Bianchet M (2005) Selection and genetic (CO)variance in bighorn sheep. *Evolution* **59**, 1372-1382.

Coltman DW, O'Donoghue P, Jorgenson JT*, et al.* (2003) Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655-658.

Davis EB, Brakora KA, Lee AH (2011) Evolution of ruminant headgear: a review. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2857-2865.

Dumont BL, Payseur BA (2008) Evolution of the genomic rate of recombination in mammals. *Evolution* **62**, 276-294.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature* **452**, 169-175.

Elshire R, Glaubitz J, Sun Q*, et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One* **6**.

Favre M, Martin JG, Festa-Bianchet M (2008) Determinants and life-history consequences of social dominance in bighorn ewes. *Animal Behaviour* **76**, 1373-1380.

Feng S, Ferlin A, Truong A*, et al.* (2009) INSL3/RXFP2 signaling in testicular descent. *Annals of the New York Academy of Sciences* **1160**, 197-204.

Ferlin A, Pepe A, Gianesello L*, et al.* (2008) Mutations in the insulin-like factor 3 receptor are associated with osteoporosis. *Journal of Bone and Mineral Research* **23**, 683-693.

Festa-Bianchet M, Coltman DW, Hogg JT, Jorgenson JT (2008) Age-related horn growth, mating tactics, and vulnerability to harvest: why horn curl limits may select for small horns in bighorn sheep. *Biennial Symposium of the Northern Wild Sheep and Goat Council* **15**, 42-49.

Festa-Bianchet M, Jorgenson J, King W, Smith K, Wishart W (1996) The development of sexual dimorphism: Seasonal and lifetime mass changes in bighorn sheep. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **74**, 330-342.

Festa-Bianchet M, Pelletier F, Jorgenson JT, Feder C, Hubbs A (2014) Decrease in horn size and increase in age of trophy sheep in Alberta over 37 years. *The Journal of Wildlife Management* **78**, 133–141.

Flicek P, Amode M, Barrell D*, et al.* (2014) Ensembl 2014. *Nucleic Acids Research* **42**, D749-D755.

García-Gámez E, Sahana G, Gutiérrez- Gil B, Arranz J-J (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* **13**.

Gautier M, Naves M (2011) Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology* **20**, 3128-3143.

Genomic Resources Development Consortium, Coltman DW, Hogg JT, Miller JM (2013) Genomic Resources Notes accepted 1 April 2013–31 May 2013. *Molecular Ecology Resources* **13**, 965-965.

Gilmour A, Gogel B, Cullis B, Thompson R (2009) *ASReml User Guide. Release 3.0.* VSN International Ltd, Hemel Hempstead, UK.

Gratten J, Pilkington JG, Brown EA*, et al.* (2012) Selection and microevolution of coat pattern are cryptic in a wild population of sheep. *Molecular Ecology* **21**, 2977–2990.

Hadfield JD, Wilson AJ, Garant D, Sheldon BC, Kruuk LEB (2010) The misuse of BLUP in ecology and evolution. *American Naturalist* **175**, 116-125.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* **7**, 1008–1025.

Hedrick P, Coltman D, Festa-Bianchet M, Pelletier F (2014) Not surprisingly, no inheritance of a trait results in no evolution. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4810-E4810.

Hengeveld PE, Festa-Bianchet M (2011) Harvest regulations and artificial selection on horn size in male bighorn sheep. *The Journal of Wildlife Management* **75**, 189-197.

Hill W, Weir B (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology* **33**, 54-78.

Hoffmann A, Griffin P, Dillon S*, et al.* (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* **2**, 1-24.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* **6**, e1000862.

Hu Z, Park C, Wu X, Reecy J (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research* **41**, D871-D879.

Hu Z-L, Fritz ER, Reecy JM (2007) AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucl. Acids Res.* **35**, D604-609.

Johnston S, McEwan J, Pickering N*, et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* **20**, 2555-2566.

Johnston SE, Beraldi D, McRae AF, Pemberton JM, Slate J (2010) Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity* **104**, 196-205.

Johnston SE, Gratten J, Berenos C*, et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93-95.

Jorgenson JT, Festa-Bianchet M, Gaillard J-M, Wishart WD (1997) Effects of age, sex, disease, and density on survival of bighorn sheep. *Ecology* **78**, 1019-1032.

Jorgenson JT, Festa-Bianchet M, Lucherini M, Wishart WD (1993) Effects of body size, population density, and maternal characteristics on age at first reproduction in bighorn ewes. *Canadian Journal of Zoology* **71**, 2509-2517.

Jorgenson JT, Festa-Bianchet M, Wishart WD (1998) Effects of population density on horn development in bighorn rams. *The Journal of Wildlife Management* **62**, 1011-1020.

Jorgenson JTF-B, M.Wishart, W. D. (1993) Harvesting bighorn ewes: consequences for population size and trophy ram production. *The Journal of Wildlife Management* **57**, 429-435.

Kawakami T, Backström N, Burri R*, et al.* (2014) Estimation of linkage disequilibrium and interspecific gene flow in Ficedula flycatchers by a newly developed 50k SNP array. *Molecular Ecology Resources* **14**, 1248–1260.

Kijas JW, Lenstra JA, Hayes B*, et al.* (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol* **10**, e1001258.

Kijas JW, Porto-Neto L, Dominik S*, et al.* (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754–757.

Kinsella R, Kahari A, Haider S*, et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database-the Journal of Biological Databases and Curation*.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**, 873-890.

Kruuk LEB, Slate J, Wilson AJ (2008) New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology Evolution and Systematics* **39**, 525-548.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-577.

Manolio T, Collins F, Cox N*, et al.* (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747-753.

Martin AM, Presseault-Gauvin H, Festa-Bianchet M, Pelletier F (2013) Male mating competitiveness and age-dependent relationship between testosterone and social rank in bighorn sheep. *Behavioral Ecology and Sociobiology* **67**, 919-928.

McFarlane S, Gorrell J, Coltman D*, et al.* (2014) Very low levels of direct additive genetic variance in fitness and fitness components in a red squirrel population. *Ecology and Evolution* **4**, 1729-1738.

McFarlane SE, Gorrell JC, Coltman DW*, et al.* (2015) The nature of nurture in a wild mammal's fitness. *Proceedings of the Royal Society of London B: Biological Sciences* **282**.

McKay S, Schnabel R, Murdoch B*, et al.* (2007) Whole genome linkage disequilibrium maps in cattle. *Bmc Genetics* **8**.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Poissant J, Kijas J, Coltman DW, TISGC (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Minikel E (2012) *Power for GWAS and extreme phenotype studies.* http://www.cureffi.org/2012/12/05/power-for-gwas-and-extreme-phenotype-studies/

Moskvina V, Schmidt K (2008) On multiple-testing correction in genome-wide association studies. *Genetic Epidemiology* **32**, 567-573.

Paterson T, Graham M, Kennedy J, Law A (2012) VIPER: a visualisation tool for exploring inheritance inconsistencies in genotyped pedigrees. *BMC bioinformatics* **13**, S5.

Pelletier F, Festa-Bianchet M, Jorgenson J (2012) Data from selective harvests underestimate temporal trends in quantitative traits. *Biology Letters* **8**, 878-881.

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2012) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity* **108**, 256–263.

Poissant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW (2008) Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proceedings of the Royal Society B-Biological Sciences* **275**, 623-628.

Porto-Neto L, Kijas J, Reverter A (2014) The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genetics Selection Evolution* **46**.

Purcell S, Cherny S, Sham P (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149-150.

Purcell S, Neale B, Todd-Brown K*, et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575.

Quinlan A, Hall I (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.

R Core Team (2012) R: A language and environment for statistical computing, reference index version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2014) R: A language and environment for statistical computing, reference index version 3.2.1. R Foundation for Statistical Computing, Vienna, Austria.

Reich DE, Cargill M, Bolk S*, et al.* (2001) Linkage disequilibrium in the human genome. *Nature* **411**, 199-204.

Robinson MR, Santure AW, DeCauwer I, Sheldon BC, Slate J (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology* **22**, 3963-3980.

Réale D, Festa-Bianchet M (2000) Quantitative genetics of life-history traits in a long-lived wild mammal. *Heredity* **85**, 593-603.

Santure A, De Cauwer I, Robinson M*, et al.* (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. *Molecular Ecology* **22**, 3949-3962.

Shafer ABA, Wolf JBW, Alves PC*, et al.* (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**, 78–87.

Sham P, Cherny S, Purcell S, Hewitt J (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *American Journal of Human Genetics* **66**, 1616-1630.

Slate J, Gratten J, Beraldi D*, et al.* (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97-107.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158-170.

Traill LW, Schindler S, Coulson T (2014) Demography, not inheritance, drives phenotypic change in hunted bighorn sheep. *Proceedings of the National Academy of Sciences* **111**, 13223–13228.

Wang J (2013) An improvement on the maximum likelihood reconstruction of pedigrees from marker data. *Heredity* **111**, 165-174.

Weeks AR, Sgro CM, Young AG*, et al.* (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* **4**, 709–725.

Wickham H (2009) *ggplot2: elegant graphics for data analysis* Springer, New York.

Wiedemar N, Tetens J, Jagannathan V*, et al.* (2014) Independent Polled Mutations Leading to Complex Gene Expression Differences in Cattle. *Plos One* **9**.

Wilson AJ, Kruuk LEB, Coltman DW (2005) Ontogenetic Patterns in Heritable Variation for Body Size: Using Random Regression Models in a Wild Ungulate Population. *The American Naturalist* **166**.

Wilson AJ, Réale D, Clements MN*, et al.* (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology* **79**, 13-26.

Yang J, Benyamin B, McEvoy BP*, et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-569.

Yang J, Lee S, Goddard M, Visscher P (2011a) GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics* **88**, 76-82.

Yang J, Manolio T, Pasquale L*, et al.* (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519-U544.

Yang J, Zaitlen N, Goddard M, Visscher P, Price A (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* **46**, 100-106.

Yuan FPL, X Lin, J Schwabe, C Büllesbach, E E Rao, C V Lei, Z M (2010) The role of RXFP2 in mediating androgen-induced inguinoscrotal testis descent in LH receptor knockout mice. *Reproduction* **139**, 759-769.

Chapter 6

HARNESSING CROSS-SPECIES ALIGNMENT TO GENERATE A DRAFT GENOME OF A BIGHORN SHEEP (*OVIS CANADENSIS*)

A version of this chapter has been accepted for publication as:

Joshua M Miller, Stephen S Moore, Paul Stothard, Xiaoping Liao, and David W. Coltman. "Harnessing cross-species alignment to discover SNPs and generate a draft genome sequence of a bighorn sheep (*Ovis canadensis*)" *BMC Genomics* 16, no. 1 (2015): doi: 10.1186/s12864-015-1618-x

6.1 Background

Widespread use of high-throughput sequencers has allowed an ever-increasing number of species to have a whole genome sequence (WGS) prepared. While many of these have been model or domestic organisms, a wide array of taxa continue to be sequenced (as reviewed in Ellegren 2014). WGS opens the door for a multitude of subsequent analyses including: 1) creation of phylogenies and assessment of broader evolutionary patterns and innovations (Prado-Martinez *et al.* 2013; Telford & Copley 2011). 2) Annotation of genes (Yandell & Ence 2012) and identification of rearrangements or gene expansions (Bourque *et al.* 2005; Zhao & Bourque 2009). 3) Discovery of large sets of markers (Dalloul *et al.* 2010; Fang *et al.* 2011). 4) Resequencing studies, including those that are genome-wide yet not full coverage (e.g. transcriptomics or reduced representation sequencing) but benefit from the presence of a reference genome (Vijay *et al.* 2013). Resequencing at any scale also allows for 'population genomics' including investigations of local adaptation or population differentiation (Angeloni *et al.* 2012; Funk *et al.* 2012), demographic history (Li & Durbin 2011; Sheehan *et al.* 2013), and the genetic basis of phenotypic traits (Daetwyler *et al.* 2014).

In the absence of a reference, construction of a WGS necessitates *de novo* methodologies (Miller *et al.* 2010). These methods require large volumes of raw sequence data which are arranged into contigs and then joined to scaffolds by either computational methods (Hunt *et al.* 2014), anchoring with outside information (e.g. a linkage map, BACs, or FISH), or continued sequencing (Ekblom & Wolf 2014). Such an endeavor is still relatively expensive and challenging in terms of the bioinformatics involved, making it beyond the capability of many research programs. However, the presence of a reference sequence enables reads to be aligned to the reference which is much faster and allows for lower sequence depths than *de novo* assembly (Ekblom & Wolf 2014; Martin & Wang 2011). Recent work has highlighted that the reference need not come from the same species the reads are from (Gnerre *et al.* 2009; Kim *et al.* 2013;

Wang *et al.* 2014) opening these methods to a wide array of 'genome-enabled' taxa (Kohn *et al.* 2006).

There are a number of reasons why we are motivated to produce a bighorn sheep (*Ovis canadensis*) WGS. First, this species has a complex demographic history in North America that has been profoundly influenced by anthropogenic activity, having experienced intense hunting, local extirpations and disease-related die-offs, as well as translocations and reintroductions throughout its range (Berger 1990; Festa-Bianchet *et al.* 2014; Hedrick 2014; Johnson *et al.* 2011; Olson *et al.* 2013; Shackleton *et al.* 1999). These events are expected to have significant genetic/genomic consequences (Coltman *et al.* 2003; Hedrick 2014; Olson *et al.* 2013) that merit further study. Second, there are several long-term study populations in which individual based questions such as the genetic basis of complex traits (Poissant *et al.* 2012; Réale *et al.* 2009) and linkages between individual genetic variation and differences in fitness (Miller *et al.* 2014; Miller *et al.* 2012b) can be addressed using genomic data. Finally, bighorn sheep are an excellent candidate for cross-species approaches since genomic resources for domestic sheep (*Ovis aries*, (Jiang *et al.* 2014; Kijas *et al.* 2014)) can be easily applied to bighorn sheep as they are a close relative (~3 million years divergent; Bunch *et al.* 2006) and are expected to share a high degree of genomic synteny (Poissant *et al.* 2010). Genomic resources have been recently developed for bighorn sheep (Genomic Resources Development Consortium *et al.* 2013; Miller *et al.* 2012a; Miller *et al.* 2011; Poissant *et al.* 2010; Poissant *et al.* 2009), but future resequencing efforts would be aided by species specific genomic sequencing data.

Here we use cross-species alignment to create a draft genome from a single ram sequenced using ABI SOLiD technology. The pros and cons of different high-throughput sequencers have been discussed at length elsewhere (Branton *et al.* 2008; Ekblom & Galindo 2011; Glenn 2011; Metzker 2010). Choice of a specific platform balances read length, the amount of sequence data output, error profiles, and cost. SOLiD technology is well-suited for resequencing studies as it returns high volumes of data and the sequence-by-ligation strategy is

able to distinguish sequencing errors from true nucleotide variants during alignment (McKernan *et al.* 2009; Ondov *et al.* 2008). Based on our alignment we called variants, annotated SNP relative to domestic sheep, and conducted enrichment analysis of those SNPs showing fixed differences.

6.2 Methods

6.2.1 Sample Collection & Sequencing

Total genomic DNA was extracted from tissue of a single bighorn sheep from Ram Mountain (Alberta, Canada), using standard phenol–chloroform extraction protocols (Sambrook & Russell 2001). From this, two libraries were constructed and sequenced. The first was a mate-paired library the details of which are provided in (Miller *et al.* 2012a). Briefly, preparation used the reagents and protocols provided by Applied Biosystems with and an expected insert size of ~1.5kb. Emulsion PCR was performed using the SOLiD EZ bead system (Life Technologies Corporation). Both forward and reverse tags were sequenced to 50 bases using an Applied Biosystems SOLiD 4 sequencer (Life Technologies Corporation). The second library was a fragment library sequenced to 75 bases using a SOLiD 5500xl sequencer (Life Technologies Corporation). The resulting xsq files were converted to csfasta and qual scores format using XSQ Tool (Life Technologies Corporation).

6.2.2 Alignment & Variant Calling

Sequence quality assessment and alignment were conducted with CLC Genomics Workbench (version 5.0; CLC bio, Cambridge, MA, USA). For each library, sequences were quality trimmed allowing for 1 ambiguous nucleotide, a quality score limit of 0.05, and minimum read length of 15 nucleotides. The resulting reads from each library were then independently aligned to domestic sheep chromosomes (version 3; Jiang *et al.* 2014). Alignment

parameters were set with no reference masking, mismatch cost of 2, insertion/deletion cost of 3, length fraction of 0.5, and similarity fraction of 0.8. Meaning at least 50% of a read must have at least 90% identity to the reference to be aligned. Non-specific matches were mapped randomly. Once mapped, PCR duplicates were removed from the alignment. We then merged the two mate-paired and fragment mappings and removed PCR duplicates from the merged file. The merged alignment was then exported both as consensus fasta sequences as well as a BAM file for use in subsequent analyses. When generating the consensus fasta sequences we allowed for ambiguities (e.g. IUPAC codes W, R, etc.) and inserted N's proportional to the length of the domestic sheep reference for regions of zero coverage. We elected to leave differences between our bighorn sheep sequence and domestic sheep reference as ambiguities in case additional sequencing reveals those sites to represent unobserved shared polymorphisms.

Variants were called from the consensus alignment using the mpileup command in SAMtools version 0.1.17 (Li *et al.* 2009) and filtered in bcftools. Specifically, variants were required to have a minimum quality of 30 and a read depth between 6 and 200. VCFfilter version 0.1.11 (Danecek *et al.* 2011) was then used to assess indel length distribution and calculate transition transversion (ti/tv) ratio using 100 basepair windows. As a quality check, genotypes from the aligned genome were compared to those generated for the same individual on the Ovine Infinium®HD SNP BeadChip, a newly developed SNP array for domestic sheep that contains 606,006 loci (Kijas *et al.* 2014). For this analysis raw intensity data were converted into genotype calls using GenomeStudio (Illumina) and SNP cluster information based on domestic sheep reference samples provided by the International Sheep Genomics Consortium. All genotype calls with GenCall scores less than 0.6, or GenTrain scores lower than 0.8, were removed from the data set. When assessing concordance between genotypes from the SNP array and the draft WGS we first positioned SNPs from the array in the reference assembly by comparing 50 nucleotides on either side of the locus position using BLAST with an E value of $1e^{-9}$. Loci with more than one match were excluded from analysis. In total this procedure

excluded 45,979 loci. To assess the effects of filtering on the recovery of chip SNPs by sequencing and on concordance between the chip and the sequence genotypes an additional set of filtering criteria was applied to the sequence-derived SNPs. In this case we increased stringency, requiring read depths greater than 5 but less than the mean plus 3 SD, at least one forward or reverse alternative allele read (where applicable), no other variants within 5 bp, and genotype quality greater than 10.

6.2.3 Annotation & enrichment analysis

SnpEff version 3.1 (Cingolani *et al.* 2012b) was used to predict functional classes (e.g. intergenic or intronic) and effect types (e.g. synonymous or non-synonymous) of the loci by comparing our SNPs to annotations from the domestic sheep genome (database oar3.1, downloaded Sept 2013). Note that within functional classes and effect types, categories are not mutually exclusive, for example a SNP can be classified as both intronic and in the 5'-UTR.

For enrichment analysis we first filtered SNPs to only those that were fixed between our bighorn sheep alignment and the domestic sheep reference using SNPsift (Cingolani *et al.* 2012a). We then split the resulting loci into two categories: 1) those with likely functional consequences (i.e. non-synonymous coding, start gained, start lost, stop gained, stop lost) and 2) those showing synonymous effects (i.e. synonymous coding, synonymous start). GO terms were added to the SNPs in these lists from the *Ovis aries* gene set (Oar v3.1) using BioMart (Kinsella *et al.* 2011) and Ensembl version 77 (Flicek *et al.* 2014). The two groups were then compared using BLAST2GO (Conesa *et al.* 2005) which employs a Fisher's Exact Test via the Gossip package (Blüthgen *et al.* 2005). Specifically, we used a two tailed test with false discovery correction of Benjamini and Hochberg (1995) set at 0.0001. Evaluation of GO enrichment among candidate genes was restricted to terms within the biological process category.

6.3 Results

6.3.1 SOLiD Sequencing & Alignment

Prior to trimming the 50 x 50bp mate-paired library contained 311,847,628 reads, while the 75bp fragment library contained 555,575,794 reads. Post-trimming, read count was reduced to 218,239,459 (70% retained) and 506,697,724 (91% retained) for the mate-paired and fragment libraries respectively. When aligned on its own the mate-paired library had 174,894,731 reads map to the reference, of which 115,727,618 were in pairs with an average distance of 1108 nucleotides between pairs, while the fragment library had 377,008,050 reads map to the reference. Once merged, the two libraries covered 95% of the reference genome with an average read depth of 12.29 (104 SD).

6.3.2 Variant Calling

In total, 15,622,884 variants (14,583,355 SNPs and 1,039,529 indels) passed our filtering thresholds and were called compared to the domestic sheep reference (Supplementary Table 1). Of the putatively bi-allelic SNPs relative to the domestic sheep reference, 9,831,700 were transitions and 4,320,985 were transversions (ti/tv = 2.275; which is similar to the 2.1 ratio observed for genomic data in many mammalian studies (Wakeley 1996)). Insertions were slightly more common than deletions (Supplementary Figure 6-S1). Of the 606,006 loci present on the Ovine Infinium®HD SNP BeadChip 422,975 loci were successfully genotyped in our bighorn sheep sample. Note that a decrease in amplification success is expected from cross-species application of SNP chips (Miller *et al.* 2012a; Sechi *et al.* 2010). 407,465 (~96%) of these chip loci were present in the list of variants identified by sequencing, and over 93% of the loci showed agreement (Table 1). To assess the effects of filtering on these results an additional set of filtering criteria was applied to the sequence-derived SNPs. Increasing our stringency thresholds for SNPs in the WGS decreased the number chip loci that were present in the list of

SNPs identified by sequencing (n = 329,690; ~78%), but increased concordance to ~95%. In both cases the major source of discordance was loci called heterozygous in the WGS but homozygous from the chip data (Table 1).

6.3.3 Annotation & Enrichment Analysis

SnpEff assigned 18,176,092 functional classes to our SNPs based on annotation of the domestic sheep genome. Note that the number of classes assigned is larger than the number of SNPs due to the fact that categories are not mutually exclusive. The vast majority of the SNPs were predicted to be intronic or intergenic and 102,231 were assigned to coding regions or have predicted functional effects (Figure 1, Supplementary Table 6-S1). Of those 102,231 loci, 52,381 SNPs were found to have fixed differences between our bighorn sheep and the domestic sheep reference, from which 25,472 were identified as non-synonymous and 27,198 were identified as synonymous. Note that sum of the number of synonymous and non-synonymous SNPs is larger than the total number of fixed differences because a locus may be classified as both synonymous and non-synonymous if a gene has more than one annotated transcript. Gene Ontology (GO) terms were available for 26,629 of the SNPs with fixed differences (9,752 non-synonymous and 16,877 synonymous) representing 6,963 genes (3,948 non-synonymous and 5,932 synonymous). We looked for functional enrichment between non-synonymous and synonymous SNPs using BLAST2GO (Conesa *et al.* 2005). When reduced to the most specific GO terms, we found 11 GO terms to be over represented and 29 to be underrepresented in the non-synonymous set compared to the synonymous set (Supplementary Table 6-S2). Note that gene length was positively correlated to the number of annotated SNPs for both the non-synonymous and synonymous sets (r= 0.43 and 0.61 respectively). But given that this association was constant between both non-synonymous and synonymous gene sets we do not think it biases our results. However, one gene, titin, was ~3 times larger than any other genes considered so we repeated the GO enrichment analysis dropping titin, which reduced the level of correlation (r= 0.37 and 0.51

respectively). When titin is removed from the datasets the number of overrepresented and underrepresented terms decrease to 9 and 15 respectively; all of which were common to the set including titin, except for one underrepresented term (cellular protein metabolic process; GO 0044267) that was unique to the second analysis (Supplementary Table 6-S2).

6.4 Discussion

Here we present a draft bighorn sheep WGS created by cross-species alignment to a domestic sheep reference sequence. Other studies have attempted *de novo* assembly with SOLiD sequencing data (Cerdeira *et al.* 2011; Genomic Resources Dev Consortium *et al.* 2014; Umemura *et al.* 2013), but this was not an option in our case due to the high read depth required by such methods for a mammalian sized genome. Our work more closely resembles that of Canavez *et al.* (Canavez *et al.* 2012) and Wang *et al.* (2014). Canavez *et al*. (2012) created a draft genome for an indicine bull (*Bos indicus*) through alignment of SOLiD reads to a taurine cow (*Bos taurus*) reference genome (divergence ~250 kya) (Canavez *et al.* 2012). While Wang *et al*. (2014) used SOLiD sequencing in a reference guided assembly of a black grouse (*Tetrao tetrix*) draft genome. However, Wang *et al.* (2014) used a combination of *de novo* and alignment methods as the large divergence time between black grouse and domestic chicken (*Gallus gallus*) used as a reference (~30-40 mya) may hinder sequences from aligning properly. In contrast, bighorn and domestic are much less divergent which allows for successful direct alignment of reads: over 76% of our quality filtered reads mapped to the reference genome. Once merged, our two sequencing libraries provided 95% coverage of the reference and average 12x (104 SD) sequence depth.

Our alignment produced a large database of SNP markers for future studies. Approximately 6% of genotypes from a high-density SNP chip were discordant with those from the genome alignment, and increasing the quality thresholds for loci discovered in the genome only marginally decreased mismatches to ~4%. In both cases the major source of discordance

was loci called heterozygous in the genome alignment but homozygous from the SNP chip. This source of discordance could be caused by incorrect joining of paralogous regions due to our procedure of randomly mapping ambiguous alignments or sequencing errors. However, given the overall high concordance between the genome aligned SNPs and those on the SNP chip we are confident that the majority of our genotypes represent real SNPs. These markers add to the set of SNPs already available for this species (Genomic Resources Development Consortium *et al.* 2013; Miller *et al.* 2011).

Genome scans of domestic sheep breeds have shown a number of regions that have been differentiated due to domestication (Jiang *et al.* 2014; Kijas *et al.* 2012). Therefore, we expect alleles associated with production traits to have been swept to or near fixation relative to a wild ancestor as well. Our GO term analysis of fixed SNP differences compared to the domestic sheep reference highlighted 40 biological process GO terms with significantly different representation in SNPs tagged as non-synonymous versus synonymous. Two of the gene ontologies that were associated with amino-acid changes relative to the domestic sheep reference involved reproduction: spermatogenesis (GO:0007283), and negative regulation of mammary gland epithelial cell proliferation (GO:0033600). This mirrors recent work that has highlighted the genetic effects domestication had on reproductive traits of several sheep breeds (Kijas *et al.* 2012; Lv *et al.* 2014). Another term that was over-represented in the non-synonymous gene set was ossification involved in bone maturation (GO:0043931). This term is noteworthy given the relationship of bones to horns which are bony projections covered by a keratinous sheath (Davis *et al.* 2011). Horns are a major determinant of reproductive success in bighorn sheep, where larger males with bigger horns win antagonistic encounters and gain access to females (Coltman *et al.* 2002; Pelletier & Festa-Bianchet 2006); however, in many breeds of domestic sheep horns have been selected against leading to gene-level consequences (Kijas *et al.* 2012). All but two of the overrepresented biological process terms (skeletal muscle adaptation (GO:0043501) and

maintenance of fidelity involved in DNA-dependent DNA replication (GO:0045005)) remained significant when titin (the largest gene in the dataset) was removed from the analysis.

For genes less likely to have amino acid changes, 14 of the 29 GO terms were related to muscles or muscle fibers, particularly cardiac muscles, e.g.: cardiac muscle hypertrophy (GO:0003300), cardiac myofibril assembly (GO:0055003), cardiac muscle fiber development (GO:0048739), adult heart development (GO:0007512), regulation of relaxation of cardiac muscle (GO:1901897), sarcomerogenesis (GO:0048769). It is interesting to note these differences associated with muscle properties, given that the domestic sheep reference genome was built from a meat-producing breed, the Texel (Clop *et al.* 2006; Jiang *et al.* 2014). As mentioned above, body size is an important life history characteristic for male bighorn sheep as it relates to access to females, while larger females have been found to have longer lifespans (Gaillard *et al.* 2000). Selective breeding for meat production in domestic sheep could favor conservation of the genes or developmental pathways that produce large muscles in bighorn sheep. However, analysis with REVIGO (Supek *et al.* 2011) indicated that there was overlap in these GO terms with 10 terms falling into two more representative terms: cardiac muscle hypertrophy (GO:0003300; containing two other terms) and cardiac muscle tissue morphogenesis (GO:0055008; containing eight other terms). In addition, nine of these terms become non-significant when titin (which has known associations with muscle properties, including body size, in cattle (Bos taurus; Lee *et al.* 2013; Sasaki *et al.* 2006) and pigs (Sus scrofa; Braglia *et al.* 2013)) is removed from the datasets.

Two factors are important to keep in mind when interpreting the results of our GO analysis. The first is that though it is tempting attribute the majority of differences we observed here to domestication and selective breeding, there are likely to be additional factors at play. In particular, natural selection, as bighorn sheep and the progenitor to domestic sheep diverged, as well as genetic drift. Second, we are only comparing SNP sites from one individual's genome to a reference sequence. This likely results in missing polymorphisms within either species, leading

to incorrect annotation of fixed differences. However, we present the results only as a preliminary analysis to highlight candidate ontologies that may contribute to differentiation between the species. Such results will need to be confirmed by additional sequencing, alternate analyses (e.g. genome scans), and perhaps functional characterization (Stinchcombe & Hoekstra 2008).

While our draft genome represents a step forward in the genomic resources available for bighorn sheep this single genome is representative of a specific demographic history, an example of the 'n = 1 constraint' (Buerkle *et al.* 2011). Future population genomic studies using additional individuals from Ram Mountain or other populations can confirm the variants we describe here, discover additional variants, and more fully examine the demographic history of bighorn sheep (Bolormaa *et al.* 2014). Expanded sequencing efforts would also allow for comparative genomics to further identify ancestral states and regions of selection relative to domestic sheep. In addition, our bighorn sheep genome can aid reference guided genome assembly (Gnerre *et al.* 2009; Kim *et al.* 2013) of other *Ovis* species using a genome that has not been subject to strong selective breeding.

6.5 Conclusion

In this study, we created a WGS for bighorn sheep using the closely related domestic sheep as a reference for alignment. This procedure was highly successful, covering 95% of the reference with an average read depth of 12 (104 SD). From this sequence we were able to call 15,622,848 variants and found 40 GO terms with significantly different representation in fixed SNPs tagged as non-synonymous versus synonymous. We hypothesize that these differences may largely be a result of selection during domestication. Our results demonstrate that cross-species alignment enables the creation of novel WGS for non-model organisms. The bighorn sheep WGS will provide a reference for future resequencing studies or comparative genomics.

Table 6 - 1 Number of loci showing concordance or discordance between the genome and the Ovine Infinium®HD SNP BeadChip

	Original Filter	Stringent Filter
Same Genotype	377129	314734
Heterozygous on Chip, Homozygous in Sequence	456	130
Homozygous on Chip, Heterozygous in Sequence	22837	9565
Alternate Homozygotes	7169	5261

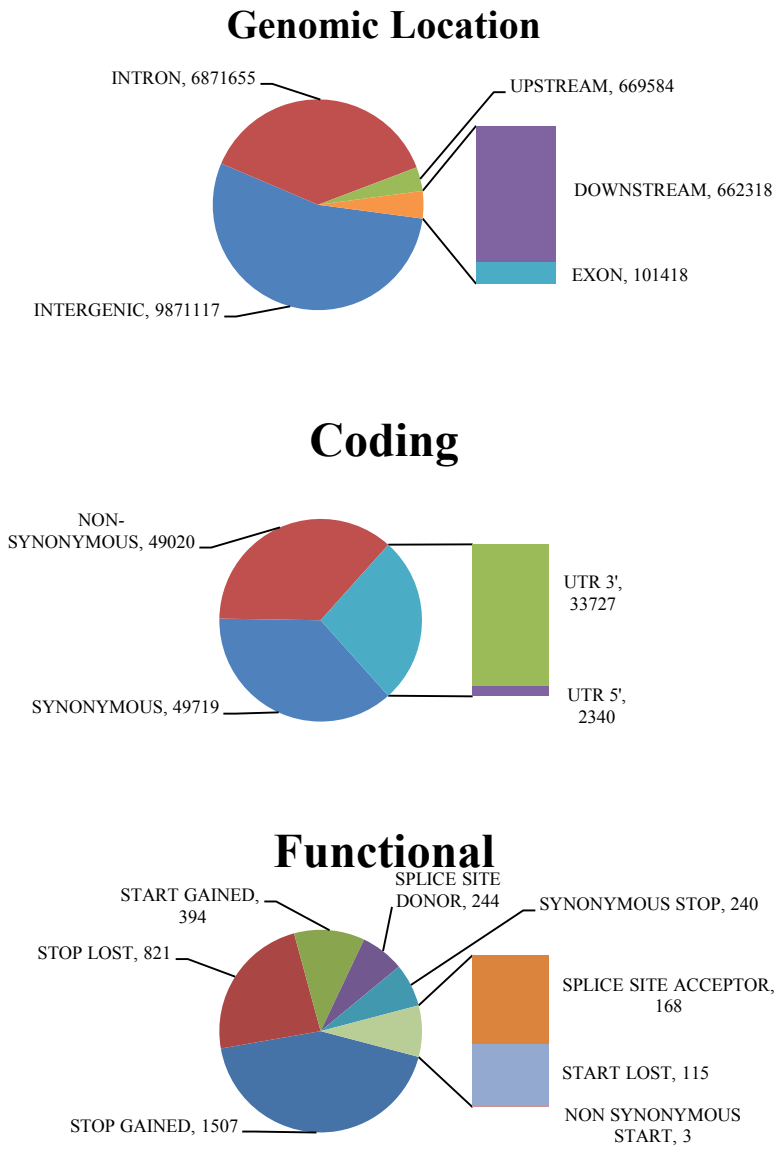


Figure 6 - 1 Distribution of SNP annotations and effect predictions
Numbers are the count of loci in each category

6.6 Bibliography

Angeloni F, Wagemaker N, Vergeer P, Ouborg J (2012) Genomic toolboxes for conservation biologists. *Evolutionary Applications* **5**, 130-143.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300.

Berger J (1990) Persistence of different-sized populations: An empirical assessment of rapid extinctions in bighorn sheep. *Conservation Biology* **4**, 91-98.

Blüthgen N, Brand K, Čajavec B*, et al.* (2005) Biological profiling of gene groups utilizing Gene Ontology. *Genome Informatics* **16**, 106-115.

Bolormaa S, Kijas J, Coltman DW, Daetwyler HD, MacLeod IM (2014) Inferring ancestral demography of domestic and wild sheep using whole-genome sequence *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*.

Bourque G, Zdobnov E, Bork P, Pevzner P, Tesler G (2005) Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Research* **15**, 98-110.

Braglia S, Davoli R, Zappavigna A*, et al.* (2013) SNPs of MYPN and TTN genes are associated to meat and carcass traits in Italian Large White and Italian Duroc pigs. *Molecular Biology Reports* **40**, 6927-6933.

Branton D, Deamer DW, Marziali A*, et al.* (2008) The potential and challenges of nanopore sequencing. *Nat Biotech* **28**, 1146-1153.

Buerkle C, Gompert Z, Parchman T (2011) The n=1 constraint in population genomics. *Molecular Ecology* **20**, 1575-1581.

Bunch T, Wu C, Zhang Y, Wang S (2006) Phylogenetic analysis of snow sheep (Ovis nivicola) and closely related taxa. *Journal of Heredity* **97**, 21-30.

Canavez FC, Luche DD, Stothard P*, et al.* (2012) Genome sequence and assembly of *Bos indicus*. *Journal of Heredity* **103**, 342-348.

Cerdeira LT, Carneiro AR, Ramos RTJ*, et al.* (2011) Rapid hybrid de novo assembly of a microbial genome using only short reads: Corynebacterium pseudotuberculosis I19 as a case study. *Journal of Microbiological Methods* **86**, 218-223.

Cingolani P, Patel VM, Coon M*, et al.* (2012a) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in Genetics* **3**.

Cingolani P, Platts A, Wang L*, et al.* (2012b) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly* **6**, 80-92.

Clop A, Marcq F, Takeda H*, et al.* (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* **38**, 813-818.

Coltman DW, Festa-Bianchet M, Jorgenson JT, Strobeck C (2002) Age-dependent sexual selection in bighorn rams. *Proceedings of the Royal Society B-Biological Sciences* **269**, 165-172.

Coltman DW, O'Donoghue P, Jorgenson JT*, et al.* (2003) Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655-658.

Conesa A, Götz S, García-Gómez JM*, et al.* (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676.

Daetwyler HD, Capitan A, Pausch H*, et al.* (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet* **46**, 858-865.

Dalloul RA, Long JA, Zimin AV*, et al.* (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *Plos Biology* **8**.

Danecek P, Auton A, Abecasis G*, et al.* (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

Davis EB, Brakora KA, Lee AH (2011) Evolution of ruminant headgear: a review. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2857-2865.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1-15.

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications* **7**, 1026–1042.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29**, 51–63.

Fang X, Zhang Y, Zhang R*, et al.* (2011) Genome sequence and global sequence variation map with 5.5 million SNPs in Chinese rhesus macaque. *Genome Biology* **12**, doi:10.1186/gb-2011-1112-1187-r1163.

Festa-Bianchet M, Pelletier F, Jorgenson JT, Feder C, Hubbs A (2014) Decrease in horn size and increase in age of trophy sheep in Alberta over 37 years. *The Journal of Wildlife Management* **78**, 133–141.

Flicek P, Amode M, Barrell D*, et al.* (2014) Ensembl 2014. *Nucleic Acids Research* **42**, D749-D755.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *TREE* **27**, 489–496.

Gaillard JM, Festa-Bianchet M, Delorme D, Jorgenson JT (2000) Body mass and individual fitness in female ungulates: bigger is not always better. *Proceedings of the Royal Society B: Biological Sciences* **267**, 471-477.

Genomic Resources Dev Consortium, Bensch S, Coltman D*, et al.* (2014) Genomic Resources Notes accepted 1 June 2013-31 July 2013. *Molecular Ecology Resources* **14**, 218-218.

Genomic Resources Development Consortium, Coltman DW, Hogg JT, Miller JM (2013) Genomic Resources Notes accepted 1 April 2013–31 May 2013. *Molecular Ecology Resources* **13**, 965-965.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**, 759-769.

Gnerre S, Lander E, Lindblad-Toh K, Jaffe D (2009) Assisted assembly: how to improve a de novo genome assembly by using related species. *Genome Biology* **10**.

Hedrick PW (2014) Conservation genetics and the persistence and translocation of small populations: bighorn sheep populations as examples. *Animal Conservation* **17**, 106–114.

Hunt M, Newbold C, Berriman M, Otto T (2014) A comprehensive evaluation of assembly scaffolding tools. *Genome Biology* **15**, R42.

Jiang Y, Xie M, Chen W*, et al.* (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1168-1173.

Johnson HE, Mills LS, Wehausen JD, Stephenson TR, Luikart G (2011) Translating effects of inbreeding depression on component vital rates to overall population growth in endangered bighorn sheep. *Conservation Biology* **25**, 1240-1249.

Kijas JW, Lenstra JA, Hayes B*, et al.* (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol* **10**, e1001258.

Kijas JW, Porto-Neto L, Dominik S*, et al.* (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* **45**, 754–757.

Kim J, Larkin D, Cai Q*, et al.* (2013) Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1785-1790.

Kinsella R, Kahari A, Haider S*, et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database-the Journal of Biological Databases and Curation*.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* **21**, 629-637.

Lee K-T, Chung W-H, Lee S-Y*, et al.* (2013) Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. *BMC Genomics* **519**.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496.

Li H, Handsaker B, Wysoker A*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

Lv F-H, Agha S, Kantanen J*, et al.* (2014) Adaptations to Climate-Mediated Selective Pressures in Sheep. *Molecular Biology and Evolution* **31**, 3324-3343.

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nat Rev Genet* **12**, 671-682.

McKernan KJ, Peckham HE, Costa GL*, et al.* (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Research* **19**, 1527-1541.

Metzker ML (2010) Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46.

Miller J, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. *Genomics* **95**, 315-327.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012a) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* **12**, 1145-1150.

Miller JM, Malenfant RM, David P*, et al.* (2014) Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity* **112**, 240–247.

Miller JM, Malenfant RM, Moore SS, Coltman DW (2012b) Short reads, circular genome: Skimming SOLiD sequence to construct the bighorn sheep mitochondrial genome. *Journal of Heredity* **103**, 140-146.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012c) Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Molecular Ecology* **21**, 1583–1596.

Miller JM, Poissant J, Kijas J, Coltman DW, TISGC (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* **11**, 314-322.

Olson ZH, Whittaker DG, Rhodes OE (2013) Translocation history and genetic diversity in reintroduced bighorn sheep. *The Journal of Wildlife Management* **77**, 1553–1563.

Ondov B, Varadarajan A, Passalacqua K, Bergman N (2008) Efficient mapping of Applied Biosystems SOLiD sequence data to a reference genome for functional genomic applications. *Bioinformatics* **24**, 2776-2777.

Pelletier F, Festa-Bianchet M (2006) Sexual selection and social rank in bighorn rams. *Animal Behaviour* **71**, 649-655.

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2012) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity* **108**, 256–263.

Poissant J, Hogg JT, Davis CS*, et al.* (2010) Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. *BMC Genomics* **11**, doi:10.1186/1471-2164-1111-1524.

Poissant J, Shafer ABA, Davis CS*, et al.* (2009) Genome-wide cross-amplification of domestic sheep microsatellites in bighorn sheep and mountain goats. *Molecular Ecology Resources* **9**, 1121-1126.

Prado-Martinez J, Sudmant PH, Kidd JM*, et al.* (2013) Great ape genetic diversity and population history. *Nature* **499**, 471-475.

Réale D, Martin J, Coltman DW, Poissant J, Festa-Bianchet M (2009) Male personality, life-history strategies and reproductive success in a promiscuous mammal. *Journal of Evolutionary Biology* **22**, 1599-1607.

Sambrook J, Russell D (2001) *Molecular cloning: a laboratory manual*, Cold Spring Harbor (NY).

Sasaki Y, Nagai K, Nagata Y*, et al.* (2006) Exploration of genes showing intramuscular fat deposition-associated expression changes in musculus longissimus muscle. *Animal Genetics* **37**, 40-46.

Sechi T, Coltman DW, Kijas JW (2010) Evaluation of 16 loci to examine the cross-species utility of single nucleotide polymorphism arrays. *Animal Genetics* **41**, 199-202.

Shackleton DM, Shank CC, Wikeem B (1999) Natural history of Rock Mountain and California bighorn sheep. In: *Mountain Sheep of North America* (eds. Valdez R, Krausman PR), pp. 78-138. The University of Arizona Press, Tuscon.

Sheehan S, Harris K, Song YS (2013) Estimating Variable Effective Population Sizes from Multiple Genomes: A Sequentially Markov Conditional Sampling Distribution Approach. *Genetics* **194**, 647-662.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158-170.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *Plos One* **6**.

Telford MJ, Copley RR (2011) Improving animal phylogenies with genomic data. *Trends in Genetics* **27**, 186-195.

Umemura M, Koyama Y, Takeda I*, et al.* (2013) Fine De Novo Sequencing of a Fungal Genome Using only SOLiD Short Read Data: Verification on Aspergillus oryzae RIB40. *Plos One* **8**.

Vijay N, Poelstra JW, Künstner A, Wolf JBW (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Molecular Ecology* **22**, 620-634.

Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in Ecology & Evolution* **11**, 158-162.

Wang B, Ekblom R, Bunikis I, Siitari H, Hoglund J (2014) Whole genome sequencing of the black grouse (Tetrao tetrix): reference guided assembly suggests faster-Z and MHC evolution. *BMC Genomics* **15**, 180.

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet* **13**, 329-342.

Zhao H, Bourque G (2009) Recovering genome rearrangements in the mammalian phylogeny. *Genome Research* **19**, 934-942.

Supplementary Table 6-S1: Summary of predicted effects of each SNP by chromosome as assigned by SNPEff.

Supplementary Table 6-S2: GO enrichment summary between loci predicted to be non-synonymous and those predicted to be synonymous.

Supplementary Figure 6-S1: Histogram of insertion/deletion lengths in the bighorn draft genome relative to the domestic sheep reference.

Chapter 7

GENERAL CONCLUSION

7.1 General Conclusion

During the course of my PhD I developed a variety of genomic resources for bighorn sheep (chapters 2 and 6). Using those resources I then addressed several questions regarding a common analysis method for estimating inbreeding in the absence of a pedigree (chapters 3 and 4) and the genetic basis of fitness characteristic in bighorn sheep (chapter 5).

In **chapter 2**, I used two parallel methodologies to discover single nucleotide polymorphisms (SNPs) in bighorn sheep. Through cross-species application of the Ovine SNP50 BeadChip I found that over 90% of the markers could be genotyped, indicating that there is no intrinsic barrier to cross-species application of the technology. However, only ~900 of the ~49,000 loci on the chip were found to be polymorphic across two populations of bighorn sheep and one population of thinhorn sheep. In tandem with this analysis, restriction-site associated DNA (RAD) sequencing of eight bighorn sheep from two populations returned over ~15,000 bighorn specific loci. These loci can serve as a resource for future studies.

In **chapter 3**, I examined the number of genetic markers that would be needed to reflect genome-wide heterozygosity in two populations of bighorn sheep. I found that in both populations, individual heterozygosity was significantly correlated between SNPs and microsatellite loci, although the strength of the correlation was weaker in a native population compared with one founded via translocation and later supplemented with additional individuals. I also noted that despite being bi-allelic, SNPs had similar correlations to genome-wide

heterozygosity as microsatellites in both populations. For both marker types, this association became stronger and less variable as more markers were considered. Both populations had significant levels of identity disequilibrium (ID, a proxy for inbreeding); however, estimates were an order of magnitude lower in the native population. As with estimates of heterozygosity, SNPs performed similarly to microsatellites when subsets of loci were used to estimate ID, and precision and accuracy of the estimates increased as more loci were considered. Together these results illustrate that genome-wide heterozygosity, and therefore heterozygosity fitness correlations (HFCs), are best measured by a large number of markers, a feat now more realistically accomplished with SNPs than microsatellites.

In **chapter 4**, I built on the results of chapter 3 by conducting a meta-analysis examining previous HFC studies to see if the level of ID (as measured by the g_2 statistic) in a population was correlated with the strength of the HFC reported. I was able to collect estimates of ID from 50 studies and found that in the majority of studies g_2 values were not significantly different than zero. Despite this, I found that the magnitude of g_2 was associated with the average effect size observed in a population, even when point estimates were non-significant. The low values of g_2 translated into low expected correlations between heterozygosity and inbreeding, and suggest that many more markers than typically used are needed to robustly detect HFCs.

In **chapter 5**, I conducted a genome wide association analysis to search for potential links between SNP variants and fitness related characteristics (three morphological and five life

history traits) in the Ram Mountain population of sheep. All three of the morphological traits and four of the life history traits were associated with one or more SNP loci, 11 loci in total. I then expanded genotyping of these candidate loci to additional individuals from the same population and found that the associations held for the morphological traits, but broke down for the life history ones. Examination of temporal trends in allele frequency at the morphological loci showed one locus (associated with average horn base circumference) to have nominally significant changes in allele frequency over time. Notably, the allele increasing in frequency was associated with reduced breeding values of horn base circumference. Expanding genotyping, using additional loci, individuals, as well as populations, could examine the consistency of these results.

In **chapter 6**, I constructed a draft whole genome sequence (WGS) from a single bighorn sheep via alignment to a domestic sheep genome as a reference. Using over 865 million reads generated from two libraries sequenced on ABI SOLiD platforms I generated a nearly complete WGS (95% coverage of the reference) at an average of 12x read depth (104 SD). From this alignment I discovered over 15 million variants and annotated the SNP loci relative to the domestic sheep reference. I then conducted an enrichment analysis of those SNPs showing fixed differences between the two species and found significant differences in a number of gene ontology terms. My results demonstrate that cross-species alignment enables the creation of

novel WGS for non-model organisms. The bighorn sheep WGS will provide a reference for future resequencing studies or comparative genomics.

The field of genomics has changed dramatically even in the short time since I started my degree. These changes have been fueled by the breathtaking pace of technological advancement (De Wit et al. 2012; Glenn 2011; Helyar et al. 2011; Morozova et al. 2009), which allow for research questions once thought intractable even in the best model systems to be addressed in wild species (Ekblom & Wolf 2014; Ellegren 2014). Through my work I have endeavored to move bighorn sheep into the genomic arena, developing several genome-wide SNP sets as well as a draft whole genome sequence.

However, this move has not been without challenges. While cross-species use of SNP chips represents a novel and cost-effective way to generate many high quality SNPs genotypes in a large sample of individuals the relatively small number of loci remaining polymorphic resulted in a lack of power to detect associations in a GWAS. Moving forward, development of a bighorn specific SNP chip seems warranted to more fully explore the genetic basis for fitness related traits in bighorn sheep. Such a SNP chip could be built by combing loci I characterized or discovered in other parts of my thesis, including those from the domestic sheep SNP chips, RAD sequencing, and heterozygous sites in the bighorn sheep draft genome. If such a chip contained 20,000 or more loci it would likely have sufficient marker density and spacing to overcome some of the power issues I encountered in chapter 5.

More broadly, as the costs of genomic methods continue to decline and new techniques are developed, additional research questions can build on the work I present here. For example: 1) transcriptome data could be used to search for functional variants underlying traits (Wolf 2013). 2) Adaptive variation could be identified through additional reduced representation sequencing of other populations or sub-species (Funk *et al.* 2012). 3) Individual genome sequencing may detect rare variants that have fitness consequences (Daetwyler et al. 2014). 4) Chromosome partitioning methods would allow investigation to see if complex traits can be attributed to polygenic gene actions (Robinson et al. 2013; Santure et al. 2013; Yang et al. 2011). 5) Runs of homozygosity analysis to examine the demographic history of bighorn sheep populations (Li & Durbin 2011). These methods represent exciting prospects to discover more about this charismatic species.

7.2 Bibliography

Daetwyler HD, Capitan A, Pausch H*, et al.* (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet* **46**, 858-865.

De Wit P, Pespeni MH, Ladner JT*, et al.* (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* **12**, 1058-1067.

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications* **7**, 1026–1042.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29**, 51–63.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *TREE* **27**, 489–496.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**, 759-769.

Helyar S, Hemmer-Hansen J, Bekkevold D*, et al.* (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**, 123-136.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496.

Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics* **10**, 135-151.

Robinson MR, Santure AW, DeCauwer I, Sheldon BC, Slate J (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology* **22**, 3963-3980.

Santure A, De Cauwer I, Robinson M*, et al.* (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. *Molecular Ecology* **22**, 3949-3962.

Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources* **13**, 559–572.

Yang J, Manolio T, Pasquale L*, et al.* (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519-U544.

BIBLIOGRAPHY

Allais-Bonnet A, Grohs C, Medugorac I*, et al.* (2013) Novel Insights into the Bovine Polled Phenotype and Horn Ontogenesis in Bovidae. *PLoS ONE* **8**.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Review Genetics* 11, 697-709.

Amos W, Wilmer J, Fullard K, et al. (2001) The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society B-Biological Sciences* 268, 2021-2027.

Andersson L (2009) Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* 136, 341-349.

Angeloni F, Wagemaker N, Vergeer P, Ouborg J (2012) Genomic toolboxes for conservation biologists. *Evolutionary Applications* 5, 130-143.

Aparicio JM, Ortego J, Cordero PJ (2008) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology* 15, 4659-4665.

Ashley MV, Willson MF, Pergams ORW*, et al.* (2003) Evolutionarily enlightened management. *Biological Conservation* **111**, 115-123.

Backström N, Qvarnström A, Gustafsson L, Ellegren H (2006) Levels of linkage disequilibrium in a wild bird population. *Biology Letters* 2, 435-438.

Baird NA, Etter PD, Atwood TS, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3, e3376.

Balloux F, Amos W, Coulson T (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* 13, 3021-3031.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265.

Bartoń K (2009) MuMIn: multi-model inference, Available at: http://r-forge.r-project.org/projects/mumin/.

Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-6. http://CRAN.R-project.org/package=lme4.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289-300.

Bensch S, Andrén H, Hansson B, et al. (2006) Selection for heterozygosity gives hope to a wild population of inbred wolves. *PLoS ONE* 1, e72.

Berger J (1990) Persistence of different-sized populations: An empirical assessment of rapid extinctions in bighorn sheep. *Conservation Biology* 4, 91-98.

Bierne N, Tsitrone A, David P (2000) An inbreeding model of associative overdominance during a population bottleneck. *Genetics* 155, 1981-1990.

Blüthgen N, Brand K, Čajavec B, et al. (2005) Biological profiling of gene groups utilizing Gene Ontology. *Genome Informatics* 16, 106-115.

Bohonak AJ, van der Linde K (2004) RMA: Software for Reduced Major Axis regression, Java version. Website: http://www.kimvdlinde.com/professional/rma.html.

Bolormaa S, Kijas J, Coltman DW, Daetwyler HD, MacLeod IM (2014) Inferring ancestral demography of domestic and wild sheep using whole-genome sequence *Proceedings, 10th World Congress of Genetics Applied to Livestock Production*.

Bonenfant C, Pelletier F, Garel M, Bergeron P (2009) Age-dependent relationship between horn growth and survival in wild sheep. *Journal of Animal Ecology* **78**, 161-171.

Borrell Y, Carleos C, Sanchez J, et al. (2011) Heterozygosity-fitness correlations in the gilthead sea bream *Sparus aurata* using microsatellite loci from unknown and gene-rich genomic locations. *Journal of Fish Biology* 79, 1111-1129.

Bourque G, Zdobnov E, Bork P, Pevzner P, Tesler G (2005) Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Research* 15, 98-110.

Braglia S, Davoli R, Zappavigna A, et al. (2013) SNPs of MYPN and TTN genes are associated to meat and carcass traits in Italian Large White and Italian Duroc pigs. *Molecular Biology Reports* 40, 6927-6933.

Brandstrom M, Ellegren H (2008) Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome Research* 18, 881-887.

Branton D, Deamer DW, Marziali A, et al. (2008) The potential and challenges of nanopore sequencing. *Nature Biotechnology* 28, 1146-1153.

Britten H (1996) Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* 50, 2158-2164.

Bro-Jørgensen J (2007) The intensity of sexual selection predicts weapon size in male bovids. *Evolution* **61**.

Browning SR, Browning BL (2011) Haplotype phasing: existing methods and new developments. *Nature Reviews Genetics* **12**, 703-714.

Buerkle C, Gompert Z, Parchman T (2011) The n=1 constraint in population genomics. *Molecular Ecology* 20, 1575-1581.

Bunch T, Wu C, Zhang Y, Wang S (2006) Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. *Journal of Heredity* 97, 21-30.

Bunch TD, Hoffmann RS, Nadler CF (1999) Cytogenetics and genetics. In: *Mountain Sheep of North America* (eds. Valdez R, Krausman PR). The University of Arizona Press, Tuscon.

Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach* Springer Science & Business Media.

Canavez FC, Luche DD, Stothard P, et al. (2012) Genome sequence and assembly of *Bos indicus*. *Journal of Heredity* 103, 342-348.

Carlson CS, Eberle MA, Rieder MJ, et al. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *The American Journal of Human Genetics* 74, 106-120.

Caspers BA, Krause ET, Hendrix R, et al. (2014) The more the better – polyandry and genetic similarity are positively linked to reproductive success in a natural population of terrestrial salamanders (*Salamandra salamandra*). *Molecular Ecology* 23, 239–250.

Cerdeira LT, Carneiro AR, Ramos RTJ, et al. (2011) Rapid hybrid de novo assembly of a microbial genome using only short reads: *Corynebacterium pseudotuberculosis* I19 as a case study. *Journal of Microbiological Methods* 86, 218-223.

Chakraborty R (1981) The distribution of the number of heterozygous loci in an individual in natural-populations. *Genetics* 98, 461-466.

Chang C, Chow C, Tellier L*, et al.* (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**.

Chapman JR, Nakagawa S, Coltman DW, Slate J, Sheldon BC (2009) A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology* 18, 2746-2765.

Charpentier M, Boulet M, Drea C (2008a) Smelling right: the scent of male lemurs advertises genetic quality and relatedness. *Molecular Ecology* 17, 3225-3233.

Charpentier M, Setchell J, Prugnolle F, et al. (2005) Genetic diversity and reproductive success in mandrills (*Mandrillus sphinx*). *Proceedings of the National Academy of Sciences of the United States of America* 102, 16723-16728.

Charpentier M, Tung J, Altmann J, Alberts S (2008b) Age at maturity in wild baboons: genetic, environmental and demographic influences. *Molecular Ecology* 17, 2026-2040.

Charpentier M, Williams C, Drea C (2008c) Inbreeding depression in ring-tailed lemurs (*Lemur catta*): genetic diversity predicts parasitism, immunocompetence, and survivorship. *Conservation Genetics* 9, 1605-1615.

Chenoweth SF, McGuigan K (2010) The genetic basis of sexually selected variation. *Annual Review of Ecology, Evolution, and Systematics* **41**, 81-101.

Cingolani P, Patel VM, Coon M, et al. (2012a) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in Genetics* 3.

Cingolani P, Platts A, Wang L, et al. (2012b) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. *Fly* 6, 80-92.

Clop A, Marcq F, Takeda H, et al. (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics* 38, 813-818.

Coates BS, Sumerford DV, Miller NJ, et al. (2009) Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *Journal of Heredity* 100, 556-564.

Collins FS, Patrinos A, Jordan E, et al. (1998) New Goals for the U.S. Human Genome Project: 1998-2003. *Science* 282, 682-689.

Coltman D, Bowen W, Wright J (1998) Birth weight and neonatal survival of harbour seal pups ape positively correlated with genetic variation measured by microsatellites. *Proceedings of the Royal Society B-Biological Sciences* 265, 803-809.

Coltman D, Pilkington J, Smith J, Pemberton J (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. *Evolution* 53, 1259-1267.

Coltman D, Slate J (2003) Microsatellite measures of inbreeding: A meta-analysis. *Evolution* 57, 971-983.

Coltman DW (2005) Testing marker-based estimates of heritability in the wild. *Molecular Ecology* **14**, 2593-2599.

Coltman DW (2008) Molecular ecological approaches to studying the evolutionary impact of selective harvesting in wildlife. *Molecular Ecology* **17**, 221-235.

Coltman DW, Festa-Bianchet M, Jorgenson JT, Strobeck C (2002) Age-dependent sexual selection in bighorn rams. *Proceedings of the Royal Society B-Biological Sciences* 269, 165-172.

Coltman DW, O'Donoghue P, Hogg JT, Festa-Bianchet M (2005) Selection and genetic (CO)variance in bighorn sheep. *Evolution* **59**, 1372-1382.

Coltman DW, O'Donoghue P, Jorgenson JT, et al. (2003) Undesirable evolutionary consequences of trophy hunting. *Nature* 426, 655-658.

Conesa A, Götz S, García-Gómez JM, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.

Cote S, Stien A, Irvine R, et al. (2005) Resistance to abomasal nematodes and individual genetic variability in reindeer. *Molecular Ecology* 14, 4159-4168.

Coulson T, Albon S, Slate J, Pemberton J (1999) Microsatellite loci reveal sex-dependent responses to inbreeding and outbreeding in red deer calves. *Evolution* 53, 1951-1960.

Coulson T, Pemberton J, Albon S, et al. (1998) Microsatellites reveal heterosis in red deer. *Proceedings of the Royal Society B-Biological Sciences* 265, 489-495.

Crestanello B, Pecchioli E, Vernesi C, et al. (2009) The genetic impact of translocations and habitat fragmentation in chamois (Rupicapra) spp. *Journal of Heredity* 100, 691-708.

Curik I, Zechner P, Solkner J, et al. (2003) Inbreeding, microsatellite heterozygosity, and morphological traits in Lipizzan horses. *Journal of Heredity* 94, 125-132.

Da Silva A, Luikart G, Yoccoz N, Cohas A, Allaine D (2006) Genetic diversity-fitness correlation revealed by microsatellite analyses in European alpine marmots (*Marmota marmota*). *Conservation Genetics* 7, 371-382.

Daetwyler HD, Capitan A, Pausch H, et al. (2014) Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics* 46, 858-865.

Dalloul RA, Long JA, Zimin AV, et al. (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biology* 8.

Dalrymple B, Kirkness E, Nefedov M, et al. (2007) Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biology* 8, R152.

Danecek P, Auton A, Abecasis G, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27, 2156-2158.

Davey JW, Hohenlohe PA, Etter PD, et al. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12, 499-510.

David P (1998) Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* 80, 531-537.

David P, Pujol B, Viard F, Castella V, Goudet J (2007) Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology* 16, 2474-2487.

Davis EB, Brakora KA, Lee AH (2011) Evolution of ruminant headgear: a review. *Proceedings of the Royal Society B: Biological Sciences* 278, 2857-2865.

De La Vega FM, Isaac H, Collins A, et al. (2005) The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Research* 15, 454-462.

De Wit P, Pespeni MH, Ladner JT, et al. (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular Ecology Resources* 12, 1058-1067.

DeWoody Y, DeWoody J (2005) On the estimation of genome-wide heterozygosity using molecular markers. *Journal of Heredity* 96, 85-88.

DiBattista JD, Feldheim KA, Gruber SH, Hendry AP (2008) Are indirect genetic benefits associated with polyandry? Testing predictions in a natural population of lemon sharks. *Molecular Ecology* 17, 783-795.

Dumont BL, Payseur BA (2008) Evolution of the genomic rate of recombination in mammals. *Evolution* **62**, 276-294.

Eberle MA, Rieder MJ, Kruglyak L, Nickerson DA (2006) Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome. *PLoS Genetics* 2, e142.

Egger M, Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315, 629-634.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107, 1-15.

Ekblom R, Wolf JBW (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications* 7, 1026–1042.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* 29, 51–63.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature* **452**, 169-175.

Elshire R, Glaubitz J, Sun Q, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos ONE* 6.

Fang X, Zhang Y, Zhang R, et al. (2011) Genome sequence and global sequence variation map with 5.5 million SNPs in Chinese rhesus macaque. *Genome Biology* 12, doi:10.1186/gb-2011-1112-1187-r1163.

Favre M, Martin JG, Festa-Bianchet M (2008) Determinants and life-history consequences of social dominance in bighorn ewes. *Animal Behaviour* **76**, 1373-1380.

Feltus FA, Wan J, Schulze SR, et al. (2004) An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments. *Genome Research* 14, 1812-1819.

Feng S, Ferlin A, Truong A*, et al.* (2009) INSL3/RXFP2 signaling in testicular descent. *Annals of the New York Academy of Sciences* **1160**, 197-204.

Ferlin A, Pepe A, Gianesello L*, et al.* (2008) Mutations in the insulin-like factor 3 receptor are associated with osteoporosis. *Journal of Bone and Mineral Research* **23**, 683-693.

Festa-Bianchet M, Coltman DW, Hogg JT, Jorgenson JT (2008) Age-related horn growth, mating tactics, and vulnerability to harvest: why horn curl limits may select for small horns in bighorn sheep. *Biennial Symposium of the Northern Wild Sheep and Goat Council* 15, 42-49.

Festa-Bianchet M, Coulson T, Gaillard JM, Hogg JT, Pelletier F (2006) Stochastic predation events and population persistence in bighorn sheep. *Proceedings of the Royal Society B-Biological Sciences* 273, 1537-1543.

Festa-Bianchet M, Jorgenson J, King W, Smith K, Wishart W (1996) The development of sexual dimorphism: Seasonal and lifetime mass changes in bighorn sheep. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **74**, 330-342.

Festa-Bianchet M, Pelletier F, Jorgenson JT, Feder C, Hubbs A (2014) Decrease in horn size and increase in age of trophy sheep in Alberta over 37 years. *The Journal of Wildlife Management* 78, 133–141.

Flicek P, Amode M, Barrell D, et al. (2014) Ensembl 2014. *Nucleic Acids Research* 42, D749-D755.

Forstmeier W, Schielzeth H (2011) Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology* 65, 47-55.

Forstmeier W, Schielzeth H, Mueller J, Ellegren H, Kempenaers B (2012) Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Molecular Ecology* 21, 3237-3249.

Fossøy F, Johnsen A, Lifjeld J (2008) Multiple genetic benefits of female promiscuity in a socially monogamous passerine. *Evolution* 62, 145-156.

Frazer KA, Eskin E, Kang HM, et al. (2007) A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 448, 1050-1053.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* 27, 489–496.

Gage M, Surridge A, Tomkins J, et al. (2006) Reduced heterozygosity depresses sperm quality in wild rabbits, *Oryctolagus cuniculus*. *Current Biology* 16, 612-617.

Gaillard JM, Festa-Bianchet M, Delorme D, Jorgenson JT (2000) Body mass and individual fitness in female ungulates: bigger is not always better. *Proceedings of the Royal Society B: Biological Sciences* 267, 471-477.

García-Gámez E, Sahana G, Gutiérrez- Gil B, Arranz J-J (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* **13**.

Gautier M, Naves M (2011) Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology* **20**, 3128-3143.

Genomic Resources Dev Consortium, Bensch S, Coltman D, et al. (2014) Genomic Resources Notes accepted 1 June 2013-31 July 2013. *Molecular Ecology Resources* 14, 218-218.

Genomic Resources Development Consortium, Coltman DW, Hogg JT, Miller JM (2013) Genomic Resources Notes accepted 1 April 2013–31 May 2013. *Molecular Ecology Resources* 13, 965-965.

Gilmour A, Gogel B, Cullis B, Thompson R (2009) *ASReml User Guide. Release 3.0.* VSN International Ltd, Hemel Hempstead, UK.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11, 759-769.

Gnerre S, Lander E, Lindblad-Toh K, Jaffe D (2009) Assisted assembly: how to improve a *de novo* genome assembly by using related species. *Genome Biology* 10.

Gratten J, Pilkington JG, Brown EA*, et al.* (2012) Selection and microevolution of coat pattern are cryptic in a wild population of sheep. *Molecular Ecology* **21**, 2977–2990.

Gray MM, Granka J, Bustamante CD, et al. (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics* 181, 1493-1505.

Gregory TR, Nicol JA, Tamm H, et al. (2007) Eukaryotic genome size databases. *Nucleic Acids Research* 35, D332-338.

Gregory, TR (2010). Animal Genome Size Database. http://www.genomesize.com

Grueber C, Nakagawa S, Laws R, Jamieson I (2011a) Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology* 24, 699-711.

Grueber C, Waters J, Jamieson I (2011b) The imprecision of heterozygosity-fitness correlations hinders the detection of inbreeding and inbreeding depression in a threatened species. *Molecular Ecology* 20, 67-79.

Grueber CE, Wallis GP, Jamieson IG (2008) Heterozygosity–fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology* 17, 3978-3984.

Guichoux E, Lagache L, Wagner S, et al. (2011) Current trends in microsatellite genotyping. *Molecular Ecology Resources* 11, 591-611.

Gunderson KL (2009) Whole-genome genotyping on bead arrays. In: *DNA Microarrays for Biomedical Research. Methods and Protocols* (ed. Dufva M), pp. 197-213. Humana Press, a part of Springer Science+Business Media, LLC.

Hadfield JD, Wilson AJ, Garant D, Sheldon BC, Kruuk LEB (2010) The misuse of BLUP in ecology and evolution. *American Naturalist* **175**, 116-125.

Hammerly SC, Morrow ME, Johnson JA (2013) A comparison of pedigree- and DNA-based measures for identifying inbreeding depression in the critically endangered Attwater's Prairie-chicken. *Molecular Ecology* 22, 5313–5328.

Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* 11, 2467-2474.

Harmegnies N, Farnir F, Davin F, et al. (2006) Measuring the extent of linkage disequilibrium in commercial pig populations. *Animal Genetics* 37, 225-231.

Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genetics* 9, e1003521.

Harrison X, Bearhop S, Inger R, et al. (2011) Heterozygosity-fitness correlations in a migratory bird: an analysis of inbreeding and single-locus effects. *Molecular Ecology* 20, 4786-4795.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* **7**, 1008–1025.

Haussler D, O'Brien SJ, Ryder OA, et al. (2009) Genome 10K: A proposal to obtain whole-genome sequence for 10 000 vertebrate species. *Journal of Heredity* 100, 659-674.

Heath D, Bryden C, Shrimpton J, et al. (2002) Relationships between heterozygosity, allelic distance (d(2)), and reproductive traits in chinook salmon, *Oncorhynchus tshawytscha*. *Canadian Journal of Fisheries and Aquatic Sciences* 59, 77-84.

Hedrick P, Coltman D, Festa-Bianchet M, Pelletier F (2014) Not surprisingly, no inheritance of a trait results in no evolution. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4810-E4810.

Hedrick PW (2014) Conservation genetics and the persistence and translocation of small populations: bighorn sheep populations as examples. *Animal Conservation* 17, 106–114.

Heifetz EM, Fulton JE, O'Sullivan N, et al. (2005) Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. *Genetics* 171, 1173-1181.

Helyar S, Hemmer-Hansen J, Bekkevold D, et al. (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* 11, 123-136.

Hengeveld PE, Festa-Bianchet M (2011) Harvest regulations and artificial selection on horn size in male bighorn sheep. *The Journal of Wildlife Management* 75, 189-197.

Herdegen M, Nadachowska-Brzyska K, Konowalik A, Babik W, Radwan J (2013) Heterozygosity, sexual ornament and body size in the crested newt. *Journal of Zoology* 291, 146-153.

Hicks J, Rachlow J (2006) Is there a genetic basis for antler and pedicle malformations in reintroduced elk in Northern Arizona? *Southwestern Naturalist* 51, 276-282.

Hill W, Weir B (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology* **33**, 54-78.

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6, 95-108.

Hoffman J, Forcada J, Amos W (2010a) Exploring the mechanisms underlying a heterozygosity-fitness correlation for canine size in the Antarctic fur seal *Arctocephalus gazella*. *Journal of Heredity* 101, 539-552.

Hoffman J, Hanson N, Forcada J, Trathan P, Amos W (2010b) Getting long in the tooth: A strong positive correlation between canine size and heterozygosity in Antarctic fur seals *Arctocephalus gazella. Journal of Heredity* 101, 527-538.

Hoffmann A, Griffin P, Dillon S*, et al.* (2015) A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* **2**, 1-24.

Hogg JT, Forbes SH (1997) Mating in bighorn sheep: frequent male reproduction via a high-risk "unconventional" tactic. *Behavioral Ecology and Sociobiology* 41, 33-48.

Hogg JT, Forbes SH, Steele BM, Luikart G (2006) Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B: Biological Sciences* 273, 1491-1499.

Hoglund J, Piertney S, Alatalo R, et al. (2002) Inbreeding depression and male fitness in black grouse. *Proceedings of the Royal Society B-Biological Sciences* 269, 711-715.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**, e1000862.

Hu Z-L, Fritz ER, Reecy JM (2007) AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Research.* **35**, D604-609.

Hu Z, Park C, Wu X, Reecy J (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Research* **41**, D871-D879.

Hunt M, Newbold C, Berriman M, Otto T (2014) A comprehensive evaluation of assembly scaffolding tools. *Genome Biology* 15, R42.

Iannuzzi L, Meo GP (1995) Chromosomal evolution in bovids: a comparison of cattle, sheep and goat G- and R-banded chromosomes and cytogenetic divergences among cattle, goat and river buffalo sex chromosomes. *Chromosome Research* 3, 291-299.

Jennions M, Moller A (2002) Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proceedings of the Royal Society B-Biological Sciences* 269, 43-48.

Jiang Y, Xie M, Chen W, et al. (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344, 1168-1173.

Johnson HE, Mills LS, Wehausen JD, Stephenson TR, Luikart G (2011) Translating effects of inbreeding depression on component vital rates to overall population growth in endangered bighorn sheep. *Conservation Biology* 25, 1240-1249.

Johnston S, McEwan J, Pickering N*, et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* **20**, 2555-2566.

Johnston SE, Beraldi D, McRae AF, Pemberton JM, Slate J (2010) Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity* **104**, 196-205.

Johnston SE, Gratten J, Berenos C*, et al.* (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93-95.

Jorgenson JT, Festa-Bianchet M, Gaillard J-M, Wishart WD (1997) Effects of age, sex, disease, and density on survival of bighorn sheep. *Ecology* 78, 1019-1032.

Jorgenson JT, Festa-Bianchet M, Lucherini M, Wishart WD (1993) Effects of body size, population density, and maternal characteristics on age at first reproduction in bighorn ewes. *Canadian Journal of Zoology* **71**, 2509-2517.

Jorgenson JT, Festa-Bianchet M, Wishart WD (1998) Effects of population density on horn development in bighorn rams. *The Journal of Wildlife Management* **62**, 1011-1020.

Jorgenson JTF-B, M.Wishart, W. D. (1993) Harvesting bighorn ewes: consequences for population size and trophy ram production. *The Journal of Wildlife Management* **57**, 429-435.

Jourdan-Pineau H, David P, Crochet P-A (2012) Phenotypic plasticity allows the Mediterranean parsley frog *Pelodytes punctatus* to exploit two temporal niches under continuous gene flow. *Molecular Ecology* 21, 876–886.

Jourdan-Pineau H, Folly J, Crochet P-A, David P (2012) Testing the influence of family structure and outbreeding depression on heterozygosity-fitness correlations in small populations. *Evolution* 66, 3624–3631.

Kardos M, Allendorf FW, Luikart G (2014) Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium? *Molecular Ecology Resources* 14, 519–530.

Karlsson EK, Baranowska I, Wade CM, et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics* 39, 1321-1328.

Kawakami T, Backström N, Burri R*, et al.* (2014) Estimation of linkage disequilibrium and interspecific gene flow in Ficedula flycatchers by a newly developed 50k SNP array. *Molecular Ecology Resources* **14**, 1248–1260.

Khatkar M, Nicholas F, Collins A, et al. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics* 9, 187.

Kijas JW, Lenstra JA, Hayes B*, et al.* (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* **10**, e1001258.

Kijas JW, Porto-Neto L, Dominik S, et al. (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics* 45, 754–757.

Kijas JW, Townley D, Dalrymple BP, et al. (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* 4, Article No.: e4668.

Kim J, Larkin D, Cai Q, et al. (2013) Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America* 110, 1785-1790.

Kinsella R, Kahari A, Haider S, et al. (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database-the Journal of Biological Databases and Curation*.

Klauke N, Segelbacher G, Schaefer H (2013) Reproductive success depends on the quality of helpers in the endangered, cooperative El Oro parakeet (*Pyrrhura orcesi*). *Molecular Ecology* 22, 2011-2027.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution* 21, 629-637.

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**, 873-890.

Kruuk LEB, Slate J, Wilson AJ (2008) New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annual Review of Ecology Evolution and Systematics* **39**, 525-548.

Küpper C, Kosztolányi A, Augustin J, et al. (2010) Heterozygosity-fitness correlations of conserved microsatellite markers in Kentish plovers *Charadrius alexandrinus*. *Molecular Ecology* 19, 5172–5185.

Laine V, Herczeg G, Shikano T, Primmer C (2012) Heterozygosity-behaviour correlations in nine-spined stickleback (*Pungitius pungitius*) populations: contrasting effects at random and functional loci. *Molecular Ecology* 21, 4872-4884.

Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10.

Laurie CC, Nickerson DA, Anderson AD, et al. (2007) Linkage disequilibrium in wild mice. *PLoS Genetics* 3, e144.

LeBas N (2002) Mate choice, genetic incompatibility, and outbreeding in the ornate dragon lizard, *Ctenophorus ornatus*. *Evolution* 56, 371-377.

Lee K-T, Chung W-H, Lee S-Y, et al. (2013) Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. *BMC Genomics* 519.

Lens L, Van Dongen S, Galbusera P, et al. (2000) Developmental instability and inbreeding in natural bird populations exposed to different levels of habitat disturbance. *Journal of Evolutionary Biology* 13, 889-896.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475, 493-496.

Li H, Handsaker B, Wysoker A, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Lieutenant-Gosselin M, Bernatchez L (2006) Local heterozygosity-fitness correlations with global positive effects on fitness in threespine stickleback. *Evolution* 60, 1658-1668.

Ljungqvist M, ÅKesson M, Hansson B (2010) Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al. (2008). *Molecular Ecology* 19, 851-855.

Loehr J, Carey J, Hoefs M, Suhonen J, & Ylönen H. (2007). Horn growth rate and longevity: implications for natural and artificial selection in thinhorn sheep (*Ovis dalli*). *Journal of Evolutionary Biology*, 20, 818-828.

Loehr J, Worley K, Grapputo A*, et al.* (2006) Evidence for cryptic glacial refugia from North American mountain sheep mitochondrial DNA. *Journal of Evolutionary Biology* **19**, 419-430.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4, 981-994.

Luquet E, Lena J, David P, et al. (2013) Within- and among-population impact of genetic erosion on adult fitness-related traits in the European tree frog *Hyla arborea*. *Heredity* 110, 347-354.

Lv F-H, Agha S, Kantanen J, et al. (2014) Adaptations to climate-mediated selective pressures in sheep. *Molecular Biology and Evolution* 31, 3324-3343.

Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-577.

Mainguy J, Côté SD, Cardinal E, Houle M (2008) Mating tactics and mate choice in relation to age and social rank in male mountain goats. *Journal of Mammalogy* 89, 626-635.

Manolio T, Collins F, Cox N*, et al.* (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747-753.

Mardis ER (2006) Anticipating the $1,000 genome. *Genome Biology* 7, 112.

Markert J, Grant P, Grant B, et al. (2004) Neutral locus heterozygosity, inbreeding, and survival in Darwin's ground finches (*Geospiza fortis and G-scandens*). *Heredity* 92, 306-315.

Marshall TC, Spalton JA (2000) Simultaneous inbreeding and outbreeding depression in reintroduced Arabian oryx. *Animal Conservation* 3, 241-248.

Martin AM, Presseault-Gauvin H, Festa-Bianchet M, Pelletier F (2013) Male mating competitiveness and age-dependent relationship between testosterone and social rank in bighorn sheep. *Behavioral Ecology and Sociobiology* **67**, 919-928.

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. *Nature Reviews Genetics* 12, 671-682.

McCarthy MI, Abecasis GR, Cardon LR, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* 9, 356-369.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* 66, 526–538.

McFarlane S, Gorrell J, Coltman D*, et al.* (2014) Very low levels of direct additive genetic variance in fitness and fitness components in a red squirrel population. *Ecology and Evolution* **4**, 1729-1738.

McFarlane SE, Gorrell JC, Coltman DW*, et al.* (2015) The nature of nurture in a wild mammal's fitness. *Proceedings of the Royal Society of London B: Biological Sciences* **282**.

McKay S, Schnabel R, Murdoch B*, et al.* (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* **8**.

McKernan KJ, Peckham HE, Costa GL, et al. (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Research* 19, 1527-1541.

McMahon BJ, Teeling EC, Höglund J (2014) How and why should we implement genomics into conservation? *Evolutionary Applications* 7, 999–1007.

McRae AF, McEwan JC, Dodds KG, et al. (2002) Linkage disequilibrium in domestic sheep. *Genetics* 160, 1113-1122.

Meadows J, Chan E, Kijas J (2008) Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* 9, 61.

Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics* 11, 31-46.

Miller J, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. *Genomics* 95, 315-327.

Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW (2012a) Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. *Molecular Ecology Resources* 12, 1145-1150.

Miller JM, Malenfant RM, David P, et al. (2014) Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity* 112, 240–247.

Miller JM, Malenfant RM, Moore SS, Coltman DW (2012b) Short reads, circular genome: Skimming solid sequence to construct the bighorn sheep mitochondrial genome. *Journal of Heredity* 103, 140-146.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012c) Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Molecular Ecology* 21, 1583–1596.

Miller JM, Poissant J, Kijas J, Coltman DW, TISGC (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources* 11, 314-322.

Minikel E (2012) Power for GWAS and extreme phenotype studies. http://www.cureffi.org/2012/12/05/power-for-gwas-and-extreme-phenotype-studies/

Monceau K, Wattier R, Dechaume-Moncharmont F, Dubreuil C, Cezilly F (2013) Heterozygosity-fitness correlations in adult and juvenile zenaida dove, *Zenaida aurita*. *Journal of Heredity* 104, 47-56.

Morin PA, Luikart G, Wayne RK, SNP Workshop Grp (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution* 19, 208-216.

Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics* 10, 135-151.

Moskvina V, Schmidt K (2008) On multiple-testing correction in genome-wide association studies. *Genetic Epidemiology* **32**, 567-573.

Nakaoka H, Inoue I (2009) Meta-analysis of genetic association studies: methodologies, between-study heterogeneity and winner's curse. *Journal of Human Genetics* 54, 615-623.

Namroud M-C, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology* 17, 3599-3613.

Narzisi G, Mishra B (2011) Comparing de novo genome assembly: the long and short of it. *PloS ONE* 6, e19175.

Neff B (2004) Stabilizing selection on genomic divergence in a wild fish population. *Proceedings of the National Academy of Sciences of the United States of America* 101, 2381-2385.

Nielsen R, Paul J, Albrechtsen A, Song Y (2011) Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* 12, 443-451.

Olano-Marin J, Mueller JC, Kempenaers B (2011a) Correlations between heterozygosity and reproductive success in the blue tit (*Cyanistes caeruleus*): an analysis of inbreeding and single locus effects. *Evolution* 65, 3175–3194.

Olano-Marin J, Mueller JC, Kempenaers B (2011b) Heterozygosity and survival in blue tits (*Cyanistes caeruleus*): contrasting effects of presumably functional and neutral loci. *Molecular Ecology* 20, 4028–4041.

Olson ZH, Whittaker DG, Rhodes OE (2013) Translocation history and genetic diversity in reintroduced bighorn sheep. *The Journal of Wildlife Management* 77, 1553–1563.

Ondov B, Varadarajan A, Passalacqua K, Bergman N (2008) Efficient mapping of Applied Biosystems SOLiD sequence data to a reference genome for functional genomic applications. *Bioinformatics* 24, 2776-2777.

Osborne A (2011) Assessment of genetic variation in the threatened New Zealand sea lion, *Phocarctos hookeri*, and its association with fitness, University of Otago.

Otter K, Stewart I, McGregor P, et al. (2001) Extra-pair paternity among Great tits *Parus major* following manipulation of male signals. *Journal of Avian Biology* 32, 338-344.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics* 26, 177-187.

Paterson T, Graham M, Kennedy J, Law A (2012) VIPER: a visualisation tool for exploring inheritance inconsistencies in genotyped pedigrees. *BMC Bioinformatics* **13**, S5.

Pelletier F, Festa-Bianchet M (2006) Sexual selection and social rank in bighorn rams. *Animal Behaviour* 71, 649-655.

Pelletier F, Festa-Bianchet M, Jorgenson J (2012) Data from selective harvests underestimate temporal trends in quantitative traits. *Biology Letters* **8**, 878-881.

Pertoldi C, Wójcik J, Tokarska M, et al. (2010) Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* 11, 627-634 .

Philippe H, Telford M (2006) Large-scale sequencing and the new animal phylogeny. *Trends in Ecology & Evolution* 21, 614-620.

Pilot M, Jedrzejewski W, Branicki W, et al. (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology* 15, 4533-4553.

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2012) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity* 108, 256–263.

Poissant J, Hogg JT, Davis CS, et al. (2010). Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. *BMC Genomics*, 11.

Poissant J, Shafer ABA, Davis CS, et al. (2009) Genome-wide cross-amplification of domestic sheep microsatellites in bighorn sheep and mountain goats. *Molecular Ecology Resources* 9, 1121-1126.

Poissant J, Wilson AJ, Festa-Bianchet M, Hogg JT, Coltman DW (2008) Quantitative genetics and sex-specific selection on sexually dimorphic traits in bighorn sheep. *Proceedings of the Royal Society B-Biological Sciences* 275, 623-628.

Porto-Neto L, Kijas J, Reverter A (2014) The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genetics Selection Evolution* **46**.

Prado-Martinez J, Sudmant PH, Kidd JM, et al. (2013) Great ape genetic diversity and population history. *Nature* 499, 471-475.

Price AL, Patterson NJ, Plenge RM, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38, 904-909.

Purcell S, Cherny S, Sham P (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* **19**, 149-150.

Purcell S, Neale B, Todd-Brown K, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81, 559-575.

Puritz JB, Matz MV, Toonen RJ, et al. (2014) Demystifying the RAD fad. *Molecular Ecology* 23, 5937-5942.

Quinlan A, Hall I (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.

R Core Team (2005) R: A language and environment for statistical computing, reference index version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2012) R: A language and environment for statistical computing, reference index version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria.

R Core Team (2014) R: A language and environment for statistical computing, reference index version 3.2.1. R Foundation for Statistical Computing, Vienna, Austria.

Réale D, Festa-Bianchet M (2000) Quantitative genetics of life-history traits in a long-lived wild mammal. *Heredity* **85**, 593-603.

Réale D, Martin J, Coltman DW, Poissant J, Festa-Bianchet M (2009) Male personality, life-history strategies and reproductive success in a promiscuous mammal. *Journal of Evolutionary Biology* 22, 1599-1607.

Reich DE, Cargill M, Bolk S, et al. (2001) Linkage disequilibrium in the human genome. *Nature* 411, 199-204.

Rezaei HR, Naderi S, Chintauan-Marquier IC, et al. (2009) Evolution and taxonomy of the wild species of the genus Ovis (Mammalia, Artiodactyla, Bovidae). *Molecular Phylogenetics and Evolution* 54, 315-326.

Rioux-Paquette E, Festa-Bianchet M, Coltman D (2010) No inbreeding avoidance in an isolated population of bighorn sheep. *Animal Behaviour* 80, 865-871.

Robinson MR, Santure AW, DeCauwer I, Sheldon BC, Slate J (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology* 22, 3963-3980.

Rubenstein D (2007) Female extrapair mate choice in a cooperative breeder: trading sex for help and increasing offspring heterozygosity. *Proceedings of the Royal Society B-Biological Sciences* 274, 1895-1903.

Ruiz-Lopez M, Ganan N, Godoy J, et al. (2012) Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *Conservation Biology* 26, 1121-1129.

Ryynänen HJ, Tonteri A, Vasemägi A, Primmer CR (2007) A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *Journal of Heredity* 98, 692-704.

Sacks BN, Louie S (2008) Using the dog genome to find single nucleotide polymorphisms in red foxes and other distantly related members of the Canidae. *Molecular Ecology Resources* 8, 35-49.

Sambrook J, Russell D (2001) *Molecular cloning: a laboratory manual*, Cold Spring Harbor (NY).

Santure A, De Cauwer I, Robinson M, et al. (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Molecular Ecology* 22, 3949-3962.

Santure A, Stapley J, Ball A, et al. (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology* 19, 1439-1451.

Sasaki Y, Nagai K, Nagata Y, et al. (2006) Exploration of genes showing intramuscular fat deposition-associated expression changes in musculus longissimus muscle. *Animal Genetics* 37, 40-46.

Sechi T, Coltman DW, Kijas JW (2010) Evaluation of 16 loci to examine the cross-species utility of single nucleotide polymorphism arrays. *Animal Genetics* 41, 199-202.

Shackleton DM, Shank CC, Wikeem B (1999) Natural history of Rock Mountain and California bighorn sheep. In: *Mountain Sheep of North America* (eds. Valdez R, Krausman PR), pp. 78-138. The University of Arizona Press, Tuscon.

Shafer ABA, Wolf JBW, Alves PC, et al. (2015) Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* 30, 78–87.

Sham P, Cherny S, Purcell S, Hewitt J (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *American Journal of Human Genetics* **66**, 1616-1630.

Sheehan S, Harris K, Song YS (2013) Estimating variable effective population sizes from multiple genomes: A sequentially markov conditional sampling distribution approach. *Genetics* 194, 647-662.

Shen R, Fan J-B, Campbell D, et al. (2005) High-throughput SNP genotyping on universal bead arrays. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 573, 70-82.

Slate J, David P, Dodds KG, et al. (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* 93, 255-265.

Slate J, Gratten J, Beraldi D, et al. (2009) Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* 136, 97-107.

Slate J, Kruuk LEB, Marshall TC, Pemberton JM, Clutton-Brock TH (2000) Inbreeding depression influences lifetime breeding success in a wild population of red deer (*Cervus*

elaphus). *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267, 1657-1662.

Slate J, Pemberton J (2002) Comparing molecular measures for detecting inbreeding depression. *Journal of Evolutionary Biology* 15, 20-31.

Slate J, Pemberton J (2007) Admixture and patterns of linkage disequilibrium in a free-living vertebrate population. *Journal of Evolutionary Biology* 20, 1415-1427.

Sterne J, Becker B, Egger M, et al. (2005a) The Funnel Plot. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 75-98.

Sterne J, Egger M, Rothstein H, Sutton A, Borenstein M (2005b) Regression Methods to Detect Publication and Other Bias in Meta-Analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 99-110.

Stewart I, Westneat D (2013) Patterns of hatching failure in the house sparrow *Passer domesticus*. *Journal of Avian Biology* 44, 69-79.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100, 158-170.

Stram DO (2004) Tag SNP selection for association studies. *Genetic Epidemiology* 27, 365-374.

Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *Plos ONE* 6.

Sutter NB, Eberle MA, Parker HG, et al. (2004) Extensive and breed-specific linkage disequilibrium in *Canis familiaris*. *Genome Research* 14, 2388-2396.

Szulkin M, Bierne N, David P (2010) Heterozygosity-fitness correlations: A time for reapprasial. *Evolution* 64, 1202-1217.

Taylor S, Sardell R, Reid J, et al. (2010) Inbreeding coefficient and heterozygosity-fitness correlations in unhatched and hatched song sparrow nestmates. *Molecular Ecology* 19, 4454-4461.

Telford MJ, Copley RR (2011) Improving animal phylogenies with genomic data. *Trends in Genetics* 27, 186-195.

The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851-861.

Townsend SM, Jamieson IG (2013) Molecular and pedigree measures of relatedness provide similar estimates of inbreeding depression in a bottlenecked population. *Journal of Evolutionary Biology* 26, 889–899.

Tozaki T, Hirota K, Hasegawa T, Tomita M, Kurosawa M (2005) Prospects for whole genome linkage disequilibrium mapping in thoroughbreds. *Gene* 346, 127-132.

Traill LW, Schindler S, Coulson T (2014) Demography, not inheritance, drives phenotypic change in hunted bighorn sheep. *Proceedings of the National Academy of Sciences* **111**, 13223–13228.

Tsitrone A, Rousset F, David P (2001) Heterosis, marker mutational processes and population inbreeding history. *Genetics* 159, 1845-1859.

Umemura M, Koyama Y, Takeda I, et al. (2013) Fine *de novo* sequencing of a fungal genome using only SOLiD short read data: Verification on *Aspergillus oryzae* RIB40. *Plos ONE* 8.

Valdez R, Krausman PR (1999) Description, distribution, and abundance of mountain sheep in North America. In: *Mountain Sheep of North America* (eds. Valdez R, Krausman PR), pp. 3-22. The University of Arizona Press, Tuscon.

Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology* 17, 3808-3817.

VanLiere JM, Rosenberg NA (2008) Mathematical properties of the r2 measure of linkage disequilibrium. *Theoretical Population Biology* 74, 130-137.

Vijay N, Poelstra JW, Künstner A, Wolf JBW (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. *Molecular Ecology* 22, 620-634.

Voegeli B, Saladin V, Wegmann M, Richner H (2013) Heterozygosity is linked to the costs of immunity in nestling great tits (*Parus major*). *Ecology and Evolution* 3, 4815–4827.

Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in Ecology & Evolution* 11, 158-162.

Wang B, Ekblom R, Bunikis I, Siitari H, Hoglund J (2014) Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution. *BMC Genomics* 15, 180.

Wang J (2013) An improvement on the maximum likelihood reconstruction of pedigrees from marker data. *Heredity* **111**, 165-174.

Weeks AR, Sgro CM, Young AG*, et al.* (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* **4**, 709–725.

Wetzel D, Stewart I, Westneat D (2012) Heterozygosity predicts clutch and egg size but not plasticity in a house sparrow population with no evidence of inbreeding. *Molecular Ecology* 21, 406-420.

White T, Searle J (2008) Mandible asymmetry and genetic diversity in island populations of the common shrew, *Sorex araneus*. *Journal of Evolutionary Biology* 21, 636-641.

Wickham H (2009) *ggplot2: elegant graphics for data analysis* Springer, New York.

Wiedemar N, Tetens J, Jagannathan V*, et al.* (2014) Independent Polled Mutations Leading to Complex Gene Expression Differences in Cattle. *Plos ONE* **9**.

Wilson AJ, Kruuk LEB, Coltman DW (2005) Ontogenetic Patterns in Heritable Variation for Body Size: Using Random Regression Models in a Wild Ungulate Population. *The American Naturalist* **166**.

Wilson AJ, Réale D, Clements MN*, et al.* (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology* **79**, 13-26.

Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources* 13, 559–572.

Worley K, Strobeck C, Arthur S, et al. (2004) Population genetic structure of North American thinhorn sheep (*Ovis dalli*). *Molecular Ecology* 13, 2545-2556.

Yandell M, Ence D (2012) A beginner's guide to eukaryotic genome annotation. *Nature Reviews Genetics* 13, 329-342.

Yang J, Benyamin B, McEvoy BP*, et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-569.

Yang J, Lee S, Goddard M, Visscher P (2011a) GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics* **88**, 76-82.

Yang J, Manolio T, Pasquale L*, et al.* (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics* **43**, 519-U544.

Yang J, Zaitlen N, Goddard M, Visscher P, Price A (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics* **46**, 100-106.

Yuan FPL, X Lin, J Schwabe, C Büllesbach, E E Rao, C V Lei, Z M (2010) The role of RXFP2 in mediating androgen-induced inguinoscrotal testis descent in LH receptor knockout mice. *Reproduction* **139**, 759-769.

Zachos F, Hartl G, Suchentrunk F (2007) Fluctuating asymmetry and genetic variability in the roe deer (*Capreolus capreolus*): a test of the developmental stability hypothesis in mammals using neutral molecular markers. *Heredity* 98, 392-400.

Zhao H, Bourque G (2009) Recovering genome rearrangements in the mammalian phylogeny. *Genome Research* 19, 934-942.

Zollner S, Pritchard J (2007) Overcoming the winner's curse: Estimating penetrance parameters from case-control data. *American Journal of Human Genetics* 80, 605-615.

APPENDECIES

Appendix 5-1: Laboratory methods for candidate loci validation

Primer sequences for Type-it multiplexing

Locus	F Primer	R Primer	Product Length
OAR14_45166067	GGGGACAGGACATGTTACAA	CGATCTGCAACAGGGGATA	119
OAR12_2916143	TGATTCGCCCAGACAATCT	AGATCTTGAGCTCTTATCTGTCATAAA	117
OAR24_23754378	TGGATGGGTTTGAAAATCTG	CAGTGATGCACAGCACAATC	116
OAR24_5898966*	CAAGTGTTTAAATATGAAACCAAGAAA	CTCCAAACCACACTGTAGCC	116
OAR3_138991772	CTTATATGCCCCCAAAACTTTC	GTGTTGGGTGTGAATGATGC	100
OAR19_7407385	TTTGTGGGAGAGAAGCCAAT	AGAGGAAGGATGGCTGACTG	100
OAR12_2915825	ATGTGAGCGAGGAGCATGTA	GGCATGAGGTCGTAGGAAAT	88
OAR10_18068194	CCACTGCATGCCAGAGTT	TTGAGTTTGAATTGCCTGCT	87
OAR15_13537505	GAAATGTTGGACAGATATAAAAGTCAT	TGGTGTCTGTGTTAAACCTTGA	85
OAR2_148529592	CAAATCTTTATTTATGGGCAACC	GTGGTGGAGGCATTTGTGAC	107
OAR10_85023560.1	TGGGTGAAAAGACTCAGAGGA	GCCTTGTCCCCACTACTGTC	99

*Locus failed to amplify in all individuals

For Type-it reactions a working primer solution was made containing all primers diluted to a final concentration of 2uM. Type-it reactions were then conduced in 10uL volumes using 5uL 2x Type-it mix (Qiagen), 1uL of 2uM primer solution, 2uL water, and 2uL genomic DNA (diluted to ~8ng/uL). The thermocycler profile for the type-it reaction was a 95C hold for 5 min, then 28 cycles of 95C for 30 sec, 57C for 90 sec, and 72C for 30 sec followed by 60C for 30min, and finally a hold a 4C. The resulting reactions were cleaned by adding 4uL of a mix of 0.6U ExoI and 2.2U SAP (New England Biolabs) to each reaction. The thermocycler profile for cleaning was 37C for 30 min, 80C for 15 min, then a 4C hold.

Probe sequences for SNaPshot assays

Locus	Probe	Probe Length	Final Length§
OAR14_45166067	CATGTTACAAAACAGGAAAC	20	29
OAR12_2916143	ACTGTGTAATAACATGCATATC	22	37
OAR24_23754378	TCTAAGTACAGGCCTGGTA	19	45

OAR24_5898966*	TCCATGACTNGAAGGG	16	53
OAR3_138991772	AAATCAAAGAAACCAAGTTAC	21	61
OAR19_7407385	CAAAGAAGATAAGAAAGTGC	20	69
OAR12_2915825	CTATCAAAGGCTCAGACC	18	77
OAR10_18068194	TCTGAAAGCCATAGGAAC	18	85
OAR15_13537505	GGGTCCCATGTACTCC	16	90
OAR2_148529592	TTTTTACCTTTGTGCAGA	18	30
OAR10_85023560.1	CCATTCTCAAGGAGCTT	17	40

§Poly-T tails were added to each probe to allow adequate separation for genotyping
*Locus failed to amplify in all individuals

For SNaPShot assays a working probe solution was used containing a combination of all interrogation probes diluted to a final concentration of 2uM. SNaPShot reactions were carried out in 10uL volumes using 5uL SNaPShot master mix (Life Technologies), 3uL Type-it PCR product, 1uL 2uM probe mix, and 1uL water. The thermocycler profile for SNaPShot was 25 cycles of 96C for 10 sec, 50C for 5 sec, 60C for 30 sec, followed by a 4C hold. This reaction was then cleaned using 2uL of a 0.33U SAP solution to each reaction. The thermocycler profile for cleaning was 37C for 30 min, 80C for 15 min, then a 4C hold. 1.8uL of the interrogation product was mixed with 0.2uL GeneScan 120 Liz and 8uL Formamide and resolved on an ABI3730 DNA Analyzer.

Appendix 5-2: Number and chromosomal distribution of markers used in the GWAS analysis.

Chromosome	No. Loci	Avg. Inter-marker Distance	SD of Inter-marker Distance
1	429	639,475.92	722,591.62
2	375	662,254.19	780,049.65
3	270	824,399.96	1,012,272.38
4	172	685,797.94	737,062.79
5	166	638,050.85	794,621.76
6	142	820,292.52	961,026.48
7	186	527,153.04	607,010.96
8	139	647,412.07	683,785.33
9	129	715,575.23	924,268.73
10	142	605,600.96	708,386.17
11	101	591,167.02	673,572.80
12	134	590,221.62	789,103.56
13	94	880,912.27	908,979.65
14	116	527,956.87	539,733.97
15	170	473,971.36	588,209.46
16	105	656,399.16	738,427.39
17	106	669,053.35	762,647.92
18	110	604,523.53	619,577.38
19	109	556,014.41	648,734.61
20	147	325,445.54	621,916.39
21	63	747,203.00	782,700.54
22	56	841,430.49	961,636.08
23	81	766,115.53	836,916.85
24	73	539,960.92	651,886.78
25	82	541,281.88	579,233.56
26	80	548,843.14	588,078.94

Appendix 5-3: Delta AICc values between null models and models containing locus genotype for life-history associated suggestive loci. Negative numbers indicate a better fit for models containing locus genotypes. All models simultaneously considered individuals genotyped on the 700k SNP chip and by SNaPshot reactions.

Locus Name	Original Association	Number of Offspring	Longevity	Age at Primiparity	Fecundity
OAR10_18068194	Fecundity	-3.4	3.1	2.9	1
OAR12_2915825	Longevity & No. Offspring	2	-2.8	2	1.6
OAR12_2916143	Longevity & No. Offspring	2	-2.8	2	1.6
OAR15_13537505	Longevity	-0.9	2	4.2	-2
OAR19_7407385	Age of Primiparity	0	-2.4	0.8	1.9
OAR24_23754378	Fecundity	-9.3	-4.9	4.2	0.8

Appendix 5-4: Gene names and biological process GO terms for genes within genomic regions surrounding the morphology associated loci.

Chr	Trait	Gene ID	Gene Name	Associated GO Terms
OAR2	Body Mass	RPL17	ribosomal protein L17	
OAR2	Body Mass	SNORA21	small nucleolar RNA, H/ACA box 21	
OAR2	Body Mass	ITGB6	integrin, beta 6	
OAR2	Body Mass	PLA2R1	phospholipase A2 receptor 1, 180kDa	cytokine production; receptor-mediated endocytosis; reactive oxygen species metabolic process; positive regulation of DNA damage response, signal transduction by p53 class mediator; replicative senescence;
OAR2	Body Mass	CD302	CD302 molecule	phagocytosis
OAR2	Body Mass	MARCH7	membrane-associated ring finger (C3HC4) 7, E3 ubiquitin protein ligase	
OAR2	Body Mass	RBMS1	RNA binding motif, single stranded interacting protein 1	
OAR3	Horn Length	SLC48A1	solute carrier family 48 (heme transporter), member 1	heme transport
OAR3	Horn Length	PCED1B	PC-esterase domain containing 1B	metabolic process
OAR3	Horn Length	AMIGO2	adhesion molecule with Ig-like domain 2	
OAR3	Horn Length	HDAC7	histone deacetylase 7	negative regulation of transcription from RNA polymerase II promoter; cell-cell junction assembly; vasculogenesis; chromatin modification; negative regulation of interleukin-2 production; negative regulation of osteoblast differentiation;
OAR3	Horn Length	RAPGEF3	Rap guanine nucleotide exchange factor (GEF) 3	establishment of endothelial barrier; cellular response to cAMP; regulation of actin cytoskeleton reorganization; regulation of protein kinase activity; small GTPase mediated signal transduction; Rap protein signal transduction; positive regulation of angiogenesis; positive regulation of Rap GTPase activity; intracellular signal transduction; regulation of small GTPase mediated signal transduction;

OAR3	Horn Length	ENDOU	endonuclease, polyU-specific	receptor-mediated endocytosis; immune response; metabolic process; proteolysis; RNA phosphodiester bond hydrolysis, endonucleolytic; female pregnancy;
OAR3	Horn Length	RPAP3	RNA polymerase II associated protein 3	
OAR10	Body Mass	5S_rRNA		
OAR10	Body Mass	FGF14	fibroblast growth factor 14	nervous system development; JNK cascade; synaptic transmission; adult locomotory behavior; neuromuscular process; positive regulation of sodium ion transport;
OAR10	Body Mass	ITGBL1	integrin, beta-like 1 (with EGF-like repeat domains)	
OAR10	Body Mass	TPP2	tripeptidyl peptidase II	proteolysis
OAR14	Horn Base	FXYD3	FXYD domain containing ion transport regulator 3	ion transport; ion transmembrane transport; regulation of catalytic activity;
OAR14	Horn Base	FXYD1	FXYD domain containing ion transport regulator 1	ion transmembrane transport; regulation of cardiac muscle cell membrane potential; positive regulation of sodium ion export from cell; ion transport;
OAR14	Horn Base	HAMP	hepcidin antimicrobial peptide	
OAR14	Horn Base	FFAR1	free fatty acid receptor 1	G-protein coupled receptor signaling pathway; positive regulation of calcium ion transport; response to fatty acid; glucose homeostasis; insulin secretion; insulin secretion;
OAR14	Horn Base	FFAR2	free fatty acid receptor 2	glucose homeostasis; fat cell differentiation; cellular response to fatty acid; positive regulation of chemokine production; mucosal immune response; lipid storage; regulation of peptide hormone secretion; leukocyte chemotaxis involved in inflammatory response; cell surface pattern recognition receptor signaling pathway; positive regulation of acute inflammatory response to non-antigenic stimulus; signal transduction; positive regulation of cytokine production involved in immune response; regulation of acute inflammatory response;
OAR14	Horn Base	KRTDAP	keratinocyte differentiation-associated protein	

OAR14	Horn Base	DMKN	dermokine	
OAR14	Horn Base	SBSN	suprabasin	
OAR14	Horn Base	TMEM147	transmembrane protein 147	
OAR14	Horn Base	ETV2	ets variant 2	Notch signaling pathway; placenta development; positive regulation of gene expression; positive regulation of mesoderm development; regulation of transcription, DNA-templated; blastocyst development; positive regulation of endothelial cell differentiation; Wnt signaling pathway; blood vessel morphogenesis;
OAR14	Horn Base	COX6B1	cytochrome c oxidase subunit VIb polypeptide 1 (ubiquitous)	hydrogen ion transmembrane transport; substantia nigra development;
OAR14	Horn Base	UPK1A	uroplakin 1A	epithelial cell differentiation
OAR14	Horn Base	IGFLR1	IGF-like family receptor 1	
OAR14	Horn Base	ZNF792	zinc finger protein 792	
OAR14	Horn Base	GRAMD1A	GRAM domain containing 1A	
OAR14	Horn Base	SCN1B	sodium channel, voltage gated, type I beta subunit	sodium ion transport; locomotion; membrane depolarization; corticospinal neuron axon guidance; neuronal action potential propagation; regulation of sodium ion transmembrane transporter activity; cardiac muscle contraction; sodium ion transmembrane transport;
OAR14	Horn Base	HPN	hepsin	proteolysis; basement membrane disassembly; negative regulation of alkaline phosphatase activity; sensory perception of sound; cholesterol homeostasis; negative regulation of apoptotic process; positive regulation of plasminogen activation; positive regulation by host of viral transcription; negative regulation of epithelial to mesenchymal transition; positive regulation of cell growth; negative regulation of epithelial cell proliferation; regulation of cell shape;
OAR14	Horn Base	FXYD7	FXYD domain containing ion transport regulator 7	ion transport; ion transmembrane transport;

OAR14	Horn Base	FXYD5	FXYD domain containing ion transport regulator 5	ion transport; ion transmembrane transport;
OAR14	Horn Base	FAM187B	family with sequence similarity 187, member B	
OAR14	Horn Base	LSR	lipolysis stimulated lipoprotein receptor	LSRlipolysis stimulated lipoprotein receptor
OAR14	Horn Base	MAG	myelin associated glycoprotein	substantia nigra development
OAR14	Horn Base	CD22	CD22 molecule	
OAR14	Horn Base	GAPDHS	glyceraldehyde-3-phosphate dehydrogenase, spermatogenic	oxidation-reduction process; sperm motility;
OAR14	Horn Base	ATP4A	ATPase, H+/K+ exchanging, alpha polypeptide	ATP biosynthetic process; response to drug; pH reduction; regulation of proton transport; cation transport; ATP hydrolysis coupled proton transport;
OAR14	Horn Base	HAUS5	HAUS augmin-like complex, subunit 5	spindle assembly
OAR14	Horn Base	KMT2B	lysine (K)-specific methyltransferase 2B	methylation; gene silencing; ovarian follicle development; histone lysine methylation;
OAR14	Horn Base	LGI4	leucine-rich repeat LGI family, member 4	adult locomotory behavior; neuron maturation; glial cell proliferation; myelination;
OAR14	Horn Base	RBM42	RNA binding motif protein 42	negative regulation of mRNA splicing, via spliceosome
OAR14	Horn Base	USF2	upstream transcription factor 2, c-fos interacting	transcription from RNA polymerase II promoter; lactation; positive regulation of transcription from RNA polymerase II promoter by glucose; lipid homeostasis; regulation of transcription from RNA polymerase II promoter; positive regulation of transcription, DNA-templated;