2013

The genetic basis of fitness: detecting inbreeding depression and selective sweeps in bighorn sheep

Martin Dennis Kardos
University of Montana, Missoula

THE GENETIC BASIS OF FITNESS: DETECTING INBREEDING DEPRESSION AND SELECTIVE SWEEPS IN BIGHORN SHEEP

By

Martin Dennis Kardos

B.S., Montana State University, Bozeman, MT, 2005

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Organismal Biology & Ecology

The University of Montana
Missoula, MT

December 2013

Approved by:

Sandy Ross, Dean of The Graduate School
Graduate School

Fred W. Allendorf, Co-Chair
Division of Biological Sciences

Gordon Luikart, Co-Chair
Division of Biological Sciences, Flathead Lake Biological Station

Jeffrey Good
Division of Biological Sciences

Jon Graham
Department of Mathematical Sciences

John McCutcheon
Division of Biological Sciences

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Kardos, Martin D., Ph.D., Fall 2013 *Organismal Biology & Ecology*

Co-Chairperson: Fred W. Allendorf
Co-Chairperson: Gordon Luikart

ABSTRACT

Understanding the fitness effects of inbreeding is a crucial and long standing goal in conservation and evolutionary biology. Many studies measure individual inbreeding (*F,* the proportion of genome that is identical by descent) and its fitness effects using either pedigrees or molecular markers. Knowing which genes most strongly affect fitness can help to explain why some individuals outperform others, and elucidate the mechanisms of inbreeding depression and adaptation. However, identifying adaptive genes is difficult in most species because of limited genomic resources.

I used simulations to evaluate the performance of marker- and pedigree-based measures of *F* and inbreeding depression. I found that F_P was less precise than marker-based measures of *F* in a broad range of scenarios. For example, the true *F* was always more strongly correlated with heterozygosity measured with 5000 single nucleotide polymorphisms (SNPs) than with F_P. *F* was also more strongly correlated with the proportion of the genome in long runs of homozygosity (F_{ROH}, estimated with 35K SNPs) than with F_P. I also show that heterozygosity-based estimates of the strength of inbreeding depression are precise in populations with high variance in *F* (e.g., $\sigma^2(F) \geq 0.002$). A potential solution to the imprecision of F_P is to use genetic markers to correct for the kinship of pedigree founders. However, I found that founder kinship-corrected values of F_P were also imprecise. These results show that *F* and inbreeding depression can be most reliably measured with genetic markers in most scenarios – countering the prevailing historical view that *F* is most reliably measured with pedigrees.

I used whole genome sequences of pooled DNA aligned to the domestic sheep genome to detect candidate adaptive genes in bighorn sheep. I detected selection signatures in 53 genomic regions containing genes. However, simulations suggest that some of these selection signatures may be false positives. Putatively selected genomic regions contained genes involved with traits known to affect fitness in bighorn sheep (e.g., horn and body growth). These results provide candidate genes for traits known to strongly influence fitness in bighorn, and illustrate the great promise of WGS for detecting selection signatures in small populations.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without contributions from many people and organizations. I am particularly grateful to my advisors, Gordon Luikart and Fred W. Allendorf for their guidance throughout my dissertation. I consider myself fortunate to have worked closely with two such talented mentors and leaders in the fields of population genetics and conservation biology. I also thank the following people and organizations for their substantial contributions to my dissertation:

My dissertation committee:
Jeff Good, Jon Graham, John McCutcheon (*current*),
Vanessa Ezenwa, L. Scott Mills, and Mark Hebblewhite (*former*)

My collaborators:
James Kijas, Jack Hogg, Sarah Dewey, John Stephenson, Rowan Bunch, Sean McWilliam, Hank Edwards, and Helen Taylor

Montana Ecology of Infectious Diseases IGERT PhD students, staff, and faculty:
Stefan Ekernas, Carin Williams, Tara Westlie, Allen Warren, Nick McClure, Michael Ceballos, Ted Cosart, Joran Elias, Erin Landguth, Julie Betch, Kellie Carim, Meredith Berthelson, Tarun Gupta, Brian Hand, Clark Kogan, Ellen Lark, Josh Marceau, Marnie Rout, Tammi Johnson, Sarah Budischak, Flo Gardipee, Robin Silverstein, Bill Holben, Jesse Johnson, Jon Graham, John Bardsley, Leonid Kalachev, John McCutcheon, Creagh Bruhner, Greg Larson, and Alden Wright

Montana Conservation Genetics Lab:
Steve Amish, Sally Painter, Angela Lodmell, Robb Leary, Dan Bingham, Ruth Shortbull, Jody Tucker, Mariah Childs, Yves Horreau, Jenna Schabacker, Helen Taylor, Kristina Ramstad

Funding:
I have been supported by the Montana Ecology of Infectious Diseases IGERT program (NSF, DGE-0504628), the Montana Institute on Ecosystems (NSF, EPS-1101342), and a teaching assistantship from the University of Montana. Agencies and individuals that supported particular chapters of my dissertation are acknowledged below.

Division of Biological Sciences Staff:
Sherrie Wright, Robin Hamilton, Janean Clark

Family:
Ali Dimond, Max Kardos, Miles Kardos, Tucker Kardos, Meleta Kardos, Steve Kardos, Connie Dimond, Paul Dimond

TABLE OF CONTENTS

LIST OF FIGURES

Figure 3-1. ROH in a chromosome from the offspring resulting from one generation of selfing with (left) and without recombination (right). A single pair of homologous chromosomes are shown from a non-inbred parent (top) and the offspring of the individual (bottom). IBD segments occur in the offspring where both chromosomes are derived from a single chromosome in the parent. The mean PGIBD will be the same in both cases, but the variance of PGIBD is much greater without recombination.

Figure 3-2. Barplots of the mean r^2 (+/- 1 s.d. across 20 simulated populations) from regressions of F_P, F_H and F_{ROH} versus PGIBD. Results shown here are from simulations of genomes with a genetic map length of 3600 cM. Results from 20 partially isolated (m = 0.05) small populations (local N_e = 20) are shown in the top row. The data shown in the bottom row are from 20 populations with a recent reduction in N_e (from N_e = 500 to N_e = 20). Horizontal dotted lines are placed at r^2 = 0.9 to aid comparison of r^2 across F_P, F_H and F_{ROH}.

Figure 3-3. F_P (A), F_H (B), and F_{ROH} (C) versus PGIBD in a representative simulation of a partially isolated population with a genetic map length of 3600 cM. We used a pedigree that included 5 generations to estimate F_P. F_H was estimated with 5K SNPs, and F_{roh} was estimated with 35K SNPs. The dashed lines have an intercept of zero and a slope of one. Points below the lines represent underestimates of PGIBD.

Figure 3-4. The bias of F_P, F_H, and F_{ROH} among simulations of genomes with a genetic map length of 3600 cM. Results from 20 simulations of partially isolated small populations are shown in the top row. Results from 20 simulations of populations with a recent reduction in N_e are shown in the bottom row.

Figure 4-1. The estimated inbreeding coefficient (F_P) versus the identical by descent (IBD) proportion of the genome (F). The results shown are from simulations of six generation pedigrees with 0-100% sibling founders (from the same pair of parents). The solid diagonal lines have intercept of zero and a slope of 1. Points below the line represent underestimates of F.

Figure 4-2. r^2 from the regressions of the pedigree inbreeding coefficient (F_P) and the corrected coefficient F_{PC} (using molecular markers) versus the true inbreeding (F; identitity by descent) plotted against the proportion of pedigree founders that were full siblings. 100 and 500 SNPs (mean $H_e \approx 0.3$) were used to estimate founder kinship coefficients which were used to estimate F_{PC}. Error bars represent the standard deviation of r^2 among twenty replicate simulated populations.

Figure 4-3. The r^2 between the estimated inbreeding coefficient (F_P) and the IBD proportion of the genome (F) ($r^2(F_E, F_T)$) versus pedigree depth. The data are from the final generation (60 individuals) of each of 50 independent simulations that were run for 70 generations with migration rates of 0.008 to 0.1 (0.5 - 6.0 migrants/generation). Error bars represent the standard deviation of r^2 among the 50 replicate simulations. Horizontal dashed lines in panel **B** represent the power to detect inbreeding depression using the true F. The vertical distance between the top of a bar and the dashed line of the same color represents the loss of power associated with imprecision of F_P.

Figure 4-4. (**A**) The r^2 between the pedigree inbreeding coefficient (F_P) and the identical by descent proportion of the genome (F) [r^2 (F_P, F), open bars] and between the pedigree inbreeding coefficient corrected for founder kinship (F_{PC}) and F [r^2 (F_{PC}, F), hatched bars] versus the pedigree depth. Error bars in panel **A** are the standard deviation or r^2 among 50 replicate simulations. (**B**) The statistical power of tests for inbreeding depression using F_P and F_{PC}. Error bars in panel **B** are 95% confidence intervals for statistical power (the proportion of 50 replicate simulations with statistically significant tests [$\alpha = 0.05$] for inbreeding depression). The dashed horizontal line in panel **B** represents the power to detect inbreeding depression when using the true individual inbreeding coefficient (F; genome identity by descent).The data shown are from simulated populations with a migration rate of 0.017 (1 migrant/generation on average). Stars indicate a statistically significant difference between the mean r^2 (F_P, F) and mean r^2 (F_{PC}, F) among 50 replicate simulations. Results from simulations of other demographic scenarios are shown in Figure 4-S1.

Figure 4-5. The estimated coefficient of kinship ($\hat{f}$) versus the true coefficient of kinship (f) for all pairs of 60 simulated individuals. The data are from the final generation of a simualtions with m = 0.017 that was run for 70 generations. The diagonal dashed lines have an intercept of 0 and a slope of 1. Points above the line represent over estimates and points below the line represent under estimates. The solid lines are fitted 2nd degree loess functions.

Figure 5-1. ZH_P for the Teton populations (bottom panel) and ZF_{ST} (top panel) across the bighorn sheep genome. Chromosomes are arranged 1-26 (left to right). The red line represents the rolling mean across 100 sliding windows. The dashed lines represent the threshold of significance of 5 standard deviations from the mean H_P and F_{ST} across the whole genome.

Figure 5-2. H_P in the Teton pool and F_{ST} across putatively selected regions on chromosomes 10 (**A**), 16 (**B**), 8 (**C**), and 2 (**D**). Orange points represent F_{ST} and blue points represent H_P. F_{ST} and H_P are shown in 100Kb windows in the top panels. F_{ST} and H_P at individual SNPs in the putatively selected regions are shown in the bottom panels. Genes labeled as "*UNC*" are uncharacterized.

Figure 5-3. ZH_P sliding window estimates from 20 simulations of neutrally evolving populations with a demographic history approximately similar to the Teton populations. Simulations were run for 100 generations with local $N_e = 30$ (top panel) and $N_e = 20$ (bottom panel). Two chromosomes were simulated for each population. The two chromosomes from each simulated population are represented by adjacent blocks of black and gray points. The simulations with 0.5 migrants per generation on average between the two Teton subpopulations and a large source population. There was a migration rate of 0.75 individual/generation on average between the two Teton subpopulations. The mean F_{ST} was 0.09 (min. = 0.05, max. = 0.12) among the 20 simulations of populations with $N_e = 30$. The mean F_{ST} across the 20 simulations with $N_e = 20$ was 0.1 (min. = 0.06, max. = 0.15). The dashed lines represent the threshold value of ZH_P that we used for our empirical data to consider a window as being putatively selected

Chapter 1: Introduction and Overview

A major goal in evolutionary biology is to understand the genetic basis of phenotypic and fitness variation in natural populations. Two related areas of intense interest in evolutionary biology are (1) understanding the fitness effects of inbreeding (mating between relatives) in natural populations (Keller & Waller 2002); and (2) identifying loci that strongly affect fitness and adaptation in natural populations (Stinchcombe & Hoekstra 2007; Ellegren & Sheldon 2008; Barrett & Hoeskstra 2011). Inbred individuals often have reduced fitness because they are more likely to be homozygous at loci carrying deleterious recessive alleles or at loci with heterozygous advantage (Keller & Waller 2002). Inbreeding depression is thought to play an important role in the evolution of mating systems and inbreeding avoidance behaviors such as sex-biased dispersal and avoidance of mating with close relatives (Charlesworth & Charlesworth 1987). Additionally, inbreeding depression can increase the extinction risk of small and isolated populations (Mills & Smouse 1994; O'Grady *et al.* 2006; Saccheri *et al.* 1998; Westemeier *et al.* 1998). Researchers have been aware of inbreeding depression for well over 100 years (Darwin 1868). However, there is currently keen interest in understanding the genetic basis of inbreeding depression, and its importance to population growth and persistence (Ouberg 2010).

Much of what we know about inbreeding depression is based on studies that have used pedigrees to measure individual inbreeding. This approach usually involves estimating the correlation between a fitness-related trait and the pedigree inbreeding coefficient (F_P, Wright 1922; Keller & Waller 2002). F_P has historically been strongly preferred over marker-based measures of F (Pemberton 2004) because of the low precision of marker-based measures when relatively few loci are used (Balloux *et al*. 2004; Slate *et al*. 2004).

Disadvantages and alternatives to pedigrees

There are several disadvantages to using pedigrees to estimate F. First, F can vary greatly among individuals with the same pedigree (e.g., siblings) because of linkage and recombination (Franklin 1977; Stam 1980; Hill & Weir 2011). Second, pedigrees are often extremely difficult and time consuming to obtain in natural populations, and inferences drawn from pedigree studies can be sensitive to errors in the assignment of parentage (Pemberton 2008). Lastly, F_P assumes that the founders of pedigrees are both non-inbred, and unrelated to one another (Keller & Waller 2002). Neither of these assumptions is likely to hold in any natural population.

There has been much interest recently in whether F could be more accurately measured with marker-based approaches that employ large number of markers (e.g., hundreds to thousands) than with pedigrees (Balloux *et al*. 2004; Slate *et al*. 2004; Szulkin *et al*. 2010; Forstmeier *et al*. 2012). Indeed, since the advent of genetic markers, researchers have used heterozygosity-fitness correlations (HFCs) to test for inbreeding depression in populations where pedigrees were unavailable (Szulkin *et al*. 2010; Chapman *et al*. 2009). HFCs can be caused by inbreeding depression because individuals with more closely related parents also have lower heterozygosity across the genome (Crow & Kimura 1970). However, there has been intense

debate about whether HFCs are generally caused by inbreeding depression or other mechanisms including selection at the genotyped markers themselves, or closely linked loci.

There has until recently been an almost unanimous view that *F* could be more accurately measured with pedigrees than with genetic markers (Pemberton 2004; 2008; Balloux *et al*. 2004; Slate *et al*. 2004; Forstmeier *et al*. 2012). The broadly held view that *F* could be more precisely measured with pedigrees than with genetic markers is rooted in the observation of very high sampling variance in estimates of individual heterozygosity based on relatively few genetic markers (e.g., 5-20 microsatellite loci). However, thousands of single nucleotide polymorphisms can now be genotyped in any organism (Davey *et al*. 2011). The sampling error of marker-based measures of *F* depends crucially on the number of markers (David *et al*. 2007; Slate *et al*. 2004; Szulkin *et al*. 2010). Therefore *F* could potentially be more precisely measured using marker-based approaches instead of pedigrees (Forstmeier *et al*. 2012; Robinson *et al*. 2013) particularly now that it is possible to type thousands of SNPs in non-model organisms.

Identifying candidate adaptive genes

Identifying genes with strong fitness effects can help to advance our understanding of the mechanisms of natural selection (Stinchcombe & Hoekstra 2007). Knowing what genes and physiological pathways most strongly affect fitness or are responsible for adaptation can help to explain why some individuals perform better than others, and how different populations or species become adapted to their habitats. Now that it is possible to type any organism at thousands of loci, scans for genomic signatures of selection are possible in most taxa. For example, genome scans for selective sweeps (a molecular signature of a response to directional selection) have been used to identify candidate adaptive genes in the wild (Turner et al. 2010), and in domesticated animals (Rubin *et al*. 2010; Axelsson *et al*. 2012).

A pervasive challenge in studies of genomic signatures of selection, particularly in small populations, is to differentiate signatures of selection from similar molecular patterns caused by genetic drift (Jensen *et al*. 2007; Pavlidis *et al*. 2012). Some approaches have been devised to limit or quantify the likelihood of false positive signatures of selection (Neilsen *et al*. 2005; Pavlidis *et al*. 2012; Hohenlohe *et al*. 2010; Qanbari *et al*. 2012). However, these approaches either require an understanding of historical population dynamics (e.g., Neilsen *et al*. 2005) or assume that the minor allele frequencies at closely linked loci are independent (e.g., Hohenlohe *et al*. 2010; Qanbari *et al*. 2012). Thus, even with the availability of rigorous statistical methods to differentiate signals of selection from those of drift, doing so in small natural populations with limited information on historical demography is challenging.

The overarching objectives of my dissertation were the following:

1. Evaluate the performance of existing methods based on identity disequilibrium to
    a. Determine if observed HFCs are caused by inbreeding depression
    b. Accurately estimate the strength of inbreeding depression

2. Evaluate the relative performance of pedigree- and marker-based measures of F when large numbers of genetic markers are available (e.g., hundreds to thousands of SNPs)

3. *A*. Evaluate the effects of related and inbred pedigree founders on the precision of F_P

   B. Determine if incorporating estimates of founder kinship into pedigree analyses can substantially increase the precision of F_P

4. *A*. Identify genomic regions bearing signatures of directional selection in bighorn sheep

   B. Determine if the genomic regions showing signatures of selection vary among populations with differences in elevation and recent exposure to disease

I used a combination of computer simulations and empirical whole genome sequences from bighorn sheep to address these objectives. Below I summarize the specific objectives, methods, and findings for Chapters 2-5.

Research Objectives and Findings

Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium?

Researchers often use HFCs to test for the presence of inbreeding depression. However the presence of a statistically significant HFC does not prove that inbreeding depression has occurred. Additionally, an HFC alone does not provide an estimate of the strength of inbreeding depression. Therefore, researchers are keenly interested in being able to determine whether an HFC is likely to have been caused by inbreeding depression.

Researchers often test for identity disequilibrium (ID, a non-random association of heterozygous genotypes between loci) to determine whether an HFC could have been caused by inbreeding depression. When F varies among individuals, ID exists between all pairs of loci across the genome. Estimates of ID (measured using the g_2 statistic) can also be used to along with HFCs to estimate the correlation between F and fitness components. In Chapter 2, I asked the following questions:

- How likely is ID to be detected when HFCs are caused by inbreeding depression?
- How many single nucleotide polymorphisms (SNPs) and microsatellite loci are required to reliably detect HFCs and ID simultaneously?
- How accurately can the correlation between heterozygosity and F, and the correlation between fitness components and F be estimated using HFCs and ID?

I addressed these questions using individual-based simulations of randomly mating and partially selfing populations with inbreeding depression for juvenile survival. I found that ID was not detected in a large proportion of populations with statistically significant HFCs when the variance of F was low (e.g., $\sigma^2(F) \approx 0.001$). Therefore, failure to detect ID should not be

interpreted as strong evidence that an HFC was not caused by inbreeding depression. The number of markers necessary to simultaneously detect HFC and ID depended strongly on $\sigma^2(F)$. Thus the mating system and demography of populations, which influence $\sigma^2(F)$, should be considered when designing HFC studies. I also found that the correlation between heterozygosity and F could be precisely estimated using estimates of ID and the variance of heterozygosity across a broad range of simulated scenarios. Additionally, the correlation between fitness and F could be precisely estimated when ID was strong and highly statistically significant. ID should be used in conjunction with HFCs to estimate the correlation between fitness and F, because HFCs alone reveal little about the strength of inbreeding depression. Chapter 2 is published *Molecular Ecology Resources* (Kardos *et al*., 2013), coauthored by Fred Allendorf and Gordon Luikart.

Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees

Inbreeding (mating between relatives) can dramatically reduce the fitness of offspring by causing a fraction of the genome to be identical by descent. Thus, measuring individual inbreeding is a crucial part of many studies in ecology, evolution, and conservation biology. The classical and most commonly preferred measure of individual inbreeding is the pedigree inbreeding coefficient (F_P). However, F_P could be an imprecise measure of the proportion of the genome that is identical by descent (F) due to physical linkage and a limited number of recombination during meiosis. I addressed the following question in Chapter 3:

- Which of three common measures of individual inbreeding best predicts the F in small populations: 1) F_P; 2) the excess of individual homozygosity relative to Hardy-Weinberg expected homozygosity (F_H); and 3) the proportion of the genome inferred to be in long runs of homozygosity (F_{ROH}).

I wrote an individual-based simulation model which accounts for physical linkage and recombination to address this question. I found that F was more strongly correlated with F_H and F_{ROH} than with F_P across a broad range of simulated scenarios when thousands of loci were used. This result demonstrates that F can be more precisely predicted with genetic markers than with pedigrees. Considering the imprecision of F_P, and the great difficulty associated with obtaining reliable pedigrees, researchers should soon adopt genomic measures of F as the necessary resources quickly become available. Chapter 3 is in review for publication at *Molecular Ecology Resources*, coauthored by Gordon Luikart and Fred Allendorf.

Cryptic pedigree founder relationships reduce the power to detect inbreeding depression: can genetic markers help?

F_P is often considered the most reliable measure of F. However, F_P assumes that the pedigree founders are unrelated and non-inbred. Neither of these assumptions can hold in any real population. A critical need for our understanding of inbreeding and its effects on fitness is to

determine how related and inbred pedigree founders affect estimates of inbreeding the power to detect inbreeding depression. I addressed the following questions in Chapter 4:

- When do inbred and/or related pedigree founders substantially reduce the precision of F_P and the power to detect inbreeding depression?
- Do genetic marker-based estimates of founder relatedness substantially improve the precision of F_P and the power to detect inbreeding depression?

I found that F_P was an imprecise measure of F, and the power to detect inbreeding depression was severely reduced when a large proportion of pedigree founders were closely related. I also found that incorporating marker-based estimates of founder relatedness into pedigree analyses only marginally increased the precision of F_P. However, accounting for founder kinship substantially improved the power to detect inbreeding depression when pedigrees included few generations (e.g., < 6 generations). Unfortunately, the power to detect inbreeding depression was still low (< 0.7) after accounting for founder kinship in shallow pedigrees. These results suggest that F_P, which has been considered the most reliable measure of individual inbreeding, is poor measure of F when founders are closely related. Additionally, this study suggests that marker-based measures of founder relatedness will not dramatically increase the imprecision of F_P when founders are related. Chapter 4 will be submitted to *Molecular Ecology Resources*, coauthored by Gordon Luikart and Fred Allendorf.

Whole genome sequencing identifies candidate adaptive loci in wild bighorn sheep

A major goal in evolutionary biology is to understand the genetic basis of phenotypic and fitness variation in natural populations (Stinchcombe & Hoekstra 2007; Ellegren & Sheldon 2008; Barrett & Hoeskstra 2011). However, identifying candidate adaptive genes or genomic regions can be challenging in species lacking whole genome reference assemblies. Additionally, it can be difficult to statistically differentiate the genomic effects of selection from similar patterns caused by genetic drift (due to small N_e). In this chapter, I addressed the following questions:

- What genomic regions bear signatures of directional selection in bighorn sheep?
- Do the genomic regions showing signatures of selection vary among populations with differences in elevation and recent exposure to disease?

My collaborators and I aligned whole genome sequences of pooled bighorn DNA to the domestic sheep whole genome reference assembly. We then scanned the bighorn genome for genomic regions bearing signatures of directional selection within pools (low heterozygosity) and strong genetic differentiation among pools (high F_{ST}) relative the genome-wide average. We identified several candidate adaptive genomic regions in bighorn. In particular, we found evidence for a selective sweep over a gene known to control horn development in domestic sheep (relaxin/insulin-like family peptide receptor 2 gene). We also identified selective sweep signatures over two genes involved with growth (the growth hormone receptor, and the insulin-like growth factor II receptor). The genomic region around that growth hormone receptor was also substantially genetically differentiated among populations occupying different elevation winter ranges – suggesting that this gene may play an important role in adaptation to life at high

elevation in bighorn sheep. We showed that false-positive selection signatures were possible in study by repeating our analyses on simulated data of populations with similar demographic histories to our study populations. Thus the fitness effects of the candidate genes identified here should be verified in future studies of the genomic basis of horn and body growth and adaptation in other populations of mountain sheep.

Chapter 5 will be submitted to *PLoS Genetics* and has several coauthors. James Kijas helped to design the study along with Gordon Luikart and me. James Kijas, Gordon Luikart, and Fred Allendorf helped to write the manuscript. Rowan bunch conducted whole genome sequencing of the DNA pools. Sean McWilliam aligned the sequence reads to the domestic sheep whole genome reference assembly and helped with variant calling. Sarah Dewey, John Stephenson, Jack Hogg, and Hank Edwards contributed tissue samples and expertise on bighorn spatial distribution and behavior.

Synthesis and Significance

Advancing our understanding of the genetic basis of fitness and adaptation has until recently been hindered by the limited availability large amounts of molecular genetic data. Fortunately, we are currently in the midst of a 'genomics revolution'. The availability of massive amounts of molecular genetic information is increasing at an almost alarming rate. For example, thousands of SNPs can now be genotyped for any organism using new sequencing technologies (Davey *et al*., 2011). However, we still do not know exactly how dramatically huge amounts of molecular genetic information will advance our understanding of evolutionary processes beyond what was possible with, say, a few dozen microsatellite loci. Additionally, the software and statistical tools to handle large amounts of genomic data are still in their infancy.

The results of my dissertation help to determine how F and inbreeding depression should be measured in this age when large numbers of genetic markers can be typed for any organism. Many studies have used tests for ID to determine whether HFC are caused by inbreeding depression. However, ID can also be used in conjunction with HFCs to estimate the strength of inbreeding depression (the correlation between F and fitness). Unfortunately, only rarely have HFCs been used in conjunction with estimates of ID to estimate the strength of inbreeding depression (Szulkin *et al*. 2010). Chapter 2 shows that the strength of inbreeding depression can be measured precisely in the populations that are most important to our understanding of the strength and genetic basis of inbreeding depression (i.e., populations with high variance in F). However, my results also suggest studies which detect HFCs in populations with low variance in F will often fail to detect ID. Therefore, the failure to detect ID should not be taken as evidence that an HFC was not caused by inbreeding depression.

Pedigrees have historically been viewed as being vastly superior to marker based measures of F (e.g., Pemberton 2004, 2008). Indeed, F_P has often been used as a standard against which to evaluate the precision of marker-based F measures (e.g., Slate *et al*. 2004, Balloux *et al*. 2004; Ellegren *et al*. 1999). Until recently, the variance in the realized F among individuals with the same pedigree has rarely been considered in studies of inbreeding depression (but see Forstmeier *et al*. 2012). This is somewhat surprising because decades-old population genetics theory shows that F is not well predicted by the pedigree relationship of parents (Franklin 1977;

Stam 1980). In Chapter 3, my simulations of small populations with realistic genomic characteristics (e.g., recombination and linkage) show that the 'old ways' of measuring individual inbreeding with pedigrees (Pemberton 2004) are no longer the best. For example, F can now almost always be more precisely measured with thousands of molecular markers than with pedigrees. Thus, future studies of inbreeding depression should make every effort to use large numbers of genetic markers. Marker-based measures of F will advance our understanding of the frequency of inbreeding depression, its genetic basis, and its consequences for conservation in natural populations.

The poor precision of F_P shown in Chapter 3 has two potential causes including 1) linkage and a limited number of recombinations during meiosis; and 2) related and inbred pedigree founders. Chapter 4 builds on the results from Chapter 3 by evaluating the effects of related and inbred pedigree founders on the accuracy of F_P. The results of Chapter 4 show that related pedigree founders can substantially weaken the precision of F_P, particularly when a large proportion of founders are closely related (e.g., in partially isolated small populations). A potential solution for the imprecision of F_P is to incorporate marker-based estimates of pedigree founder kinship into pedigree analyses. Unfortunately, I found that doing so only marginally increases the precision of F_P and the power to detect inbreeding depression. Thus, Chapter 4 further supports the idea that F and inbreeding depression can be most precisely estimated purely with molecular markers, rather than with F_P.

In Chapter 5, I show that genomic resources from model or agricultural species can be very useful for identifying candidate adaptive genomic regions in closely related non-model species in the wild. I identify strong selection signatures at genes that are related to traits known to affect fitness in bighorn horn (horn and body growth). Therefore, these genes should be considered as strong candidate genes in future studies of the genetic basis of fitness, horn morphology, and body growth in mountain ungulates. Our simulation results show that separating the genomic signatures of selection and genetic drift is difficult in small, isolated natural populations. Thus, this study illustrates the importance of considering realistic effective populations sizes when evaluating the possibility of false positive selection signatures.

Dissertation Format

The following chapters are the result of a collaborative efforts. Therefore, I use the collective term "we" to reflect the substantial contributions of my coauthors. Chapter formats vary according to specific journal editorial requirements. Appendices are intended for publication alongside the main body of each paper. Supplements are intended for publication on the worldwide web only.

CHAPTER 2: Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium?

Abstract

Heterozygosity-fitness correlations (HFCs) have been observed for several decades, but their causes are often elusive. Tests for identity disequilibrium (ID, correlated heterozygosity between loci) are commonly used to determine if inbreeding depression is a possible cause of HFCs. We used computer simulations to determine how often ID is detected when HFCs are caused by inbreeding depression. We also used ID in conjunction with HFCs to estimate the proportion of variation (r^2) in fitness explained by the individual inbreeding coefficient (F). ID was not detected in a large proportion of populations with statistically significant HFCs unless the variance of F was high ($\sigma^2(F)\geq 0.005$) or many loci were used (100 microsatellites or 1000 SNPs). For example, with 25 microsatellites, ID was not detected in 49% of populations when HFCs were caused by six lethal equivalents and $\sigma^2(F)$ was typical of vertebrate populations ($\sigma^2(F)\approx 0.002$). Estimates of r^2 between survival and F based on ID and HFCs were imprecise unless ID was strong and highly statistically significant ($P\approx 0.01$). These results suggest that failing to detect ID in HFC studies should not be taken as evidence that inbreeding depression is absent. The number of markers necessary to simultaneously detect HFC and ID depends strongly on $\sigma^2(F)$. Thus the mating system and demography of populations, which influence $\sigma^2(F)$, should be considered when designing HFC studies. ID should be used in conjunction with HFCs to estimate the correlation between fitness and F, because HFCs alone reveal little about the strength of inbreeding depression.

Introduction

Correlations between fitness-related traits and individual heterozygosity (so called heterozygosity-fitness correlations, HFCs) have been observed for several decades in many species (Chapman *et al.* 2009). Despite the frequent detection of HFCs, surprisingly little is known about the relative importance of locus-specific versus genome-wide heterozygosity (i.e. individual inbreeding) as mechanisms causing HFCs. Therefore an important goal of studies that detect HFCs is to identify the most likely underlying mechanism(s) causing the observed correlations.

There are three main hypotheses to explain the occurrence of HFCs. The hypothesis that HFC is caused by inbreeding depression is referred to as the 'general effect' hypothesis (David 1997; David 1998; Slate *et al*. 2004). Individuals with related parents have lower heterozygosity because many loci are 'identical by descent' (IBD, derived from a single gene copy in a common ancestor of the parents). Under the general effect hypothesis, heterozygosity is informative of the proportion of the genome that is identical by descent (F) in the presence of inbreeding depression. The reduced fitness of individuals that are IBD at loci with heterozygous advantage or deleterious recessive alleles causes a correlation between heterozygosity and the affected fitness component(s).

The 'local effect' hypothesis suggests that HFCs are caused by linkage disequilibrium between genotyped markers and nearby loci displaying overdominance or carrying deleterious recessive alleles (Hill & Robertson 1968; Ohta 1971; David 1998; Chapman *et al*. 2009; Szulkin *et al*. 2010). The 'direct effect' hypothesis proposes that HFCs are caused by selection at the genotyped loci themselves. Testing predictions of these three hypotheses is crucial to understanding the importance of inbreeding and locus-specific genetic variation to individual fitness and population performance.

A requirement of the general effect hypothesis is that F (and thus genome-wide heterozygosity, H) varies among individuals. When F varies among individuals, heterozygosity measured in one part of the genome will be correlated with heterozygosity measured in any other part of the genome (Weir & Cockerham 1973; Slate *et al*. 2004; Szulkin *et al*. 2010). Correlated heterozygosity between loci is referred to as identity disequilibrium (ID) (Weir & Cockerham 1973; Szulkin *et al*. 2010). Tests for ID between pairs of loci throughout the genome can be used to test if F varies among individuals and thus whether an observed HFC could potentially be caused by inbreeding depression. The two most commonly used methods to detect ID are the heterozygosity-heterozygosity correlation (HHC) test and the g_2 estimator of identity disequilibrium.

The HHC test repeatedly (e.g., hundreds of times) estimates the correlation coefficient between heterozygosity at one randomly chosen half of loci with heterozygosity calculated with the other half of the loci in a data set (Balloux *et al.* 2004). A mean HHC correlation coefficient that is statistically significantly greater than zero suggests that genome-wide ID is present. The HHC approach has been widely used (e.g., Acevedo-Whitehouse *et al*. 2006; Gage *et al*. 2006; Reid *et al*. 2007; Alho *et al*. 2009) but it has recently been criticized for lacking statistical rigor and for being only periferally related to HFC theory (Szulkin *et al.* 2010). The g_2 statistic directly estimates ID as the excess of doubly heterozygous genotypes at all pairs of loci relative to a random distribution of heterozygous genotypes among individuals at each locus (David *et al*. 2007; Szulkin *et al*. 2010; Ruiz-López *et al*. 2012; Luquet *et al*. 2012). The procedure described by David *et al*. (2007) uses a randomization test to evaluate the statistical significance of g_2.

In addition to detecting variation in F, ID (as quantified by g_2) can be used in conjunction with HFCs to estimate how strongly heterozygosity and fitness components are correlated with F (Szulkin *et al.* 2010). Szulkin *et al*. (2010) derived the correlation coefficient between heterozygosity and F as

$$r(H_S, F) = \frac{\sqrt{g_2}}{\sigma\ (H_S)}$$

where H_S is standardized heterozygosity (Coltman *et al.* 1999), and $\sigma(H_S)$ is the standard deviation of H_S among individuals. The correlation between a fitness component (W) and H_S ($r(W, H_S)$) is the product of the correlation between fitness and F ($r(W, F)$) and $r(H_S, F)$:

$$r(W, H_S) = r(W, F)r(H_S, F)$$

(Szulkin *et al*. 2010). When HFC is caused solely by variation in F, $r(W, H_S)$ will be weaker than $r(W, F)$ because $r(H_S, F) < 1$. The previous equation can be rearranged to solve for $r(W, F)$. Thus, combining estimates of HFC and ID provides a potentially powerful way to quantify the effects of inbreeding on individual fitness in non-pedigreed natural populations. Most recent HFC studies have tested for ID, but researchers have rarely used ID to quantify how strongly heterozygosity and fitness components are related to F (see the Supplementary Materials in Szulkin *et al*. (2010) for examples).

Our objective was to evaluate and improve the usefulness of ID tests to detect and quantify inbreeding depression in HFC studies. We asked the following questions in order to address this objective: When HFCs are caused by inbreeding depression, how likely is ID to be detected? How many single nucleotide polymorphisms (SNPs) and microsatellite loci are required to reliably detect HFCs and ID simultaneously? How accurately can $r(H_S, F)$ and $r(W, F)$ be estimated using HFCs and ID? We addressed these questions using individual-based simulations of randomly mating and partially selfing populations with inbreeding depression for juvenile survival.

Methods

Simulations of random mating populations

We wrote an individual-based simulation model of a sexually reproducing, non-selfing species, with non-overlapping generations and random mating within populations. Our simulation model is written for the program R version 3.0.1 (R Core Team 2013). A simulation script is available in the Dryad data repository (see Data Accessibility). Population size was held constant at 60 individuals to allow inbreeding to accumulate relatively quickly. Simulations were run for 50 generations. Population founders and immigrants were non-inbred ($F = 0$) and unrelated to one another, and immigrants were unrelated to all residents. The sex ratio was 1:1, with each individual's sex determined with a random number generator. Our simulations assumed Wright's infinite island model of migration (Wright 1931).

The expected F of an individual is the pedigree inbreeding coefficient (F_P). However, the observed F can deviate substantially from F_P because of Mendelian segregation of chromosomes and recombination within chromosomes during meiosis (Franklin 1977; Hill & Weir 2011; Stam 1980). The variance of F around the expected F_P is higher among organisms with short genetic map lengths (i.e., fewer crossovers per meiosis) or with few chromosomes (Franklin 1977). We wanted our simulations to include realistic variation in F. Therefore we included linkage and recombination in our simulations. We simulated two different genomic architectures. First, we simulated 3 Gb genomes, with 20 chromosomes and a recombination rate of 1.2 cM/Mb (genetic map length = 3600 cM total). We also simulated 3 Gb genomes with 10 chromosomes and a recombination rate of 0.33 (genetic map length = 1000 cM). Our model of recombination follows Fisher's theory of junctions (Chapman & Thompson 2003; Fisher 1965). The simulations assumes no interference, and a constant recombination rate across the genome and among individuals and sexes.

In order to generate variation in $\sigma^2(F)$ among simulated scenarios, we used eight different immigration rates (m) including m = (0.004, 0.006, 0.008, 0.017, 0.033, 0.1, 0.25, 0.75), where m is the probability of an individual being an immigrant with a genotype drawn from the same allele frequency distribution as the population founders (see below). We simulated 100 independent replicate populations using each of the eight values of m. We calculated the F for each individual as the sum of the lengths of all IBD chromosome segments divided by the total genome size.

Tests for ID and HFCs were limited to individuals born during generations 49-50 in order to achieve a realistic sample size of 120 individuals, which is in the middle of the range of sample sizes used in past HFC studies (Chapman *et al.* 2009). The distribution of $\sigma^2(F)$ in generations 49-50 for each value of m is shown in Figure 2-1. The range of simulated values of $\sigma^2(F)$ (Figure 2-1) were typical of values of $\sigma^2(F_P)$ observed in studies of wild vertebrates (Grueber *et al.* 2011).

To assess the effects of the type of genetic marker used, we simulated highly variable microsatellite-like loci, and less variable diallelic SNPs. Simulations were parameterized so that microsatellite loci had mean expected heterozygosity (H_e) of 0.65 (mean number of alleles per locus was 3.7), and SNPs had H_e = 0.30 on average in the final (50^{th}) generation of the simulations. The microsatellite genotypes of population founders and immigrants were determined by randomly assigning two alleles selected from a pre-determined allele frequency distribution. Founder and immigrant SNP genotypes were determined by randomly choosing one of two alleles at each locus with a pre-determined minor allele frequency. Details of how founder and immigrant allele frequencies were determined are described in Appendix 1. The genotypes of all other subsequent individuals (descendants) were determined by random mating and Mendelian inheritance of chromosomes. We calculated H_S for each individual using 100, 250, 500, and 1000 SNPs, and 10, 25, 50 and 100 microsatellite loci.

We conducted separate simulations using different numbers of lethal equivalents to explore the importance of the strength of inbreeding depression to how useful tests of ID are to detect variation in F in HFC studies. We simulated populations with 6 and 12 diploid lethal equivalents to model inbreeding depression for survival to reproduction. All non-inbred individuals (F = 0) survived to reproduction, but we reduced the survival probability of inbred individuals using the lethal equivalents model of Morton *et al*. (1956). The model of Morton *et al*. (1956) represents a linear decline in log(fitness) with increasing F (Keller & Waller 2002). We used the Morton *et al*. (1956) model of inbreeding depression instead of simulating the loci responsible for inbreeding depression in order to hold the strength of inbreeding depression constant. Simulating the phenotypically active loci would have created uncontrolled variation in the strength of inbreeding depression among populations (i.e., simulation replicates) and through time due to genetic drift and purging of deleterious recessive alleles. We used only six lethal equivalents when simulating populations with a short genetic map length (1000 cM) and 10 chromosomes.

We compared simulation model output to results predicted by population genetic theory in order to verify that simulations were consistent with theoretical expectations. We compared the theoretical expectations and the observed values of the relationship between heterozygosity

and F, the mean F_P at migration-drift equilibrium, and the relationship between F and F_P. Simulation model output agreed closely with theoretical expectations (Figures 2-S1 – 2-S3).

Simulations of partially selfing populations

To simulate partially selfing populations, we changed the simulation model described above so that all individuals were self-compatible hermaphrodites. We simulated 100 replicate closed populations containing 60 individuals using each of six selfing rates (S = 0.0, 0.04, 0.08, 0.12, 0.16, 0.2), where each individual had a self fertilization probability of $S + 1/N_s$ (where N_s is the number of individuals that survived to reproduce). The range of simulated values of S are typical of herbaceous and woody perennial plants (Barrett *et al.* 1996). Simulations of partially selfing populations were run for 15 generations, but we estimated HFCs and ID only among individuals in the last two generations of the simulations. All other aspects of the simulations of partially selfing populations were identical to the simulations of random mating populations.

g_2 and HHC test procedures

We implemented the g_2 estimator of ID as described by David *et al*. (2007) in a script written for the program R. The P-value for tests of the null hypothesis $g_2 = 0$ was determined using the randomization procedure described by David *et al.* (2007). We also wrote an R script implementing the HHC method described by Balloux *et al.* (2004). For each simulated population, the loci were randomly split into two groups. The proportion of heterozygous genotypes (multiple locus heterozygosity, or 'MLH') was calculated for each of the two groups of loci. Then the correlation coefficient between MLH at one set of markers and MLH at the other set of markers was calculated using the cor function in the program R. This was repeated 500 times for each simulated dataset. The P-value from the HHC test was defined as the proportion of the 500 repetitions with an HHC correlation coefficient ≤ 0.

Statistical analyses

We used H_s as a measure of individual heterozygosity because it is one of the most commonly used measures of heterozygosity. We used generalized linear models (GLMs) with a logit link function in the glm function (binomial error distribution) of the program R to test if survival to reproduction was significantly related to H_s in each simulated population (i.e., to test for HFC). We estimated the proportion of variance in survival explained by H_s ($\hat{r}^2(survival, H_S)$) as the proportional reduction in deviance (sum of the squared deviance residuals) in the model relative to an intercept only model (Zheng 2000). We estimated the proportion of simulated populations with statistically significant HFCs that also had statistically significant tests for ID using the prop.test function in the stats package of the program R.

Taking the square of $r(H_S, F)$ yields $r^2(H_S, F)$, a parameter describing the true proportion of variance in H_S explained by variation in F. We calculated $r^2(H_S, F)$ for each simulated population from a linear regression of H_S versus F. We then estimated $r^2(H_S, F)$ for each simulated population as would be done in an HFC study without perfect information on F:

$$\hat{r}^2(H_S, F) = \frac{\hat{g}_2}{\hat{\sigma}^2(H_S)}$$

where $\hat{g}_2$ is an estimate of the parameter g_2, and $\hat{\sigma}^2(H_S)$ is an estimate of the variance in H_S (Szulkin *et al*. 2010).

Squaring $r(survival, F)$ yields $r^2(survival, F)$, a parameter which describes the proportion of variance in survival explained by variation in F. We calculated $r^2(survival, F)$ as the proportional reduction in the deviance of a logistic regression model of survival versus F relative to an intercept only model (Zheng 2000). We then estimated $r^2(survival, F)$ as

$$\hat{r}^2(survival, F) = \frac{\hat{r}^2(survival, H_S)}{\hat{r}^2(H_S, F)}$$

(Szulkin *et al.* 2010).

We present our results on the precision and bias of $\hat{r}^2(H_S, F)$ and $\hat{r}^2(survival, H_S)$ from different simulations categorized according to $\sigma^2(F)$. We retained only simulated populations with $\sigma^2(F) > 0$. We split the remaining simulations as evenly as possible into three groups based on $\sigma^2(F)$ for simulations with random mating. Specifically, we sorted the simulations according to $\sigma^2(F)$. We then grouped simulations that were in the bottom, middle and top thirds (based on the number of simulations) with respect to $\sigma^2(F)$. In our results, we label the simulations categorized according to $\sigma^2(F)$ with the mean $\sigma^2(F)$ for each category. For simulated partial selfing populations, we split all of the simulations as evenly as possible (as with simulations of random mating populations) into four groups based on $\sigma^2(F)$.

We evaluated the precision of $\hat{r}^2(H_S, F)$ by comparing its standard deviation to the standard deviation of $r^2(H_S, F)$ for each combination of number and type of loci and $\sigma^2(F)$. A standard deviation of $\hat{r}^2(H_S, F)$ that is substantially higher than the standard deviation of $r^2(H_S, F)$ would suggest that $\hat{r}^2(H_S, F)$ was imprecise. We used *F*-tests to evaluate the statistical significance of the difference between standard deviations of $\hat{r}^2(H_S, F)$ and $r^2(H_S, F)$. We used the same approach to test if the standard deviations of $\hat{r}^2(survival, F)$ and $r^2(survival, F)$ were statistically significantly different.

To evaluate the bias of $\hat{r}^2(H_S, F)$, we calculated the difference between the mean $\hat{r}^2(H_S, F)$ and the mean $r^2(H_S, F)$ for each combination of number and type of loci and $\sigma^2(F)$. We used the same approach to evaluate the bias of $\hat{r}^2(survival, F)$. We used a randomization procedure (Manly 2007) to test for a statistically significant difference between the mean

$r^2(survival, F)$ and mean $\hat{r}^2(survival, F)$ for simulations in each category of $\sigma^2(F)$. We used the same randomization procedure to test if the mean $\hat{r}^2(H_S, F)$ and $r^2(H_S, F)$ were significantly different for simulations in each category of $\sigma^2(F)$. The randomization test procedure is described in detail in the Supplementary Materials.

Results

Here we focus on the results from simulations with random mating and 6 diploid lethal equivalents only, 20 chromosomes, and a genetic map length of 3600 cM. The conclusions reached from analyses of other simulation scenarios are qualitatively similar, regardless of the mating system, strength of inbreeding depression, or genomic characteristics. The results from other simulation scenarios are provided in detail in the Supplementary Materials.

The results of tests for ID using the $\hat{g}_2$ and HHC procedures were equivalent in more than 90% of simulated populations. Having found that the $\hat{g}_2$ and HHC tests almost always provide the same results, we only present results from analyses using $\hat{g}_2$, as unlike HHC, $\hat{g}_2$ allows estimation $r^2(H_S, F)$ and $r^2(survival, F)$ (Szulkin *et al*. 2010).

How often is ID detected in populations with HFCs caused by inbreeding depression?

The probability of detecting ID given that an HFC was detected depended strongly on both the number and type of loci used, and $\sigma^2(F)$ (Figure 2-2). Approximately 4-8 times more SNPs than microsatellite loci were necessary to achieve a given probability of detecting ID in populations with detected HFCs. $\sigma^2(F) \geq 0.005$ was necessary for the proportion of populations with HFCs with significant tests for ID to be above 0.8 for 25, 50, or 100 microsatellite loci. This proportion never exceeded 0.5 when 10 microsatellite loci were used. $\sigma^2(F) \geq 0.002$ was necessary for the proportion of HFCs with significant ID to exceed 80% for 500 and 1000 SNPs. $\sigma^2(F) = 0.005$ was necessary for 0.8 of simulations with detected HFCs to have significant tests for ID when 250 SNPs were used. This proportion never exceeded approximately 0.75 when only 100 SNPs were used. The statistical power of HFC and ID tests (independent from one another) is described in the Supplemental Materials.

Performance of methods to estimate $r^2(H_s, F)$ and $r^2(survival, F)$

The mean $\hat{r}^2(H_S, F)$ and $r^2(H_S, F)$ were statistically significantly different in 5 out of 24 combinations of the number and type of marker and $\sigma^2(F)$ ($P \leq 0.05$, randomization tests) (Figures 2-3, 2-S5). The mean $\hat{r}^2(H_S, F)$ was slightly lower than the mean $r^2(H_S, F)$ in 17 out of 24 combinations of the number and type of marker and $\sigma^2(F)$. The differences between the mean $\hat{r}^2(H_S, F)$and $r^2(H_S, F)$ were small in all cases, ranging from 0.001 with 250 SNPs and $\sigma^2(F) = 0.005$, to 0.06 with 1000 SNPs and $\sigma^2(F) = 0.001$.

The standard deviation of $\hat{r}^2(H_S, F)$ was statistically significantly higher than the standard deviation of $r^2(H_S, F)$ in 15 out of 24 combinations of the number and type of marker and $\sigma^2(F)$ ($P \leq 0.05$, F-tests) (Figures 2-3, 2-S5). This difference ranged from 0.009 when 1000 SNPs were used and $\sigma^2(F)$ was 0.005, to 0.06 when 1000 SNPs were used and $\sigma^2(F)$ was 0.001 (Figure 2-S5).

The mean $\hat{r}^2(survival, F)$ was always slightly lower than the mean $r^2(survival, F)$ when $\sigma^2(F)$ was 0.002 or greater (Figure 2-4). Conversely, $\hat{r}^2(survival, F)$ was almost always slightly higher than $r^2(survival, F)$ (in 7 out of 8 cases) when $\sigma^2(F)$ was 0.001. The mean $\hat{r}^2(survival, F)$ and $r^2(survival, F)$ were statistically significantly different only when 1000 SNPs were used and $\sigma^2(F)$ was 0.001 (Figure 2-S6). Differences between the mean $\hat{r}^2(survival, F)$ and $r^2(survival, F)$ ranged from 0.007 with 50 microsatellite loci and $\sigma^2(F) = 0.002$, to 0.07 with 1000 SNPs and $\sigma^2(F) = 0.001$.

The standard deviation of $\hat{r}^2(survival, F)$ was statistically significantly larger than the standard deviation of $r^2(survival, F)$ in 17 out of 24 combinations of the number and type of marker and $\sigma^2(F)$. The difference between the standard deviations of $\hat{r}^2(survival, F)$ and $r^2(survival, F)$ was always substantially smaller when $\sigma^2(F)$ was 0.005 than when $\sigma^2(F)$ was 0.001 (Figures 2-4 and 2-S6). For example, the difference between the standard deviations of $\hat{r}^2(survival, F)$ and $r^2(survival, F)$ was 0.1 when $\sigma^2(F)$ was 0.001 and 500 SNPs were used (Figure 2-4). However, when $\sigma^2(F)$ was 0.005 and 500 SNPs were used, the difference between the standard deviations of $\hat{r}^2(survival, F)$ and $r^2(survival, F)$ was 0.0002.

Considering that difference between the standard deviations of $\hat{r}^2(survival, F)$ and $r^2(survival, F)$ depended strongly on $\sigma^2(F)$, estimates of $\hat{g}_2$ may be helpful to evaluate the reliability of a particular estimate of $\hat{r}^2(survival, F)$. We assessed this with linear regression models of the running standard deviation of the error of $\hat{r}^2(survival, F)$ ($\hat{r}^2(survival, F) - r^2(survival, F)$) versus $\hat{g}_2$ and the P-values from $\hat{g}_2$ tests. The running standard deviation of $\hat{r}^2(survival, F)$ was strongly associated with $\hat{g}_2$ ($P < 0.001$, $r^2 = 0.35$) and its P-value ($P < 0.001$, $r^2 = 0.66$). 95% of the estimation errors of $\hat{r}^2(survival, F)$ were ≤ 0.2 units from one another when when $\hat{g}_2$ was $\geq$ approximately 0.002 and when P-values from g_2 tests were ≤ 0.06 (Figure 2-5).

Discussion

We used simulations to evaluate the effectiveness of identity disequilibrium (ID) tests to help assess the role that inbreeding depression plays in causing HFCs. We often detected HFCs in populations where ID was not detected when $\sigma^2(F)$ was relatively low. ID and HFC exist in all populations where F varies and inbreeding depression occurs. However our results show that

tests for HFC and tests for ID will often fail to pass the statistical significance threshold simultaneously. This result suggests that studies will commonly detect HFCs that are caused by inbreeding depression but fail to detect ID, unless $\sigma^2(F)$ is high (e.g., in small recently admixed or partially selfing populations) or very large numbers of loci are used. Therefore, failure to detect ID should not be taken as strong evidence that inbreeding depression did not cause an HFC.

Accuracy and precision of effect size estimates

Previous studies (Slate *et al*. 2004; Szulkin *et al*. 2010) compared the correlation between H_S and F_P determined from pedigrees ($\hat{r}(H_S, F_P)$) to the correlation between H_S and F ($\hat{r}(H_S, F)$) predicted based on population genetic theory similar to that employed in our study. Our study builds on these previous studies by evaluating the precision and bias of $\hat{r}^2(H_S, F)$ and $\hat{r}^2(survival, F)$ using simulated data with error-free knowledge of the F.

Our simulations showed that $\hat{r}^2(H_S, F)$ was essentially unbiased. There were several simulated scenarios with statistically significant differences between the mean $\hat{r}^2(H_S, F)$ and the mean $r^2(H_S, F)$ (Figure 2-S5), but we believe the magnitude of the differences were too small to be meaningful in practice. This finding is in contrast to previous results (Slate *et al*. 2004; Szulkin *et al*. 2010) based on empirical data from intensively studied populations of mammals and birds. Despite $\hat{r}(H_S, F)$ being significantly correlated with $\hat{r}(H_S, F_P)$ in Slate *et al*. (2004) and Szulkin *et al*. (2010), $\hat{r}(H_S, F_P)$ was often weaker than $\hat{r}(H_S, F)$ on average.

This difference between the results of our study and of Slate *et al*. (2004) is not surprising. Using simulations, we were able to compare $\hat{r}^2(H_S, F)$ to $r^2(H_S, F)$ where F is the error-free proportion of the genome that is identical by descent. Because they used empirical data from real populations, Slate *et al*. (2004) and Szulkin *et al*. (2010) were only able to compare $\hat{r}(H_S, F)$ to $\hat{r}(H_S, F_P)$, where in F_P is an imperfect pedigree-based estimator of the F due only to the known (recent) common ancestors of parents. Errors in pedigree construction, related or inbred pedigree founders (Slate *et al*. 2004), and variation in the proportion of the genome that is IBD among individuals with the same F_P (Franklin 1977; Hill & Weir 2011) is expected to cause $\hat{r}(H_S, F_P)$ to be weaker than $r(H_S, F)$ on average. These results demonstrate that the error prone nature of F_P should be considered when using pedigrees to evaluate the performance of marker-based measures of inbreeding and inbreeding depression.

The standard deviation of $\hat{r}^2(H_S, F)$ was only slightly higher than the standard deviation of $r^2(H_S, F)$ on average among simulation repetitions across all scenarios. This suggests that $\hat{r}^2(H_S, F)$ can be estimated precisely in a wide range of scenarios. Therefore $\hat{g}_2$ appears to be a very useful tool to determine how strongly heterozygosity is related to individual inbreeding across a broad range of scenarios. Thus, we believe that estimating $\hat{r}^2(H_S, F)$ should be standard practice in HFC studies.

We found that $\hat{r}^2(survival, F)$ tended to be only slightly biased in all simulated scenarios. $\hat{r}^2(survival, F)$ was upwardly biased when $\sigma^2(F)$ was 0.001, and downwardly biased when $\sigma^2(F)$ was 0.002 or higher. We believe that the magnitude of this bias is so small such that it is not likely to be important in practice (Figures 2-4, 2-S6). Despite having low bias, $\hat{r}^2(survival, F)$ was imprecise (i.e., high standard deviation among replicates relative to $r^2(survival, F)$) when $\sigma^2(F)$ was low. This suggests that HFC-based estimates of the strength of inbreeding depression will be unreliable in many study populations. Fortunately, we found that both the magnitude and statistical significance of $\hat{g}_2$ are good predictors of the error of $\hat{r}^2(survival, F)$ (Figure 2-5). We suggest that $\hat{r}^2(survival, F)$ should have acceptably low sampling error when $\hat{g}_2$ is greater than approximately 0.005 and highly statistically significant ($P \approx 0.01$).

Number of loci

The ability to reliably infer the effects of inbreeding depression in HFC studies depends strongly on the number of loci used and $\sigma^2(F)$. When the variance of F is very high (e.g., in partially selfing or small recently admixed populations) relatively few loci will be sufficient. For example, our simulations show that 10-25 microsatellite loci are sufficient to reliably detect both HFC and identity disequilibrium, and to precisely estimate the strength of inbreeding depression when $\sigma^2(F)$ is very high (e.g., with selfing rates of approximately 12% or greater and $\sigma^2(F) \geq 0.017$, Figures 2-S13, 2-S17). Conversely, when $\sigma^2(F)$ is relatively low ($\sigma^2(F) \approx 0.001$), as in some random mating populations, even very large numbers of markers (100 microsatellite loci or 1000 SNPs) are often insufficient to reliably detect ID when HFCs are detected. The recent historical demography and mating systems of study populations (when known) may provide some clues regarding whether $\sigma^2(F)$ is likely to be high or low, and thus whether large numbers of markers are necessary to achieve high power to detect HFC and ID.

The necessary number of loci to detect both identity disequilibrium and HFC should vary with the number of sampled individuals in addition to $\sigma^2(F)$. The effects of sample size versus the number of loci on the detection of HFC and ID is beyond the scope of this study. We chose the sample size of 120 individuals (two generations) for HFC and ID tests because it is typical of previous HFC studies (Chapman *et al*. 2009). A given investment in increasing sample size could potentially increase the power to detect both ID and HFC more than a similar investment in increasing the number of loci. An interesting avenue of research for the future would be to evaluate the relative influence of sample size versus number of markers on the power to detect ID and HFC.

When is $\sigma^2(F)$ expected to be high?

The $\sigma^2(F)$ of a particular population is affected by the mating system, the local N_e (genetic drift), and gene flow from other populations. Our simulations covered a limited range of demographic and mating system scenarios so it is worth exploring how these factors are expected to affect $\sigma^2(F)$. Partially selfing populations (e.g., of some plant species) probably have some of the

highest $\sigma^2(F)$ in natural populations due to a mixture of the highly inbred offspring resulting from self fertilization ($F_P \geq 0.5$) and the relatively non-inbred offspring of randomly mated parents. Populations with small N_e (strong drift) and occasional pulses of immigrants should also produce high $\sigma^2(F)$ (Figure 2-1). The mean F of residents in populations with small N_e should be high due to strong genetic drift, and offspring from matings between residents and unrelated immigrants will be non-inbred ($F \approx 0$), creating a mixture of highly inbred and relatively outbred individuals.

Grueber *et al*. (2011) compiled estimates of $\sigma^2(F_P)$ from empirical studies of vertebrates. The highest observed values of $\sigma^2(F_P)$ were from very small, highly inbred populations. $\sigma^2(F)$ was 0.015 in a population of large ground finches (*Geospiza magnirostris*) that was founded by two individuals and consisted of a mixture of highly inbred residents ($F_P \geq 0.25$) and relatively non-inbred offspring of immigrants (Grant *et al*. 2001). A captive population of grey wolves (*Canis lupus*) founded with only eight individuals (Hedrick *et al*. 2001) had $\sigma^2(F_P)$=0.019. These examples illustrate that $\sigma^2(F)$ or $\sigma^2(F_P)$ can be very high in non-selfing populations with strong local genetic drift followed by admixture or gene flow. Bottleneck and assignment tests and estimates of N_e might together be applied to help identify populations that experienced strong local drift and recent admixture (Rannala & Mountain 1997; Luikart *et al*., 1998, 2010; Pritchard *et al*. 2000).

Conclusions

1. Studies that detect HFCs caused by inbreeding depression will commonly fail to detect ID, which is often tested for to detect variation in individual inbreeding (and thus inbreeding depression as a cause of HFC). This results from tests for ID and HFC failing to simultaneously cross the statistical significance threshold. Therefore, failure to detect ID should not be interpreted as strong evidence that an HFC was not caused by inbreeding depression.

2. We suggest that using ID to estimate the association between heterozygosity and F should be standard practice in HFC studies. This procedure performs well across a wide range of scenarios, and reliably elucidates how informative heterozygosity is of F. Researchers can easily estimate g_2 with the RMES program (David *et al*. 2007) available at http://www.cefe.cnrs.fr/en/genetique-et-ecologie-evolutive/patrice-david .

3. We suggest focusing on the $\hat{r}^2$ between fitness components and F, while considering the the magnitude and statistical significance of $\hat{g}_2$. A combination of relatively high $\hat{g}_2$ (e.g., $\hat{g}_2 \geq 0.005$) and high statistical significance (e.g., $P \approx 0.01$) strongly suggests that the strength of inbreeding depression can be precisely estimated. HFCs alone (even with associated positive tests for identity disequilibrium) provide little information about how strongly inbreeding affects fitness components.

4. The number of loci needed to reliably detect an HFC and ID depends strongly on $\sigma^2(F)$ and thus the mating system and recent demography (e.g., bottleneck and admixture history) of the study population. We believe that researchers should carefully consider these factors when designing and interpreting the results of HFC studies.

The combined use of the results and guidelines presented here, along with increasingly large genetic data sets, has enormous potential to increase our understanding of the effects of inbreeding on fitness in natural populations.

Acknowledgements

This material is based on work supported within the Montana Institute on Ecosystems by the National Science Foundation EPSCoR program grant EPS-1101342 (M. Kardos' support). M. Kardos was also supported by a Montana Ecology of Infectious Diseases NSF IGERT traineeship under Grant No. DGE-0504628, and a teaching assistantship from the University of Montana. GL received support from and NSF DEB-1067613 and DEB-0723928. We thank Patrice David, Jon Graham, L. Scott Mill helpful comments and discussions. Wolfgang Forstmeier and two anonymous reviewers also provided helpful criticisms.

Data Accessibility

Our R simulation scripts are available on the Dryad data repository (doi:10.5061/dryad.g7qs2). The Dryad DOI will not be active until the type-set version of the paper is published online. However, the R scripts can be accessed for review through Dryad here: http://datadryad.org/review?wfID=20710&token=fb87ed26-6d2e-467c-89d7-07d5eb70fd21

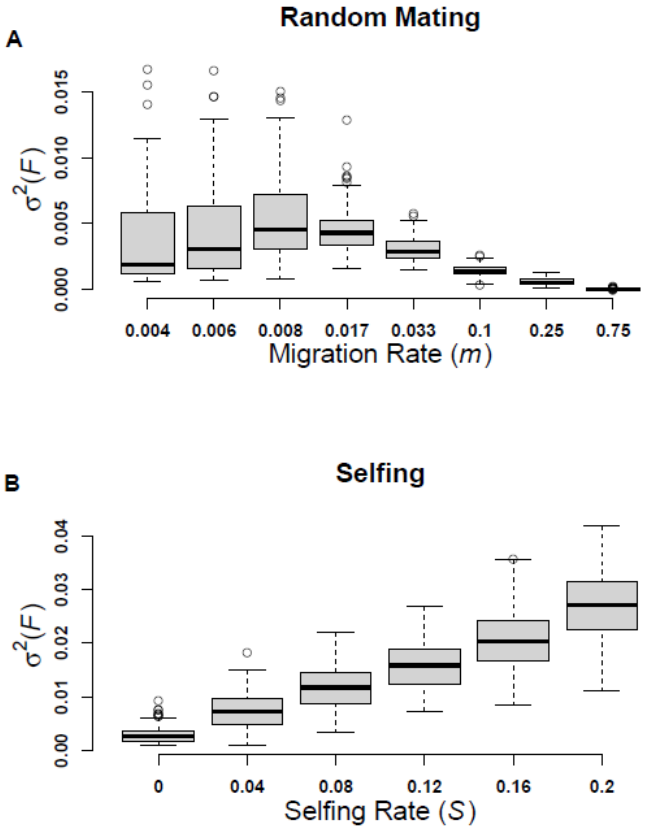

Figure 2-1. The variance of F ($\sigma^2(F)$) in simulated random mating (**A**) and partially selfing (**B**) populations. The data are shown from the last two generations of 100 simulations for each of eight migration rates (m) and six selfing rates (S).

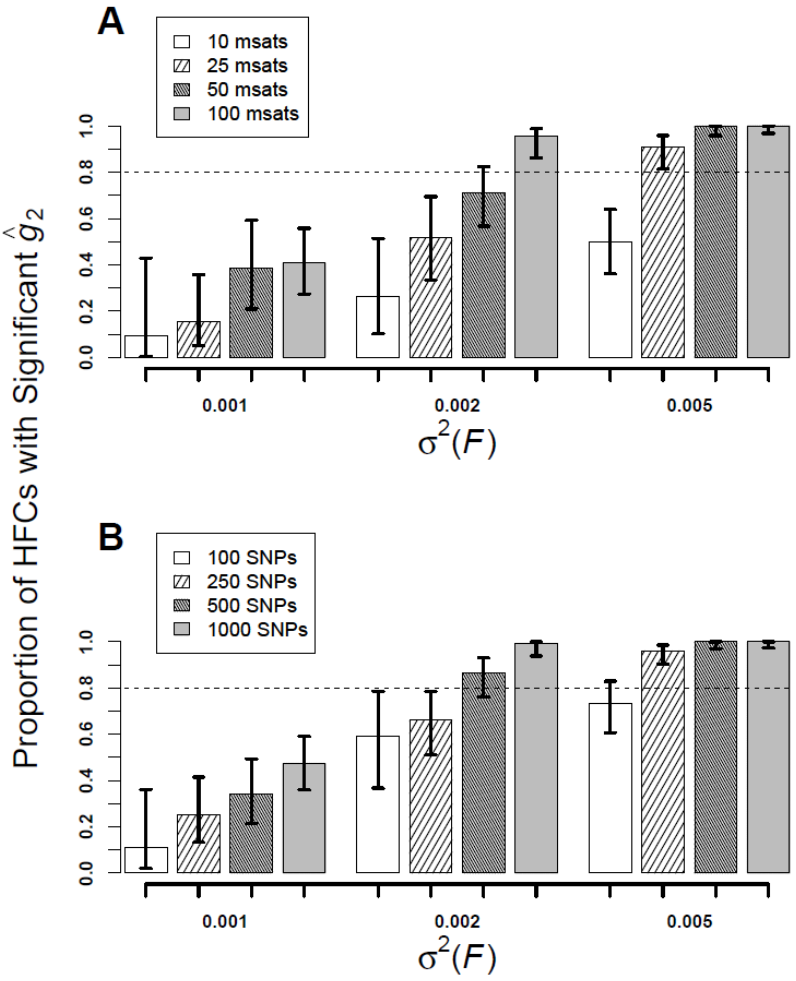

Figure 2-2. The proportion of simulations with statistically significant HFCs that also had statistically significant tests for identity disequilibrium plotted against the variance of F ($\sigma^2(F)$). Results are shown from simulations using six diploid lethal equivalents, microsatellite loci (A), and SNPs (B). Error bars are 95% confidence intervals. The dashed lines represent 80% of simulations with statistically significant HFC also having statistically significant g_2.

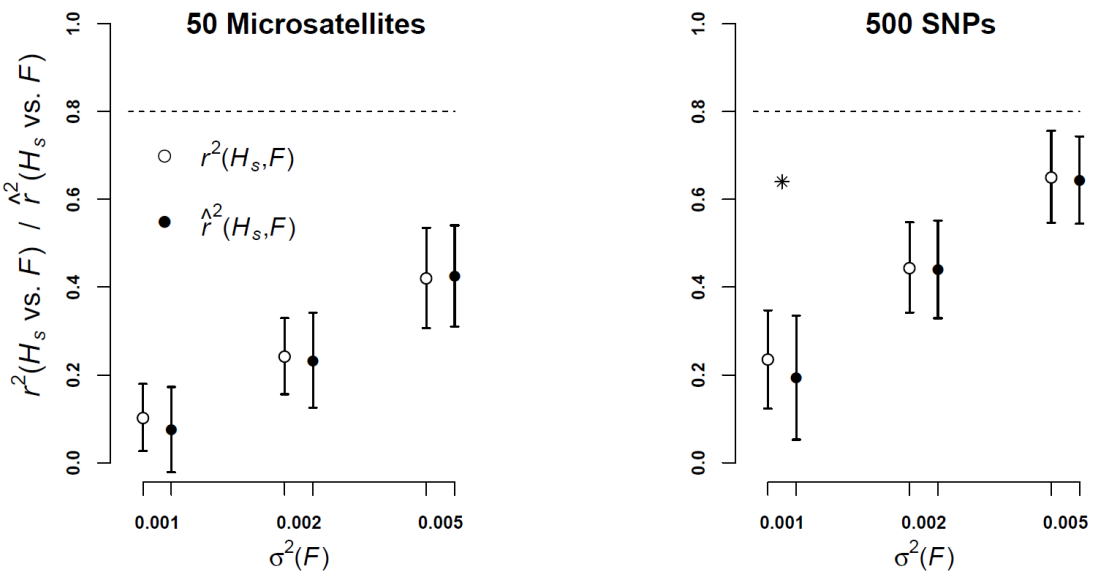

Figure 2-3. The true ($r^2(H_S,F)$) and estimated ($\hat{r}^2(H_S,F)$) proportion of variance in H_s explained by F (+/- one standard deviation) versus the variance of F ($\sigma^2(F)$). The data shown are from simulations of random mating populations with 6 diploid lethal equivalents, and using 100 SNPs (top row) and 500 SNPs (bottom row). Results are shown from simulations of genomes with 20 chromosomes and 3600 cM (left column) and 10 chromosomes and 1000 cM (right column). Filled circles represent $\hat{r}^2(H_S,F)$, and open circles represent $r^2(H_S,F)$. The dashed line represents $r^2 = 0.8$. Asterisks indicate a statistically significant difference between the mean $r^2(H_S,F)$ and the mean $\hat{r}^2(H_S,F)$. Results from analyses based on microsatellite loci and different numbers of loci and are shown in the supplementary materials.

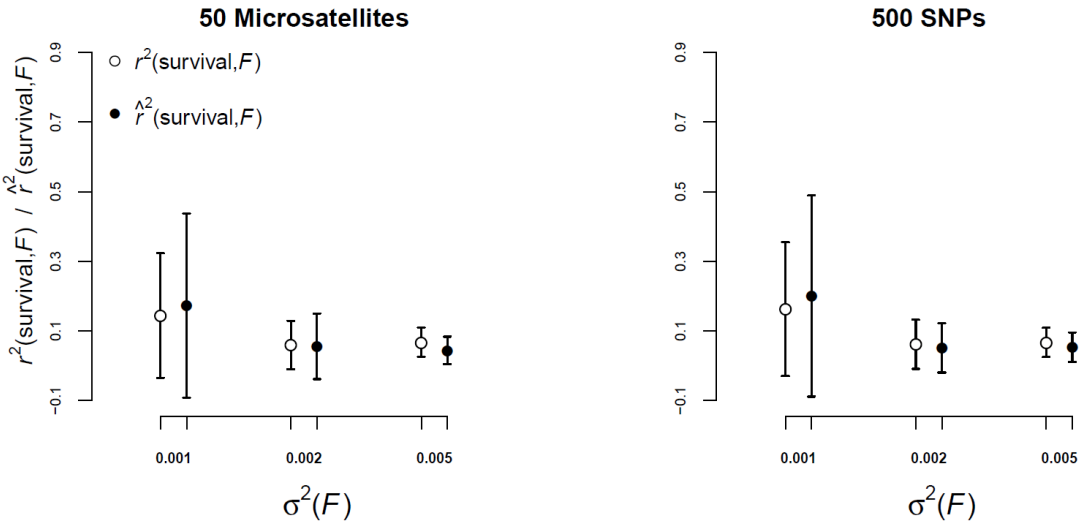

Figure 2-4. The mean true ($r^2(survival, F)$) and estimated ($\hat{r}^2(survival, F)$) proportion of variance in survival due to variation in F (+/- one standard deviation) versus the variance of F (σ^2(F)). The data shown are from simulations of random mating populations with 6 diploid lethal equivalents, and using 50 microsatellite loci and 500 SNPs. Asterisks indicate statistically significant differences between the median $r^2(survival, F)$ and $\hat{r}^2(survival, F)$. Open circles represent the true values of $r^2(survival, F)$, and closed circles represent $\hat{r}^2(survival, F)$. Results for different numbers of loci show a similar pattern, and are shown in Figure 2-S6.

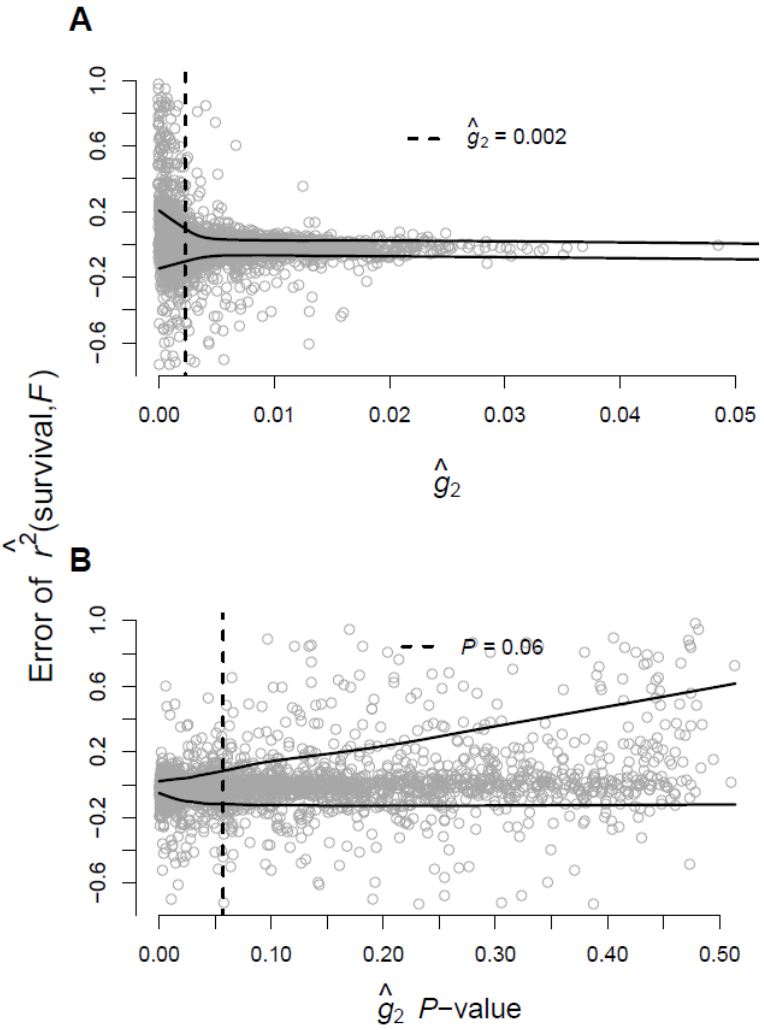

Figure 2-5. Errors of the estimated proportion of variance in fitness due to variation in F ($\hat{r}^2(survival, F)$) versus estimates of g_2 (A) and the P-values from g_2 tests for identity disequilibrium (B). Data are from all simulations with random mating and six lethal equivalents. The individual estimation errors are shown as gray points. The solid black lines are loess functions fit to the 5th and 95th running quantiles of the estimation errors, and show how the precision of $\hat{r}^2(survival, F)$ varies with g_2 (A) and its associated P-value (B). The dashed black lines represent the value of g_2 above which 95% of errors are within 0.2 units from one another (A), and the P-value below which 95% of estimation errors of $\hat{r}^2(survival, F)$ are within 0.2 units of one another (B).

CHAPTER 3: Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees

Abstract

Inbreeding (mating between relatives) can dramatically reduce the fitness of offspring by causing a fraction of the genome to be identical by descent. Thus, measuring individual inbreeding is a crucial part of many studies in ecology, evolution, and conservation biology. We used simulations to determine which of three common measures of individual inbreeding best predicts the proportion of the genome that is identical by descent (PGIBD) in small populations: 1) the pedigree inbreeding coefficient (F_P); 2) the excess of individual homozygosity relative to Hardy-Weinberg expected homozygosity (F_H); and 3) the proportion of the genome inferred to be in long runs of homozygosity (F_{ROH}). PGIBD was more strongly correlated with F_H and F_{ROH} than with F_P across a broad range of simulated scenarios when thousands of SNPs were used. For example, PGIBD was more strongly correlated with F_{ROH} (estimated with ≥35000 SNPs) and F_H (estimated with ≥5000 SNPs) than with F_P, regardless of pedigree depth, in small partially isolated populations. PGIBD was always more strongly correlated with F_H than with F_{ROH} when 25K or fewer SNPs were used. However, the correlation between PGIBD and F_{ROH} was at least as strong as the correlation between PGIBD and F_H when 35K or more SNPs were used. Our results suggest that PGIBD can be more precisely predicted with genetic markers than with pedigrees. Considering the imprecision of F_P, and the great difficulty associated with obtaining reliable pedigrees, we encourage researchers to soon adopt genomic measures of PGIBD as the necessary resources quickly become available.

Introduction

Biologists have long recognized that inbred individuals (those whose parents are closely related) often have lower fitness than the offspring of unrelated parents (Darwin 1868). The cumulative effects of inbreeding on individual fitness can reduce the population growth rate and the probability of persistence (Madsen *et al.* 1999; O'Grady *et al.* 2006; Saccheri *et al.* 1998; Westemeier *et al.* 1998). Consequently, measuring individual inbreeding is a crucial part of many studies in ecology, evolution, and conservation biology.

Inbred individuals have lower genome-wide heterozygosity because a fraction of loci are 'identical by descent'. A locus is identical by descent if it carries two gene copies that both originated from a single copy in a common ancestor of the parents. All measures of individual inbreeding seek to predict the proportion of the genome that is identical by descent (PGIBD). Unfortunately, it is has historically been very difficult to precisely measure PGIBD using either pedigrees or genetic markers.

The classical measure of PGIBD is the pedigree inbreeding coefficient F_P (Keller & Waller 2002; Wright 1922). F_P predicts PGIBD due to the known common ancestors of parents. F_P has often been considered the best measure of individual inbreeding (Pemberton 2004; Pemberton 2008). However, pedigrees are often impractical because they require multiple generations of accurate parentage assignments. Additionally, some classical (Franklin 1977) and

recent research (Forstmeier *et al.* 2012; Hill & Weir 2011) cast doubt on the superiority of F_P over marker-based approaches.

PGIBD can very substantially among individuals with the same pedigree (Franklin 1977; Hill& Weir 2011). The standard deviation of PGIBD (σ_{IBD}) around the expected value of F_P is higher among organisms with fewer chromosomes or a shorter genetic map length (Franklin 1977; Hill & Weir 2011). For example, consider the offspring produced from one generation of selfing (Figures 3-1 & 3-S1). All of the offspring have $F_P = 0.5$. When there is one chromosome without recombination, each gamete is derived from a single chromosome copy in the parent. In this case, half of the offspring have PGIBD = 0, the other half have PGIBD = 1, and σ_{IBD} among the offspring is 0.5. When the chromosome has a genetic map length is 100 cM (one recombination per meiosis on average), σ_{IBD} among the offspring is 0.31 (Franklin 1977). If we split the genome into 20 equally sized chromosomes and hold genetic map length at 100 cM, the expected σ_{IBD} among the offspring is then reduced to 0.096 (Franklin 1977). This example illustrates that F_P can be an imprecise measure of PGIBD, particularly for organisms with few chromosomes or a short genetic map. Additional imprecision of F_P can result from errors in pedigrees (e.g., due to false parentage assignments and missing individuals) or violating the assumptions of unrelated and non-inbred founders. Therefore F_P may often be an imprecise measure of the PGIBD, even when pedigrees are free of errors.

The increasing availability of genomic resources (e.g., many thousands of mapped SNPs) is making it possible to more precisely measure PGIBD with genetic markers than ever before. In particular, genome-scale molecular genetic data might make it possible to more precisely estimate PGIBD with molecular genetic data than with pedigrees. For example, PGIBD can be estimated as the reduction of multiple-locus heterozygosity (MLH) relative to Hardy-Weinberg expected homozygosity (F_H) (Purcell *et al.* 2007). This is possible because of the expected decline in MLH among individuals whose parents are more closely related. Heterozygosity-based measures of inbreeding are expected to be imprecise when relatively few markers are used because of high sampling variance (Balloux *et al.* 2004; Slate *et al.* 2004; Szulkin *et al.* 2010). Fortunately, the precision of marker-based measures of PGIBD will increase as more molecular markers (e.g., thousands of SNPs) rapidly become available.

IBD sites in the genome occur in contiguous homozygous chromosome segments (Figure 3-1) commonly referred to as 'runs of homozygosity' (ROH) (Chapman & Thompson 2003; Fisher 1965; McQuillan *et al*. 2008). The distribution of ROH lengths within an individual is determined by the number of generations since the common ancestor(s) of the parents and the recombination rate (Chapman & Thompson 2003). Inbreeding due to recent ancestors creates very long ROH on average (e.g., up to dozens of megabases in length). Inbreeding due to very distant ancestors creates shorter ROH on average, because of an increased number of meioses separating an inbred individual from it parents' common ancestor(s). ROH are expected to be longer in species with lower recombination rates due to fewer crossovers along a given chromosome throughout an inbred individual's ancestry.

It is possible to identify ROH as long consecutive runs of homozygous genotypes at mapped single nucleotide polymorphisms (SNPs). The fraction of an individual's genome that is inferred to be in ROH (F_{ROH}) can then be used as an estimate of PGIBD (McQuillan *et al.* 2008).

F_{ROH} has been widely used in studies of individual inbreeding and inbreeding depression in humans (Kirin *et al.* 2010; McQuillan *et al.* 2012; McQuillan *et al.* 2008). Runs of homozygosity have seldom been analyzed in natural populations (e.g., Pollinger *et al*. 2011), but hold great promise for our understanding of inbreeding and its fitness effects in the wild.

Little is known about the relative performance of F_{ROH}, F_H and F_P as measures of PGIBD. One previous study found that the number of homozygous rare alleles within an individual (a proxy for PGIBD) was more strongly correlated with F_{ROH} than with other marker-based inbreeding measures or F_P (Keller *et al.* 2011). However, their study was focused only on large populations ($N_e \geq 100$) and did not account for the effects of genetic map length, chromosome number (e.g., they simulated 220 cM genomes with two chromosomes) or SNP density across the genome. As noted above, the genetic map length and chromosome number can strongly affect the precision of F_P. It follows that F_P should be more weakly correlated with PGIBD in organisms with shorter map lengths. We believe that a thorough evaluation of the relative performance of F_P, F_{ROH}, and F_H in small populations (where inbreeding depression is most likely to occur) is badly needed. Such an analysis would advance our understating of how PGIBD (and inbreeding depression) should be measured as we quickly move into an era when tens of thousands of genetic markers will be commonplace in genetic studies of natural populations.

In this study, we asked whether PGIBD was better predicted by F_P, F_{ROH}, or F_H in small populations which are of the greatest concern for conservation. Specifically we evaluated the precision and bias of F_P, F_{ROH}, and F_H as measures of PGIBD in small populations with genomic characteristics typical of mammals. We addressed this objective while accounting for the depth of the pedigree used to estimate F_P, the number of markers used to estimate F_{ROH} and F_H, genetic map length of the genome, and demographic history.

Methods

Computer Simulation model

We wrote a stochastic, individual-based simulation program for R version 3.0.1 (R Core Team 2013). Below we describe the major components of the simulation program. An R script for our simulation model is available in the Dryad data repository (see Data Accessibility).

We simulated a sexually reproducing, hermaphroditic, non-selfing species with non-overlapping generations. We included recombination in our simulations and kept track of the ancestral origin of each chromosome segment. We simulated 3Gb genomes with 20 chromosomes of equal size. We simulated only 20 chromosomes because the number of chromosomes is not expected to strongly affect σ_{IBD} when genetic map lengths are in the range we simulated (Figure 3-S1, Franklin 1977). We used two different recombination rates in order to evaluate the effect of genetic map length on the precision of F_P, F_{ROH}, or F_H. First, we simulated a recombination rate of 1.2 cM/Mb which is similar to humans (Jensen-Seaman *et al.* 2004) and resulted in a genetic map length of 3600 cM. We also simulated a recombination rate of 0.27 cM/Mb which is typical of the lower end of the distribution of recombination rates among mammals (Dumont & Payseur 2008) and resulted in a genetic map length of 800 cM.

Most mammals have large genomes in terms of physical size (e.g., ≥ 2.5Gb) (Dumont & Payseur 2008). However, the genetic map length varies considerably among mammals. For example, genetic map lengths ranged from approximately 650 cM in short-tailed opossum (*Monodelphis domestica*) to > 3600 cM in humans (Dumont & Payseur 2008).

Our model of recombination is conceptually identical to the simulations of Chapman & Thompson (2012). This model of recombination is based on Fisher's theory of junctions (Fisher 1965). We assume no interference among chiasmata, and that the number of crossovers along a chromosome is Poisson distributed. We also assume that the recombination rate is constant across the genome and among individuals.

Population founders (those in the first simulated generation) were unrelated and carried two unique copies of each chromosome; thus PGIBD for an individual is relative to the non-inbred founders of the population. Our model is flexible in terms of the migration rate (m, the probability of an individual within a population being an immigrant) and temporal changes in population size. Simulation output included the true PGIBD for each individual in the final simulated generation. Simulation output also included genotypes at 100,000 SNPs with mean H_e = 0.3. We included only 100,000 SNPs because adding additional loci did not substantively affect the precision of the marker based measures of PGIBD in preliminary simulations (data not shown). SNP positions in the genome were randomly distributed across the genome using a random number generator.

Simulated demographic scenarios

We focused on two demographic scenarios for small populations. First we simulated partially isolated small populations (N_e = 20) receiving an occasional unrelated immigrant (m = 0.05, one immigrant per generation on average). This first scenario represents small populations on habitat islands with occasional immigration from a large source population – e.g., populations on oceanic islands with intermittent immigration from a large mainland population. Immigrants were unrelated to each other and to residents, and had genotypes drawn from the same allele frequency distribution as population founders. We ran simulations for this first scenario for 150 generations. Second, we simulated closed populations that were initially large (N_e = 500) but then experienced a reduction in size to N_e = 20. This second scenario represents populations that have recently experienced a sharp reduction in N_e. We ran these simulations for 90 generations with N_e = 500, then reduced N_e to 20 for an additional 10 generations. We ran twenty replicate simulations for both scenarios. The mean and variance of PGIBD were 0.16 and 0.01 on average in simulated partially isolated populations. In simulated populations with recently reduced N_e, the mean and variance of PGIBD were 0.22 and 0.003 on average.

We also simulated an additional demographic scenario to determine if the relative performance of the different PGIBD estimators was sensitive to population size. For this demographic scenario, we simulated twenty replicates of populations with a constant size of N_e = 100 and m = 0.01.

Measuring PGIBD

We defined PGIBD so that we would account for inbreeding due to recent ancestors, excluding the effects of inbreeding due to very distant ancestors – an approach consistent with most studies of PGIBD in humans (McQuillan *et al.* 2012; McQuillan *et al.* 2008). We adjusted the minimum length of the ROH used to calculate PGIBD according to the genetic map length. The genetic map length strongly affected the distribution of ROH lengths in our simulations. For example, the mean ROH length was 23.8 Mb in partially isolated populations with a genetic map length of 800 cM. However, in partially isolated populations with a genetic map length of 3600 cM, the mean ROH length was 6.2 Mb. Therefore, we defined PGIBD as the fraction of the genome in ROH that were 8 Mb or longer for simulations of 800 cM genomes. PGIBD was defined as the fraction of the genome in ROH that were 2 Mb or longer for simulations of 3600 cM genomes.

We used the program PLINK (Purcell *et al*. 2007) to detect ROH in the genomes of simulated individuals. We chose PLINK because it is commonly used in ROH studies and it has been shown to more reliably detect ROH than other algorithms (Howrigan *et al.* 2011). PLINK slides a window of 50 SNPs across the genome to identify ROH. PLINK allows the user to define the criteria for assigning a chromosome segment as being in an ROH. Specifically, the user can define the minimum number of contiguous homozygous SNPs, a minimum marker density (expressed as Kb/SNP), and the minimum length in Kb for a homozygous region to be considered in an ROH. Additionally the user can specify the maximum gap between adjacent SNPs allowed within an ROH. We adjusted the ROH detection parameters in PLINK based on the number of loci being used and the length of ROH being evaluated. The PLINK ROH detection settings we used are shown in Table S1. We used PLINK ROH detection parameters similar to previous studies of humans that used different numbers of SNPs to detect ROH (Kirin *et al.* 2010; McQuillan *et al.* 2012).

We estimated F_{ROH} as the sum of the lengths of all detected ROH that were 2 Mb or longer divided by the physical genome size (3 Gb) for simulations of 3600 cM genomes (typical of many vertebrates). For simulations of 800 cM genomes, we estimated F_{ROH} considering only detected ROH 8 Mb or longer. We used 15K-100K SNPs to estimate F_{ROH} and $F_{H.}$ We estimated F_H as described in Purcell *et al*. (2007) using PLINK. There are several other heterozygosity-based measures of individual inbreeding that could be used (Chapman *et al.* 2009). However, all of these measures are highly correlated and non-independent (Chapman *et al.* 2009). Thus using a different marker-based measure of PGIBD is unlikely to have substantively affected our results. We used the kinship2 package in the program R (Therneau *et al.* 2011) to estimate F_P for each individual using 3-20 generations of pedigree information.

Some studies put SNP genotype data through a process called 'linkage disequilibrium pruning' (LD pruning) before attempting to detect ROH (e.g., McQuillan *et al*. 2012). This involves removing from a data set one SNP from each pair of loci in a window of 50 SNPs that are above a user defined LD significance threshold. This is done in order to avoid detecting common ROH caused by high frequency ancestral haplotypes (e.g., due to natural selection). We chose not to use LD pruning here because our simulations did not include selection, so all long ROH should be reflective of parental relatedness and not of deep historical selection events.

Nevertheless, we reran analyses of partially isolated populations with a genetic map length of 3600 cM after LD pruning, and found that our results did not change substantively (Figure 3-S2).

Comparing the performance of F_P, F_H, and F_{ROH}

We used the proportion of variance in PGIBD explained by F_{ROH}, F_H and F_P (r^2) from simple linear regression models to evaluate the precision of each measure of PGIBD. We conducted separate regressions of F_{ROH}, F_H and F_P versus PGIBD for the individuals in the final generation of each simulated population. We used a natural log transformation to normalize the distribution of F_P before conducting regressions of F_P versus PGIBD. We then used two tailed *t*-tests to determine if the mean r^2 with PGIBD (among 20 replicate simulations) was statistically significantly different for F_{ROH}, F_H and F_P. We compared all possible combinations of pedigree depth (for F_P) and number of SNP used (for F_H and F_{ROH}) when testing if the mean r^2 with PGIBD was statistically significantly different for F_P versus F_H or F_{ROH}. Likewise, we tested all possible combinations of the number of SNPs used to estimate F_{ROH} and F_H when testing if the correlation with PGIBD was statistically significantly different for F_{ROH} and F_H. We also used *t*-tests to determine if the mean r^2 with PGIBD was statistically significantly different for F_{ROH}, F_H and F_P when the genetic map length was 800 cM instead of 3600 cM.

We measured bias as the mean amount by which F_P, F_H, and F_{ROH} over- or under-estimated PGIBD. As with r^2, we measured the bias of F_P, F_H, and F_{ROH} only considering individuals in the last generation of each simulated population.

Results

We first present results on the precision and bias of F_P, F_H, and F_{ROH} from simulations using a genetic map length of 3600 cM. We then describe the effects of the genetic map length on the precision and bias of the PGIBD estimators.

Precision in partially isolated small populations

The correlation of PGIBD with F_H and F_{ROH} was almost always stronger than the correlation between PGIBD and F_P in partially isolated populations (Figure 3-2). The correlation between F_P and PGIBD was very weak when pedigrees included few (e.g., ≤5) generations in partially isolated populations. For example, the mean r^2 between F_P and PGIBD (across 20 simulation repetitions) was 0.32 when three pedigree generations were used (i.e., when using complete pedigrees three generations deep). However, when pedigrees included twenty generations, the r^2 between PGIBD and F_P was 0.92. The r^2 between F_H and PGIBD was ≥ 0.88 when ≥ 1000 SNPs were used. The r^2 between PGIBD and F_{ROH} was > 0.97 when ≥ 35K SNPs were used.

The mean r^2 between PGIBD and F_H estimated with ≥5K SNPs was statistically significantly higher than the mean r^2 between PGIBD and F_P regardless of how many generations of pedigree were used to estimate F_P ($P < 0.002$ for all comparisons, *t*-tests). F_H estimated with 1K SNPS had a statistically significantly higher mean r^2 with PGIBD than F_P estimated with five or fewer generations of pedigree ($P < 0.0001$ for all comparisons, *t*-tests). The r^2 between PGIBD and F_{ROH} estimated with 50K SNPs was statistically significantly higher

than the mean r^2 between PGIBD and F_P regardless of the number of pedigree generations used to estimate F_P ($P < 0.0001$ for all comparisons *t*-tests).

PGIBD was more strongly correlated with F_H than with F_{ROH} when relatively few SNPs were used (Figure 3-2). For example, the r^2 between F_H and PGIBD was statistically significantly higher than the mean r^2 between F_{ROH} and PGIBD when 35K or fewer SNPs were used ($P < 0.0001$ for all comparisons, *t*-tests). However, the r^2 between F_H and PGIBD (mean $r^2 = 0.99$) was only slightly higher than the r^2 between PGIBD and F_{ROH} (mean $r^2 = 0.97$) when 35K SNPs were used. Plots of F_P, F_H and F_{ROH} versus PGIBD from a representative simulation of a partially isolated population are shown in Figure 3-3.

Bias in partially isolated small populations

F_P and F_H consistently underestimated PGIBD in partially isolated populations (Figure 3-4). F_P estimated with three generation pedigrees underestimated PGIBD by 0.14 on average. However, F_P underestimated PGIBD by less than 0.05 on average when pedigrees included twenty generations. F_H underestimated PGIBD by approximately 0.2 on average. The number of loci used to estimate F_H had no effect on bias. F_{ROH} underestimated PGIBD when relatively few loci were used but this bias decreased as more SNPs were used (Figure 3-4). For example, F_{ROH} underestimated PGIBD by > 0.15 on average when 15K SNPs were used. However, F_{ROH} estimated with 100K SNPs was an unbiased estimator of PGIBD.

Precision in populations with recently reduced N_e

PGIBD was almost always more strongly correlated with F_H and F_{ROH} than with F_P in populations with recently reduced N_e. Increasing the depth of the pedigree beyond five generations had no effect on the r^2 between F_P and PGIBD in populations with recently reduced N_e (Figure 3-2). The mean r^2 between F_P and PGIBD was never higher than 0.66 in this demographic scenario. The mean r^2 between PGIBD and F_P estimated with a three generation pedigree was 0.49.

The mean r^2 between PGIBD and F_H estimated with 1K SNPs was statistically significantly higher than the mean r^2 between PGIBD and F_P estimated with three generation pedigrees ($P < 0.0001$, *t*-test). PGIBD was more strongly correlated with F_H estimated with 5K SNPs than with F_P regardless of pedigree depth ($P < 0.0001$ for all comparisons, *t*-tests). The r^2 between PGIBD and F_{ROH} estimated with 35K SNPs was statistically significantly higher than the mean r^2 between PGIBD and F_P regardless of pedigree depth.

The relative performance of F_H and F_{ROH} was similar in populations with a recent reduction in N_e compared to partially isolated populations. For example, the mean r^2 between PGIBD and F_H was statistically significantly higher than the mean r^2 between PGIBD and F_{ROH} when 35K or fewer SNPs were used ($P < 0.0001$ for all comparisons, *t*-tests). However, the magnitude of the difference in the correlation with PGIBD for F_H and F_{ROH} was small when 35K SNPs were used. The r^2 from regressions of PGIBD versus F_H and F_{ROH} were slightly lower when 25-35K SNPs were used in populations with recently reduced N_e than in partially isolated populations (Figure 3-2).

Bias in populations with recently reduced N_e

F_P, F_H, and F_{ROH} had very similar bias in populations with recently reduced N_e compared to small partially isolated populations (Figure 3-4). F_P estimated with three generation pedigrees underestimated PGIBD by approximately 0.2 on average. F_P again underestimated PGIBD by less than 0.05 on average when pedigrees included twenty generations. F_H underestimated PGIBD by approximately 0.28 and the number of loci used to estimate F_H again had no effect on bias. F_{ROH} underestimated PGIBD by > 0.2 on average when only 15K SNPs were used. However, F_{ROH} underestimated PGIBD by < 0.03 when 100K SNPs were used.

Effects of the genetic map length on the precision and bias

F_P was a less precise measure of PGIBD when the genetic map length was 800 cM instead of 3600 cM (Figures 3-2 & 3-S2). The difference in the mean r^2 between PGIBD and F_P across different genetic map lengths was statistically significant ($P < 0.05$ for all comparisons, *t*-tests) except when pedigrees included 3-5 generations in partially isolated populations. The difference in the precision of F_P across different genetic map lengths was particularly large among simulations of populations with recently reduced N_e. For example, the mean r^2 between PGIBD and F_P estimated with a three generation pedigree was 0.49 when the genetic map length was 3600 cM in populations with recently reduced N_e (Figure 3-2). However, when the genetic map length was 800 cM, the mean r^2 between PGIBD and F_P estimated with three generation pedigrees was only 0.24 in populations with recently reduced N_e (Figure 3-S2).

There was a smaller difference in the precision of F_P across different genetic map lengths in partially isolated populations (Figures 3-2 & 3-S2). For example, the mean r^2 between PGIBD and F_P estimated with three generation pedigrees was 0.32 when the genetic map length was 3600 cM in partially isolated populations. However, the mean r^2 between PGIBD and F_P estimated with three generation pedigrees was 0.27 when the genetic map length was 800 cM in partially isolated populations.

The genetic map length had little effect on the r^2 between F_H and PGIBD (Figures 3-2 & 3-S2). For example, the difference in the r^2 between PGIBD and F_H across different map lengths was > 0.01 in magnitude and statistically significant ($P < 0.05$) only when 1000 SNPs were used. However, the r^2 between F_{ROH} and PGIBD was higher when the genetic map length was 800 cM rather than 3600 cM and ≤ 35K SNPs were used. The mean r^2 between PGIBD and F_{ROH} estimated 25K SNPs was 0.91 when the genetic map length was 800 cM and only 0.64 with a genetic map length of 3600 cM in populations with recently reduced N_e.

The bias of F_P and F_H were unaffected by the genetic map lengths of the simulated genomes (Figures 3-4 & 3-S3). However, F_{ROH} tended to be more downwardly biased among simulations with a genetic map lengths of 3600 cM; this difference in bias was similar for both demographic scenarios (Figures 3-4 and 3-S3). For example, the F_{ROH} estimated with 35K SNPs had a mean downward bias of 0.17 when the genetic map lengths was 3600 cM in populations with recently reduced N_e. However, the mean downward bias of F_{ROH} estimated with 35K SNPs was only 0.05 when the genetic map length was 800 cM.

The results from our simulations of partially isolated populations with $N_e = 100$ were qualitatively similar to the results presented above (Figure 3-S5) and are thus not presented in detail.

Discussion

In this study, we asked whether F_P, F_H, or F_{ROH} better predicts PGIBD in small populations. We found that F_P had the weakest correlation with PGIBD, particularly among simulations of genomes with short genetic maps. Additionally, PGIBD was more strongly correlated with F_H than with F_{ROH} when fewer than 25-35K SNPs were used. Our results suggest that marker-based measures of PGIBD based on thousands of loci should be preferred over F_P.

Bias and precision

The bias and imprecision of F_P have two major sources. First, as discussed above, there can be high variance in PGIBD among individuals with the same F_P (Franklin 1977; Hill & Weir 2011; Stam 1980). This imprecision may partially explain why MLH is usually only weakly related to F_P, and why heterozygosity-fitness correlations are sometimes observed when F_P-fitness correlations are absent (Forstmeier *et al.* 2012). Mendelian sampling and a finite genetic map length is expected to weaken the correlation between F_P and PGIBD, but not to cause F_P to underestimate PGIBD. The most likely cause of downward bias in F_P in our study is close relatedness and inbreeding of pedigree founders. When founders are related, inbreeding due to common ancestors deeper in an individual's ancestry is not accounted for, thus causing F_P to underestimate PGIBD. Having high variance in founder relatedness could weaken the correlation between PGIBD and F_P by causing variation among individuals in the magnitude of the underestimation of PGIBD. F_P will also underestimate PGIBD for individuals whose parents have a common ancestor that is an inbred founder. A focus of future research should be to determine the relative influence of related or inbred pedigree founders versus Mendelian segregation and recombination on the weak correlation between F_P and PGIBD.

Pedigrees are not only an imprecise and biased measure of PGIBD, they will also often be more difficult to obtain than sufficient molecular genetic data to precisely measure PGIBD. Very large numbers of SNPs can currently be typed for any organism, for example by restriction-site-associated DNA sequencing (RAD-seq) (Davey *et al.* 2011), or targeted resequencing (Cosart *et al.* 2011). Such new sequencing technologies, along with the increasing number of whole genome reference sequences (Haussler *et al.* 2009) are making it possible to precisely measure PGIBD with molecular markers. Obtaining genotypes at very large numbers of SNPs is still expensive, but this cost will almost always pale in comparison to the difficulty of obtaining accurate, multi-generation pedigrees for free-ranging natural populations. Considering the imprecision of F_P, and the great difficulty associated with obtaining reliable pedigrees, we encourage researchers to soon adopt genomic measures of PGIBD as the necessary resources quickly become available.

Effects of genetic map length on the relative performance of PGIBD estimators

Our results suggest that the advantages of F_{ROH} and F_H over F_P are greatest in organisms with short genetic map lengths. The shorter simulated genetic map length (800 cM compared to 3600 cM) reduced the precision of F_P in our simulations (Figures 3-2 & 3-S2). The reduced precision of F_P in organisms with shorter genetic maps was particularly strong in our simulations of populations with recently reduced N_e (Figures 3-2 and 3-S2). The reduced precision of F_P among organisms with shorter map lengths is easily attributed to the higher variance in PGIBD among individuals with the same F_P because of fewer crossovers per meiosis on average (Franklin 1977; Hill & Weir 2011). The correlation between PGIBD and F_H was unaffected by map length. F_{ROH} had a stronger correlation with PGIBD when the genetic map length was 800 cM instead of 3600 cM and relatively few SNPs (e.g., 25-35K) were used (Figures 3-2 & 3-S2). Longer ROH are expected to be more easily detected with a given number of markers. Thus the increased precision of F_{ROH} for genomes with a short map length is most likely due to a higher mean ROH length. These findings suggest that F_P should be viewed with particularly strong skepticism when used as a measure of PGIBD in organisms with very short genetic map lengths.

Effects of the number of loci on the performance of F_{ROH} and F_H

F_{ROH} was always weakly correlated with PGIBD when fewer than 25-35K SNPs were used. Additionally, F_{ROH} consistently underestimated PGIBD when fewer than 100K SNPs were used. The weak correlation with PGIBD and the downward bias of F_{ROH} are easily explained by a failure of low density SNPs to detect a large fraction of ROH. A larger fraction of ROH are detected when more markers are used, thus elevating the correlation with PGIBD and decreasing the downward bias of F_{ROH} (Figures 3-2 & 3-4). These results show that more markers are needed to obtain unbiased estimates of PGIBD than to obtain a strong correlation between F_{ROH} and PGIBD. The precision of F_{ROH} (i.e., its r^2 with PGIBD) is probably more important than its bias to most researchers. The r^2 between F_{ROH} and PGIBD is informative of whether individuals are correctly ranked with respect to PGIBD. Correctly ranking individuals with respect to PGIBD is the key to detecting inbreeding depression. However, as discussed below, unbiased estimates of PGIBD are also important in studies seeking to identify genes involved with inbreeding depression (Leutenegger *et al.* 2006).

The number of SNPs necessary for unbiased and precise estimation of F_{ROH} in a given study system should depend on marker heterozygosity and genetic map length. First, mean heterozygosity of SNPs will strongly affect the precision and bias of F_{ROH}. Higher heterozygosity markers should provide higher power to correctly identify ROH. Long stretches of homozygous genotypes are less likely to occur just by chance outside of IBD chromosome segments when mean heterozygosity is high. Second, our results suggest that the genetic map length of the genome will affect the necessary number of markers. Organisms with longer genetic maps will require more SNPs to achieve the necessary minimum marker density to confidently detect ROH caused by recent ancestors. Organisms with shorter genetic map lengths will require fewer markers to confidently detect ROH caused by recent ancestors because of an increased average ROH length. Researchers could conduct simulations similar to the ones used in this study to evaluate the precision and bias of F_{ROH} given a particular genomic architecture, recombination rate, set of markers, and a range of likely demographic and inbreeding scenarios.

F_H was always very strongly correlated with PGIBD ($r^2 > 0.9$) when 5K or more SNPs were used in our study. However, F_H consistently underestimated PGIBD and this magnitude of this bias was unaffected by the number of SNPs used. This bias can be explained by a discrepancy between the assumed and actual base populations used to estimate F_H. F_H and other similar heterozygosity-based statistics (e.g., Carothers *et al.* 2006) measure the proportional reduction of MLH relative to Hardy-Weinberg expected heterozygosity (H_e) (Purcell *et al.* 2007). Individuals that have MLH < H_e are then inferred to be inbred. This approach implicitly assumes that an individual with MLH = H_e is non-inbred (i.e., its parents are unrelated). For this assumption to hold, the allele frequencies used to estimate H_e must be derived from a base population where all parents would be unrelated if mating occurred randomly. For example, consider an extremely large population (e.g., $N_e \approx \infty$) with partial selfing. Here, genetic drift is essentially absent, and H_e would approximately equal the predicted MLH of an individual whose parents are unrelated. The proportional reduction in MLH relative to H_e would then provide an unbiased estimate of PGIBD (i.e., F_H would equal zero for a non-inbred individual). However, in studies of real populations, allele frequencies are usually only available from one or a few small populations where inbreeding depression is of interest to a researcher. In this common scenario, H_e will not equal the MLH of a non-inbred individual because of historical genetic drift. Therefore, F_H and other heterozygosity-based measures of inbreeding should be interpreted only after careful consideration of the individuals used to estimate allele frequencies. See Appendix 1 for a mathematical treatment of the bias in F_H.

Relative benefits of F_{ROH} and F_H

F_{ROH} has several attractive qualities not shared by F_H. First, analyses based on ROH can distinguish inbreeding due to recent versus distant ancestors (Kirin *et al.* 2010) whereas F_H cannot. Inbreeding depression could often be caused mainly by inbreeding due to very recent ancestors if most deleterious recessive alleles are purged over many generations. F_{ROH} can specifically measure PGIBD due to recent ancestors by restricting analyses to very long ROH – thereby ignoring PGIBD due to distant ancestors which may not strongly affect fitness. Heterozygosity-based measures of PGIBD are affected by an individual's entire ancestry. Thus the power and accuracy of tests for inbreeding depression might often by higher when PGIBD is measured with F_{ROH} instead of F_H. However, it could be important to account for variation in PGIBD due to distant ancestors when purging is inefficient at removing a large proportion of the genetic load (e.g., when inbreeding depression is caused by many deleterious recessive alleles with small effect sizes). Deleterious recessive alleles with small effects could be found in ROH of all lengths as we expect them to persist for many generations. Thus, measuring inbreeding due to distant ancestors by using F_H or by including short ROH in estimates of F_{ROH} could increase the power of tests for inbreeding depression in some studies. A logical approach would be to use both F_H and F_{ROH}, and to test multiple minimum ROH length thresholds for inclusion in estimates of F_{ROH}.

A second strength of ROH-based methods in studies of inbreeding depression is that homozygosity mapping can be used to identify recessive variants responsible for inbreeding depression (Leutenegger *et al.* 2006). Homozygosity mapping takes advantage of the fact that affected inbred individuals should have ROH surrounding the genomic positions carrying

phenotypically active mutations (Leutenegger *et al.* 2006; McQuillan *et al.* 2008). Homozygosity mapping requires enough genetic markers to accurately estimate of genome-wide PGIBD in order to sufficiently control the type I error rate (McQuillan *et al.* 2008). Homozygosity mapping should be a powerful tool in the search for the genomic regions associated with inbreeding depression in natural populations.

The current onslaught of DNA sequence data from non-model organisms is providing the tools necessary to use F_{ROH} in natural populations. Estimating F_{ROH} requires a reliable physical genome map. High quality reference genome assemblies from model or agricultural organisms can often be used as a physical genome map for closely related non-model study organisms (Cosart *et al.* 2011). Additionally, whole genome sequences are being constructed for many non-model organisms (Haussler *et al.* 2009; Levine 2011). However, most species will lack reliable physical genome maps for some time. ROH-based approaches to studying inbreeding are not possible in such organisms. Fortunately, heterozygosity-based measures of PGIBD which do not require information on the location of markers in the genome, can precisely measure PGIBD when using only a few thousand loci. Additionally, our results suggest that F_H will be substantially more precise measure of PGIBD than F_{ROH} when relatively few SNPs are available. When a reliable physical genome map and large numbers of SNPs are available, we suggest that F_{ROH} should be used to measure PGIBD because of its advantages discussed above.

Conclusions

In summary, our results suggest that genomic marker-based estimates of PGIBD are substantially more precise than F_P. PGIBD is likely to be more precisely predicted with F_H than with F_{ROH} when fewer than approximately 25-35K SNPs are available in organisms with genomic characteristics typical of mammals. When larger numbers of SNPs are available (> approximately 35K), F_{ROH} has several advantages including the ability to use homozygosity mapping to identify the loci causing inbreeding depression, and differentiating between inbreeding due to recent versus distant ancestors. The increased precision of F_{ROH} and F_H over F_P is greatest in organisms with short genetic maps. We encourage researchers to adopt F_{roh} and F_H (or other heterozygosity based measures of inbreeding) as the preferred measures of PGIBD as large numbers of markers and physical genome maps quickly become available for non-model organisms.

Acknowledgements

This material is based on work supported by the National Science Foundation through the Montana Institute on Ecosystems (EPS-1101342) and the Montana Ecology of Infectious Diseases IGERT program (DGE-0504628) (MK's support). GL and FWA were supported by National Science Foundation grant DEB-074218. GL was also supported by DEB-0723928. We thank Helen Taylor for comments on a previous version of the manuscript.

Data Accessibility

An R script for the simulation program used in this paper is available in the Dryad data repository (doi:10.5061/dryad.54g7b). The Dryad DOI will not be active until the final version of the paper is published online. However, the script can be accessed for review here: http://datadryad.org/review?wfID=20036&token=b7e4c889-a18c-46c3-9eeb-3910cb21a756

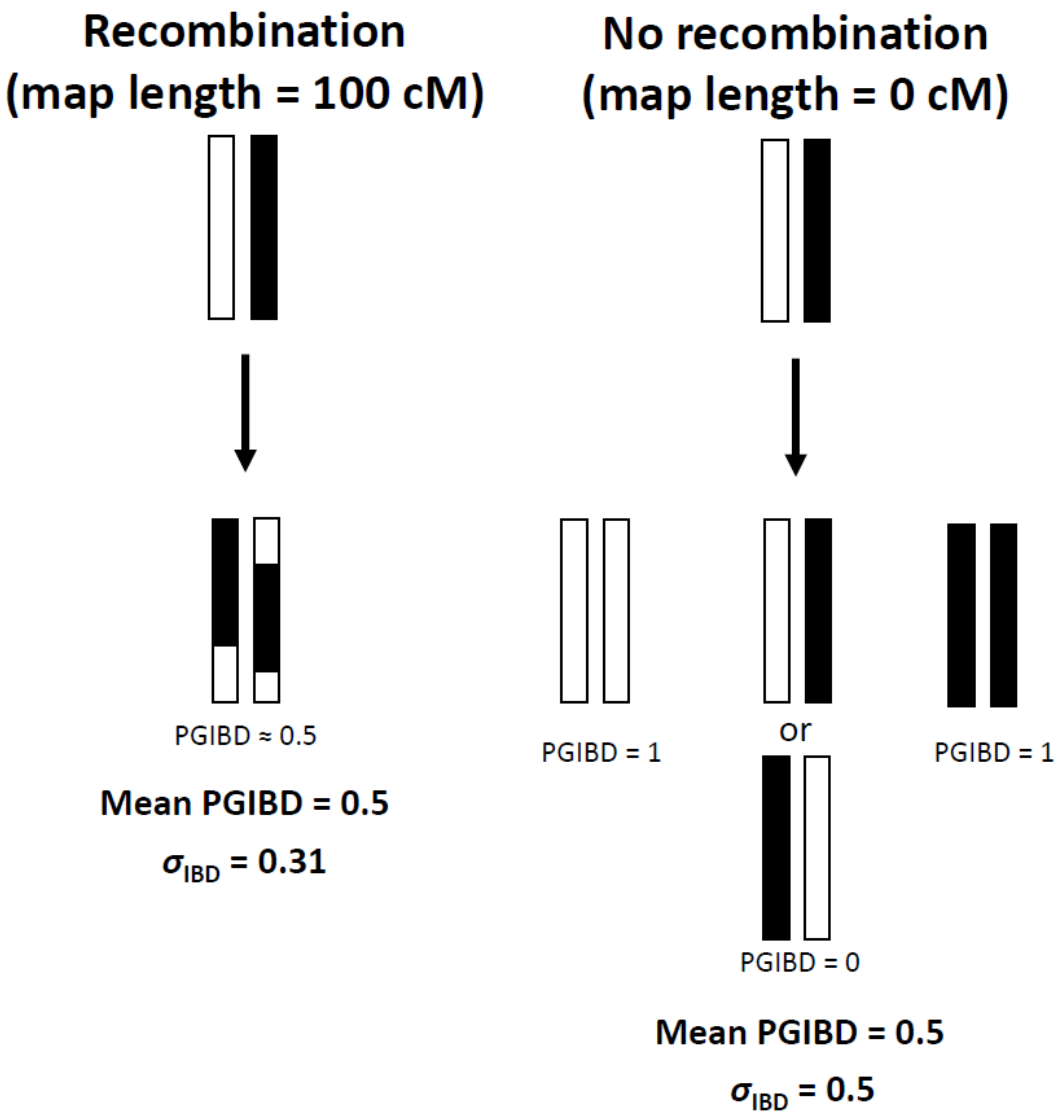

Figure 3-1. ROH in a chromosome from the offspring resulting from one generation of selfing with (left) and without recombination (right). A single pair of homologous chromosomes are shown from a non-inbred parent (top) and the offspring of the individual (bottom). IBD segments occur in the offspring where both chromosomes are derived from a single chromosome in the parent. The mean PGIBD will be the same in both cases, but the variance of PGIBD is much greater without recombination.

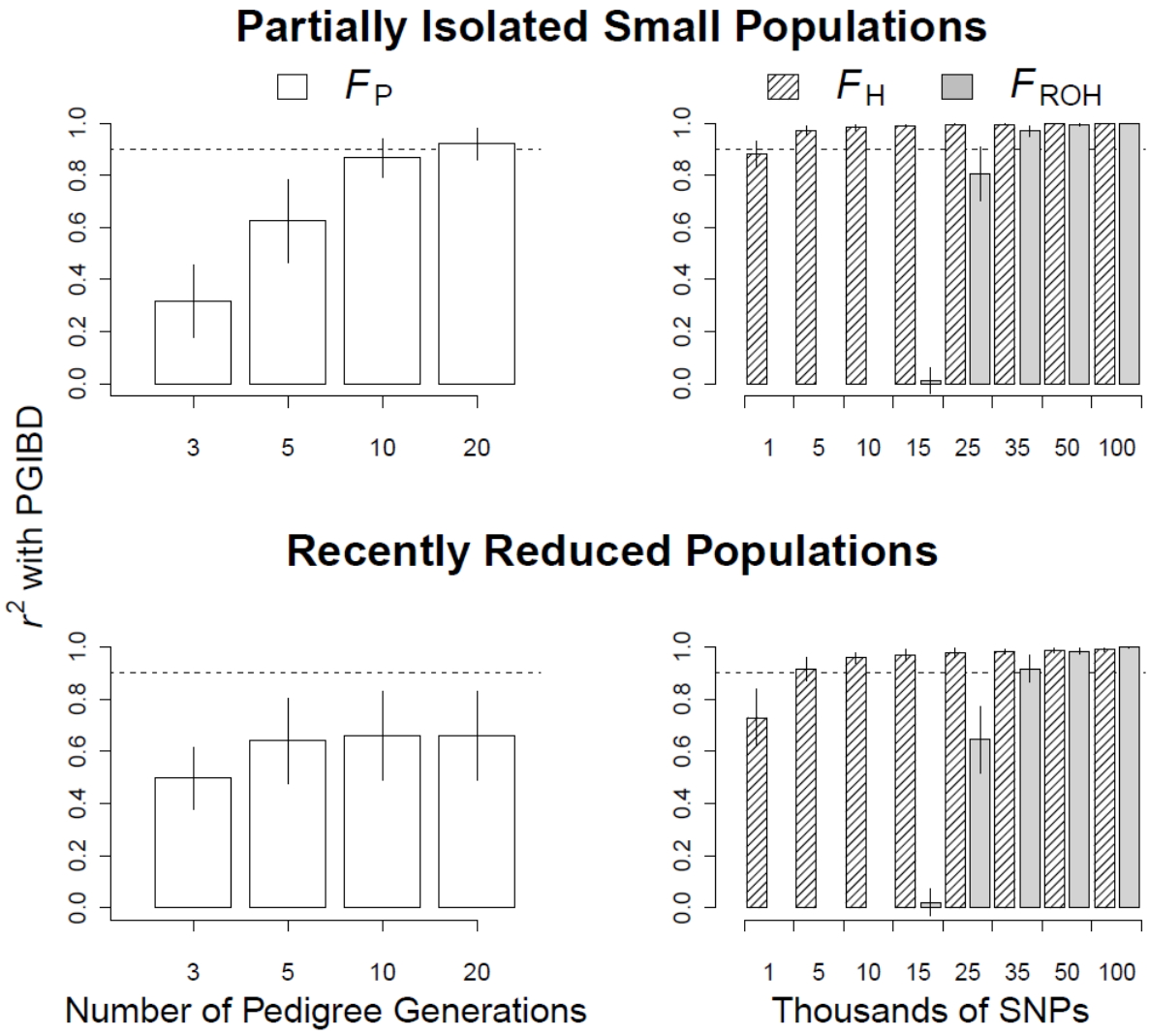

Figure 3-2. Barplots of the mean r^2 (+/- 1 s.d. across 20 simulated populations) from regressions of F_P, F_H and F_{ROH} versus PGIBD. Results shown here are from simulations of genomes with a genetic map length of 3600 cM. Results from 20 partially isolated (m = 0.05) small populations (local N_e = 20) are shown in the top row. The data shown in the bottom row are from 20 populations with a recent reduction in N_e (from N_e = 500 to N_e = 20). Horizontal dotted lines are placed at r^2 = 0.9 to aid comparison of r^2 across F_P, F_H and F_{ROH}.

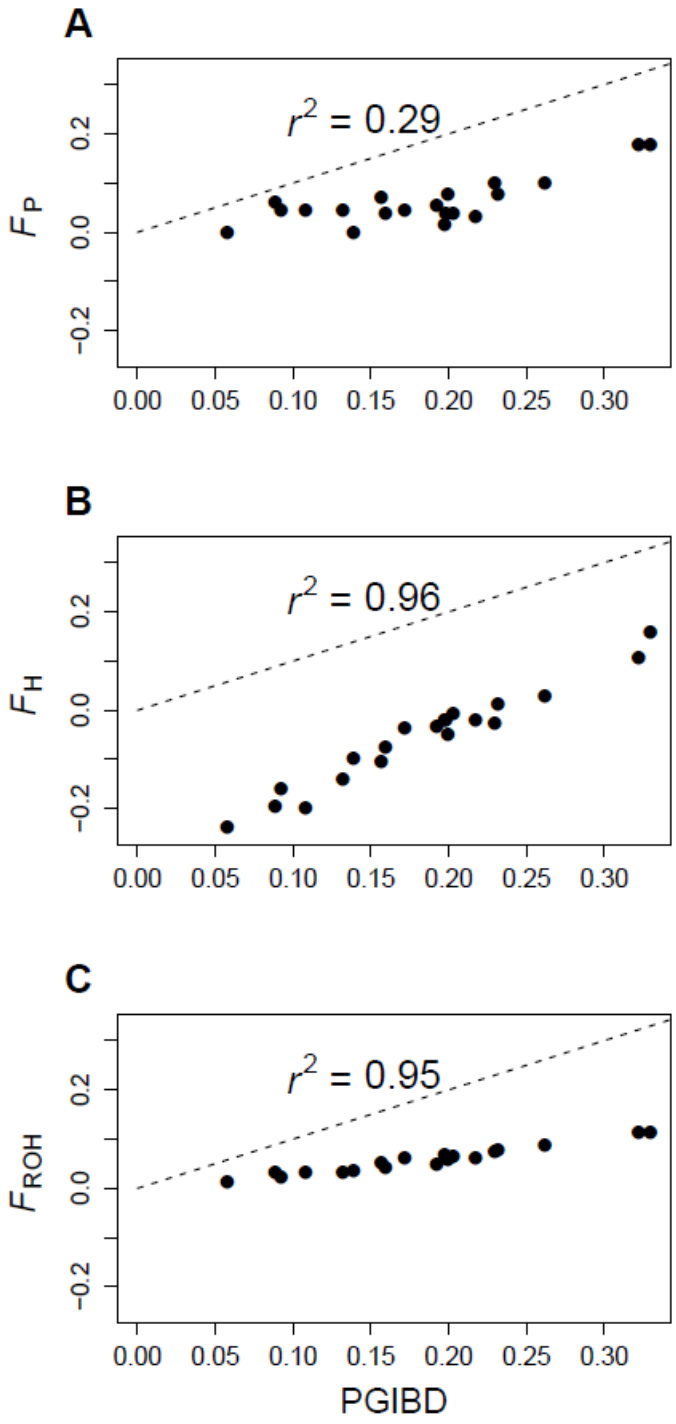

Figure 3-3. F_P (A), F_H (B), and F_{ROH} (C) versus PGIBD in a representative simulation of a partially isolated population with a genetic map length of 3600 cM. We used a pedigree that included 5 generations to estimate F_P. F_H was estimated with 5K SNPs, and F_{roh} was estimated with 35K SNPs. The dashed lines have an intercept of zero and a slope of one. Points below the lines represent underestimates of PGIBD.

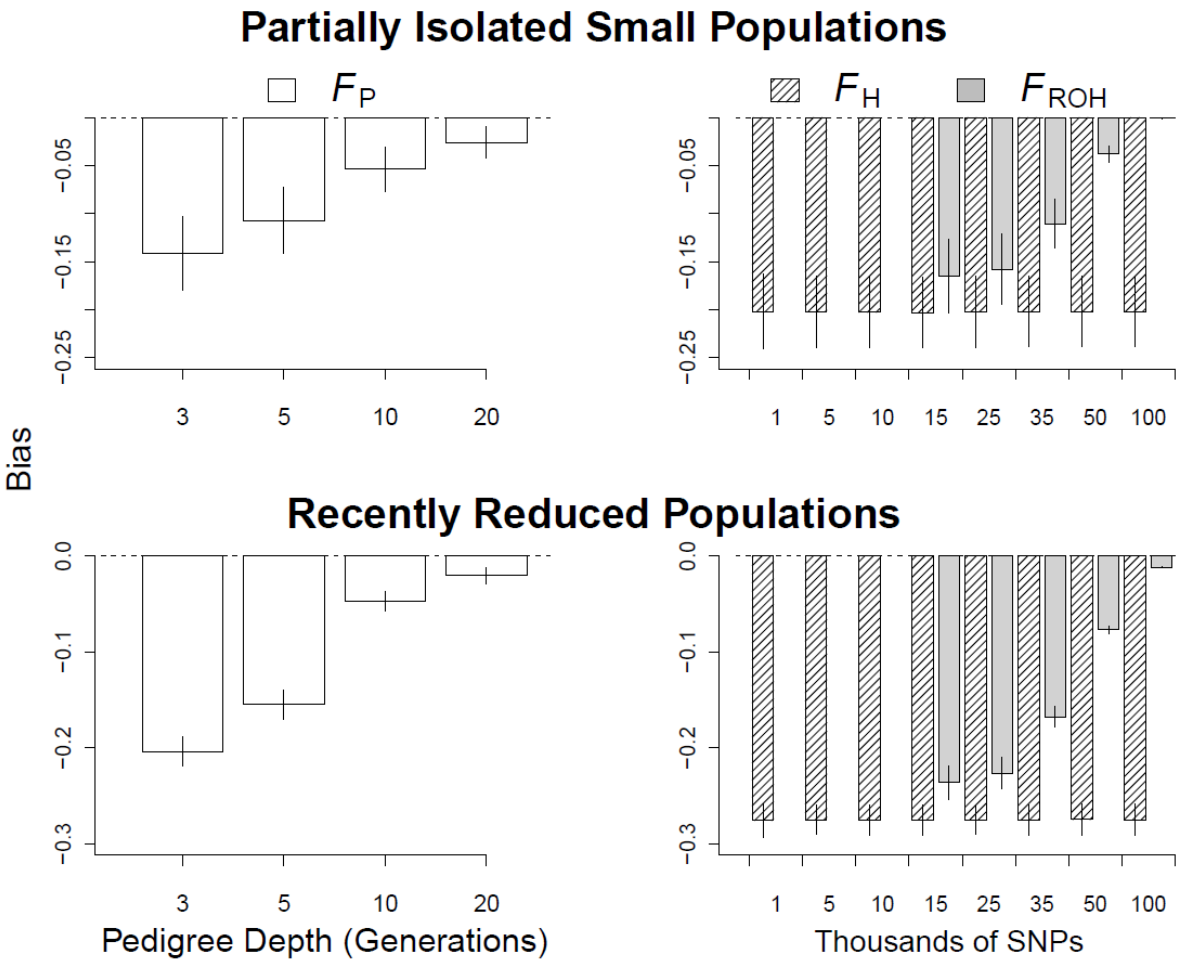

Figure 3-4. The bias of F_P, F_H, and F_{ROH} among simulations of genomes with a genetic map length of 3600 cM. Results from 20 simulations of partially isolated small populations are shown in the top row. Results from 20 simulations of populations with a recent reduction in N_e are shown in the bottom row.

CHAPTER 4: Cryptic pedigree founder relationships reduce the power to detect inbreeding depression: can genetic markers help?

Abstract

The pedigree inbreeding coefficient (F_P) is often a poor predictor of the identical by descent proportion of the genome (F). Indeed, F can vary greatly among individuals with the same F_P because of a limited number of recombinations during meiosis. Additionally, F_P assumes that pedigree founders are unrelated and non-inbred. We used simulations to assess the precision of F_P and the power to detect inbreeding depression when pedigree founders were inbred and related. We also assessed whether incorporating estimates of founder kinship (based on 100-500 single nucleotide polymorphisms) into pedigree analyses substantially increased the precision of F_P and the power to detect inbreeding depression. Inbred pedigree founders had no effect on the correlation between F_P and F. However, F_P was weakly correlated with F and the power to detect inbreeding depression was very low when most pedigree founders were related (e.g., in small partially isolated populations). Incorporating marker-based estimates of founder kinship into pedigree analyses increased the r^2 between F_P and F by < 20% in all simulated scenarios. However, accounting for founder kinship increased the power to detect inbreeding depression (caused by 10 lethal equivalents) by up to 25 times when pedigrees included ≤ 6 generations. Unfortunately, the power to detect inbreeding depression was still low (power < 0.7) after accounting for founder kinship in shallow pedigrees. We attribute the poor performance of F_P corrected for founder kinship to strong downward bias and low precision of maximum likelihood kinship estimators. These results suggest that accounting for founder kinship will rarely result in precise pedigree-based estimates of F or high power to detect inbreeding depression. Future studies should consider using pure marker-based estimates of F and tests for inbreeding depression, particularly in populations with shallow or incomplete pedigrees.

Introduction

Individuals with related parents often have dramatically reduced fitness (Keller & Waller 2002). This phenomenon, known as inbreeding depression, can reduce population growth and the probability of population persistence (O'Grady *et al.* 2006; Saccheri *et al.* 1998; Westemeier *et al.* 1998). Additionally, inbreeding depression is thought to affect the evolution of inbreeding avoidance behaviors such as dispersal and kin avoidance (Charlesworth & Charlesworth 1987). Thus, measuring individual inbreeding is a crucial component of many studies in medicine, ecology, evolutionary genetics, and conservation biology.

The classical measure of individual inbreeding is Wright's pedigree inbreeding coefficient (F_P) (Crow & Kimura 1970; Keller & Waller 2002; Pemberton 2004; Pemberton 2008; Wright 1922). F_P predicts the proportion of the genome that is identical by descent (F) due the known common ancestors of parents (Jacquard 1975; Keller & Waller 2002; Templeton & Read 1994). F_P has historically been strongly preferred over marker-based measures of individual inbreeding (Pemberton 2004; Pemberton 2008). The general preference for F_P is rooted in the perceived imprecision (high variance) of marker-based measures relative to F_P

when relatively few loci are used (Balloux *et al.* 2004; Pemberton 2004; Pemberton 2008; Slate *et al.* 2004).

However, F_P itself can be an imprecise measure of F. First, F_P assumes that the founders of a pedigree and immigrants are non-inbred and unrelated. Unfortunately, neither of these assumptions is likely to hold in any real population, particularly in small populations where inbreeding depression is a concern. When founders are related, pedigrees including few generations will fail to include many recent common ancestors of parents. Moving the reference population further back in time by adding more distant generations to the pedigree increases the F_P of individuals in more recent generations. Additionally, F will be underestimated for those individuals whose parents share a common ancestor that is an inbred founder.

F_P can also be a poor predictor of F because F can vary substantially among individuals with the same pedigree (i.e., siblings) due to a limited number of recombination events during meiosis (Franklin 1977; Hill & Weir 2011). The variance of F among individuals with the same pedigree is higher in organisms with shorter genetic map lengths or fewer chromosomes (i.e., fewer recombination events per meiosis) (Franklin 1977). Recent simulations have shown that F is almost never precisely predicted by F_P, and that pure marker-based measures of F are typically more strongly correlated with F when very large numbers of loci (e.g., thousands of single nucleotide polymorphisms, SNPs) are used (Kardos et al. *in review*, Chapter 3).

A potential solution to the imprecision of F_P is to incorporate genetic marker-based estimates of the kinship (f) of all founder pairs into pedigree analysis. Accounting for the f of founders might increase the precision of F_P substantially if a large proportion of the imprecision of F_P is due to related pedigree founders. Several methods use individual genotypes and allele frequencies to estimate pair wise relatedness ($2\times f$) (Blouin 2003; Wagner *et al.* 2006; Wang 2007; Weir *et al.* 2006). Replacing the assumed f of zero among founders with estimates based on genetic markers might adequately account for founder relatedness (Rudnick & Lacy 2008). Incorporating estimates of founder f into pedigree estimates of inbreeding is easy, and should become standard practice if doing so substantially increases the precision of F and the power to detect inbreeding depression.

In this study, we asked the following questions: When do inbred and/or related pedigree founders substantially reduce the precision of F_P and the power to detect inbreeding depression? Do genetic marker-based estimates of founder f substantially improve the precision of F_P and the power to detect inbreeding depression?

Methods

The simulation model

We wrote an individual-based simulation model for the program R version 3.0.1 (R Core Team 2013). The model simulates a hermaphroditic, sexually reproducing, non-selfing, random-mating species with non-overlapping generations, and produces full pedigrees and simulated genotypes at 1500 diallelic SNP loci. The first simulated individuals in a population were given genotypes by randomly choosing two alleles with frequencies chosen so that $H_e = 0.3$ on average in the final

simulated generation. We simulated 3 Gb genomes with 20 chromosomes and a genetic map length of 3600 cM, which are typical genomic parameters among mammals (Dumont & Payseur 2008). Population size was held constant at 60 individuals so that inbreeding would accumulate quickly, and so that the simulated effective population size would be similar to small natural populations where inbreeding depression is a concern. The F of each simulated individual was calculated as the proportion of the physical genome that was IBD. The details of our simulation program, and a comparison of simulation output with theoretical expectations are described elsewhere (Kardos *et al.* 2013).

Simulating inbred pedigree founders

To model the effects of inbred founders, we began by simulating 6 populations with 60 unrelated founders. Each simulated population had a different proportion of inbred founders. Specifically, we simulated populations with 0-100% of founders with $F_P = 0.25$. We then simulated random mating in the each population for five additional generations. We used the pedigrees to calculate F_P for each individual in the last (sixth) generation of the simulated populations.

Simulating related pedigree founders

We simulated pedigrees with extreme variation in the relatedness of founders in order to assess how related founders can affect estimates of inbreeding. We began each simulation by assigning 60 non-inbred founders, with some proportion (0-100%) of them being full siblings. We then simulated random mating in the population for five additional generations. F_P was calculated for each individual in the final simulated (sixth) generation.

We conducted twenty replicate simulations for each of six simulated proportions of sibling or inbred pedigree founders. We used the squared correlation (r^2) from linear regression models of F_P versus F ($r^2(F_P, F)$) to evaluate the precision of F_P.

Simulating pedigrees in approximate equilibrium populations with inbreeding depression

Pedigrees built in natural populations will often have both inbred and related founders. The distribution of relatedness and inbreeding among individuals in a random mating population is determined by a combination of genetic drift within the population, and immigration of individuals from other (genetically differentiated) populations. Immigration produces a mixture of immigrants, hybrids, and pure residents. Higher immigration rates result in lower mean inbreeding and relatedness among individuals. Lower immigration results in a larger proportion of 'resident' individuals and higher mean inbreeding and relatedness among individuals in a population.

We simulated populations with different migration rates (m, the probability of an individual being an immigrant) to mimic the distributions of founder relatedness and inbreeding expected with different amounts of population subdivision. We ran 50 replicate simulations using each of four migration rates (m = 0.0083, 0.017, 0.033, 0.1). These values of m are equivalent to 0.5-6 migrants arriving in a population per generation (on average). Each simulation was initiated with unrelated individuals, and ran for 70 generations to allow the

populations to reach approximate migration-drift equilibrium. Immigrant individuals were unrelated to residents, and had genotypes that were randomly drawn from the same allele frequency distribution as the individuals in the first generation of the simulation. We simulated inbreeding depression for survival to reproduction with 10 lethal equivalents per diploid genome, following the inbreeding depression model of Morton *et al.* (1956). We did not simulate the loci responsible for inbreeding depression so that the strength of inbreeding depression would be known, and constant across simulation repetitions. The distributions of the mean and variance of F for our simulated partially isolated populations are shown in Figure 4-S1.

We used pedigrees including 2-20 generations to calculate F_P only for individuals in the last generation of each simulation. We conducted statistical analyses only on individuals from the last simulated generation in order to hold sample size constant, and so that the amount of pedigree information used to estimate F_P was constant among individuals in each analysis. We used $r^2(F_P, F)$ to measure how informative F_P was of F. We used generalized linear models to test for inbreeding depression on survival ($\alpha = 0.05$) (Armstrong & Cassey 2007). We estimated statistical power as the proportion of simulated populations where tests for inbreeding depression were statistically significant ($\alpha = 0.05$).

Incorporating marker-based estimates of founder kinship

We used 100 - 500 SNPs ($H_e \approx 0.3$) to estimate the pair wise kinship (f) of each pair of pedigree founders in our simulations of populations with sibling founders. The nine parameter relatedness estimator implemented in the program Coancestry (Wang 2007; Wang 2011) was multiplied times 0.5 to yield an estimate of pairwise kinship ($\hat{f}$) (Blouin 2003). The original assumed f for each pair of founders (which was zero) was then replaced with the marker-derived $\hat{f}$. We then calculated the pedigree inbreeding coefficient corrected for founder kinship (F_{PC}). We estimated r^2 from linear regressions of F_{PC} versus F ($r^2(F_{PC}, F)$) and tested for inbreeding depression using F_{PC} in addition to F_P.

We used t-tests to determine if the mean $r^2(F_P, F)$ and $r^2(F_{PC}, F)$ were statistically significantly different among the replicate simulations for each simulated migration rate. The prop.test function in the program R (Harrell 2012) was used to test for a statistically significant difference in the statistical power of tests for inbreeding depression that used F_P and F_{PC}.

Results

Effects of inbred and related founders on $r^2(F_P, F)$

The proportion of inbred ($F_P = 0.25$) founders had very little effect on the correlation between F_P and F in our simulations. There was no indication of substantive changes in the bias or precision of F_P as the proportion of inbred founders increased (Figure 4-S2). Given the close agreement between F_P and F when founders were inbred, we focus on the effects of related pedigree founders throughout the rest of the paper.

F_P became less precise as the proportion of sibling founders increased (Figures 4-1 & 4-2). The mean $r^2(F_P, F)$ across 20 replicate simulations was approximately 0.8 when all pedigree founders were unrelated (Figure 4-2). However when ≥ 60% of founders were siblings, the mean $r^2(F_P, F)$ was always < 0.5. As expected, F_P also became increasingly downwardly biased as the proportion of sibling pedigree founders increased (Figure 4-1). For example, F_P was rarely > 0.0 for individuals with $F < 0.2$ when all pedigree founders were siblings.

How useful are marker-based estimates of founder kinship in populations with sibling founders?

$r^2(F_{PC}, F)$ was higher than $r^2(F_P, F)$ on average when an intermediate proportion of pedigree founders were siblings (Figure 4-2). For example, the mean $r^2(F_{PC}, F)$ among 20 replicate populations (using 100-500 SNPs to estimate founder *f* values) was statistically significantly higher than the mean $r^2(F_P, F)$ when 40-80% of founders were full siblings. However, when either 0% or 100% of founders were full siblings there was no difference between the mean values of $r^2(F_{PC}, F)$ and mean $r^2(F_P, F)$. The largest increase in $r^2(F_{PC}, F)$ relative to $r^2(F_P, F)$ (21%) occurred when 60% of founders were full siblings (Figure 4-2).

Effects of inbred and related founders on $r^2(F_P, F)$ and the power to detect inbreeding depression in populations at approximate equilibrium

The migration rate and the pedigree depth strongly affected $r^2(F_P, F)$ and the power to detect inbreeding depression (Figure 4-3). The mean $r^2(F_P, F)$ was ≤ 0.8 when pedigrees included fewer than 6 generations in all simulated demographic scenarios (Figure 4-3). The mean $r^2(F_P, F)$ was never higher than approximately 0.7 for populations with the lowest migration rate ($m = 0.008$, 0.5 migrants/generation) regardless of the depth of the pedigree (Figure 4-3). However, the mean $r^2(F_P, F)$ was > 0.8 when pedigrees included six or more generations in populations with the highest migration rate ($m = 0.1$, six migrants per/generation).

The power to detect inbreeding depression was always statistically significantly lower for F_P than for F when pedigrees included 10 or fewer generations (Figure 4-3). Power to detect inbreeding depression using F_P was always < 0.2 when pedigrees included on two generations. Power was < 0.8 for all simulated demographic scenarios when pedigrees included six or fewer generations. Power was never higher than 0.8 for simulations with $m = 0.008$ (0.5 migrants per generations).

How useful are marker-based estimates of founder kinship in population at approximate equilibrium?

Incorporating estimates of *f* for all pairs of founders into pedigree analyses only modestly increased (by < 22%) the precision of pedigree-based F estimates in all demographic scenarios (Figures 4-4 & 4-S2). We focus our results on simulations of populations with $m = 0.017$ (1 migrant per generation on average), because the difference between $r^2(F_{PC}, F)$ and $r^2(F_P, F)$ was similar across all demographic scenarios. Complete results from simulations of other demographic scenarios are shown in Figure 4-S3.

The mean $r^2(F_{PC}, F)$ across 50 replicate simulations was statistically significantly higher than the mean $r^2(F_P, F)$ when the pedigree depth was less than 20 generations in simulations with $m = 0.017$ (Figure 4-4). For example, $r^2(F_P, F)$ and $r^2(F_{PC}, F)$ were 0.16 and 0.22, respectively, when 2 generation pedigrees were used. When six generation pedigrees were used, $r^2(F_P, F)$ and $r^2(F_{PC}, F)$ were 0.42, and 0.59, respectively. $r^2(F_{PC}, F)$ was > 0.80 in simulations with $m = 0.017$ only when pedigrees included 15 or more generations. $r^2(F_{PC}, F)$ was never > 0.8 in simulations with $m = 0.008$, regardless of the depth of the pedigree (Figure 4-S2).

The power to detect inbreeding depression was dramatically higher for F_{PC} than for F_P when pedigrees included four or fewer generations in all demographic scenarios (Figures 4-4 & 4-S3). For example, the power to detect inbreeding depression was 25 times higher for F_{PC} (power = 0.5) than for F_P (power = 0.02) when pedigrees included three generations in simulations with $m = 0.017$ (Figure 4-4). Despite a large increase in the power to detect inbreeding depression with F_{PC} compared to F_P, power was still very low in most demographic scenarios when pedigrees included relatively few generations. For example, 10 to 15 generations of pedigree were necessary to achieve power > 0.8 for tests using F_{PC} in simulations with $m = 0.017$ and $m = 0.033$ (Figure 4-S3).

Statistical performance of $\hat{f}$

Having found that incorporating founder $\hat{f}$ values into pedigree analyses only marginally increased the correlation between the estimated and true F, we evaluated the statistical performance of $\hat{f}$. We simulated a population as described above with $m = 0.017$, and $N = 60$ individuals for 70 generations. Then we calculated the true f – the actual proportion of alleles shared identical by descent – between each pair of individuals in the 70 generation of the simulation. We calculated $\hat{f}$ as described above using 100-1000 SNPs. Plots of $\hat{f}$ versus f are shown in Figure 4-5. Loess smoothing functions were fit to the data to determine how $\hat{f}$ and its precision varied with f.

$\hat{f}$ had very low precision when relatively few SNPs were used (Figure 4-5). The mean error (square root of the mean squared residual) from the loess model fits were 0.025, 0.041, 0.009, and 0.007 for analyses based on 100, 250, 500, and 1000 SNPs, respectively. Additionally, we found that $\hat{f}$ was extremely downwardly biased and insensitive to variation in f when f was below approximately 0.2 regardless of the number of SNPs used (Figure 4-5).

Discussion

Our results show that related pedigree founders can strongly reduce the precision of F and the power to detect inbreeding depression under many demographic scenarios. Unfortunately, incorporating marker-based estimates of founder kinship into pedigree analyses only marginally improved the precision of F. Although, the power to detect inbreeding depression was dramatically increased by accounting for founder kinship in pedigree analyses. Power was still often too low (< 0.7) after accounting for founder kinship to reliably detect inbreeding

depression. Below we discuss the implications of these findings for our understanding of inbreeding and inbreeding depression in natural populations.

Effects of inbred founders on the correlation between F_P and F

We found that having inbred (but unrelated) pedigree founders rarely affected estimates of *F*. This makes sense because *F* can only be affected by founder inbreeding for individuals whose parents have a common ancestor who is a founder. This combination of events should be relatively rare compared to the frequency of related founders in large pedigrees. However, there are some circumstances when founder inbreeding might be very important. For example, if a pedigree consisted of only three generations, the only way for non-zero inbreeding to be detected is when a founder is a common ancestor of both the parents of an individual in the last (third) generation in the pedigree. Here the inbreeding of founders could be important as all known inbreeding loops end at individuals in the founding generation.

Effects of related founders on $r^2(F_P, F)$

We found that $r^2(F_P, F)$ was ~ 0.8 on average when all pedigree founders were unrelated. However, $r^2(F_P, F)$ was < 0.5 when the majority of pedigree founders were siblings (Figures 4-1 & 4-2). The imperfect correlation between F_P and *F* when founders are unrelated and non-inbred is caused purely by Mendelian segregational variance in *F* due to meioses that occurred after the founding generation.

We believe there are two ways that related founders can reduce $r^2(F_P, F)$. First, F_P is an unbiased estimator of *F* for individuals whose parents do not have related founder ancestors (Figure 4-1). However, F_P underestimates *F* on average for individuals who have closely related founder ancestors (Figure 4-1). When an intermediate proportion of founders are closely related, F_P will be downwardly biased for some individuals and unbiased for others. Therefore, a mixture of closely related and unrelated pairs of pedigree founders should increase the variance of F_P among individuals with the same *F*, thus weakening the correlation between F_P and *F*.

Second, the presence of related founders could increase the variance in *F* among individuals with the same F_P due to Mendelian segregation. When founders are all unrelated, there is no variation in the proportion of the genome shared IBD among all pairs of founders. However, there can be high variance in the proportion of the genome shared IBD among founders when some founders are related (Hill & Weir 2011). Therefore related pedigree founders could introduce additional error variance in *F* due to Mendelian segregation – further weakening $r^2(F_P, F)$.

The correlation between F_P and true F and the power to detect inbreeding depression in populations at approximate equilibrium

$r^2(F_P, F)$ was weak and the power to detect inbreeding depression was low in all simulated demographic scenarios when pedigrees included fewer than 6 generations. The precision of F_P and the power to detect inbreeding depression were the lowest in populations with the lowest migration rate. There are two ways higher migration rates could increase the

precision of F_P. First, immigration into a population reduces the frequency of closely related individuals in a population. Therefore pedigrees built in populations with frequent immigration from genetically differentiated populations should have fewer closely related pedigree founders. Second, higher immigration should reduce the mean number of generations that inbreeding loops extend through a pedigree. Therefore, the number of meioses contributing to Mendelian segregational variation in F should be lower in populations with more immigration.

How useful are genetic marker-based estimates of founder relatedness?

Incorporating estimates f for all pairs of founders does not solve the problem of related pedigree founders. We found that correcting for founder kinships only marginally increased $r^2(F_P, F)$ for all simulated demographic scenarios. The benefits of using estimates of founder kinship were modest in most of the scenarios we simulated. $r^2(F_{PC}, F)$ was never >20% higher than $r^2(F_P, F)$ in our simulated populations. The power to detect inbreeding depression was substantially higher with F_{PC} than with F_P when shallow pedigrees (<6 generations) was used, which are typical in studies of natural populations. However the power to detect inbreeding depression with F_{PC} was still very low compared to when the true F was used. Thus, our results suggest that using marker-based estimates of founder kinship in pedigree analyses will not increase the precision of F estimates or the power to detect inbreeding depression to acceptable levels in most circumstances.

Our results also suggest that F is hardest to measure precisely with pedigrees in isolated populations with little immigration. Such populations with high mean and variance in F (Figure 4-S1) are arguably the most important to the research community because they offer the greatest potential to elucidate the genetic basis of inbreeding depression and its importance to population growth. Additionally, inbreeding depression is more likely to reduce the mean fitness in very small populations where genetic drift is strong and the mean F is high. Our results suggest that F_P will usually be highly imprecise in such populations, even when pedigrees include many generations (e.g., 10-15 generations) and estimates are corrected for founder kinship.

The limited benefit of using marker-based estimates of founder kinship in pedigree analyses is unfortunate, but not surprising. The triadic likelihood-based pair wise f estimator we used has been found to outperform other methods (Wang 2007). However we found that Wang's 2007 f estimator was dramatically downwardly biased and insensitive to kinship of individuals with f less than approximately 0.2 (Figure 4-5). Additionally the f estimator had low precision when few loci were used.

The underestimation of the kinship of closely related individuals results from an assumption that allele frequencies were measured in a hypothetical base population where all genes are non-identical by descent (Wang 2007). Measuring allele frequencies in the current population being studied (which is normally the only option in empirical studies) makes individuals appear more distantly related than they really are. Using much larger panels of markers (e.g., high density SNP chips) could increase the precision of marker-based kinship estimates. However, increasing the number of markers should not decrease the bias of the likelihood kinship estimators.

When should purely marker-based measures of F be preferred over pedigrees?

Pedigrees including many generations will often be impractical in studies of inbreeding in natural populations. Molecular measures of inbreeding have emerged that could provide a more practical and accurate way to measure *F* than pedigrees. For example, *F* has been found to be more strongly correlated with individual heterozygosity based on as few as 1000 SNPs than with F_P, even when large numbers of generations are included in pedigrees (Kardos et al., *in review*, Chapter 3). Additionally, associations between heterozygosity estimated with small numbers of markers and fitness have been detected in studies where inbreeding depression was not detected via pedigree analyses (e.g., Forstmeier *et al.* 2012). New marker based measures of inbreeding that use physically mapped SNPs to identify IBD chromosome segments (Leutenegger *et al.* 2006; McQuillan *et al.* 2008) can precisely estimate *F* when tens of thousands of markers are available (Kardos et al., *in review*, Chapter 3). Furthermore, methods using mapped SNPs can be used to identify chromosome segments contributing to inbreeding depression (Leutenegger *et al.* 2006), which could greatly advance our understanding of the genetic basis of inbreeding depression.

The results presented here, combined with previous findings (Kardos *et al. in review*, Chapter 3) suggest that pedigree-based measures of inbreeding and tests for inbreeding depression will rarely, if ever, have higher precision or power than currently available molecular approaches. Deep pedigrees (e.g., > 15 generations), used in conjunction with molecular estimates of founder kinships will provide precise estimates of *F* in partially isolated populations (Figures 4-3 4-4, Kardos *et al*. *in review*, Chapter 3). However, the precision of *F* appears to always be low in closed populations, even when very deep pedigrees are used (Kardos *et al. in review*, Chapter 3). Thus, we believe that future studies of inbreeding depression should employ pure marker-based measures of *F* using thousands of SNPs when possible. Tens of thousands of SNPs can now be genotyped in almost any organism using new sequencing technologies (Davey *et al*. 2011). The ability to easily type huge numbers of markers should facilitate the widespread use of molecular estimates of *F* in the near future.

Conclusions

The presence of related pedigree founders can dramatically recue the precision (variance) of the pedigree inbreeding coefficient (F_P) and the power to detect depression when using F_P. We found that incorporating marker-based estimates of founder kinship into pedigree analyses can dramatically increase power. However, using genetic markers to account for founder kinship did not result in acceptably high precision of the pedigree inbreeding coefficient or high enough power to reliably detect inbreeding depression unless pedigrees included many (> 6) generations. Using marker-based approaches to estimate founder kinship should not be considered an adequate solution to the imprecision of pedigree-based measures of inbreeding. We encourage researchers to consider pure marker-based measures of true *F* when large numbers of genetic markers are available (L > 1000), which is now feasible for virtually any species.

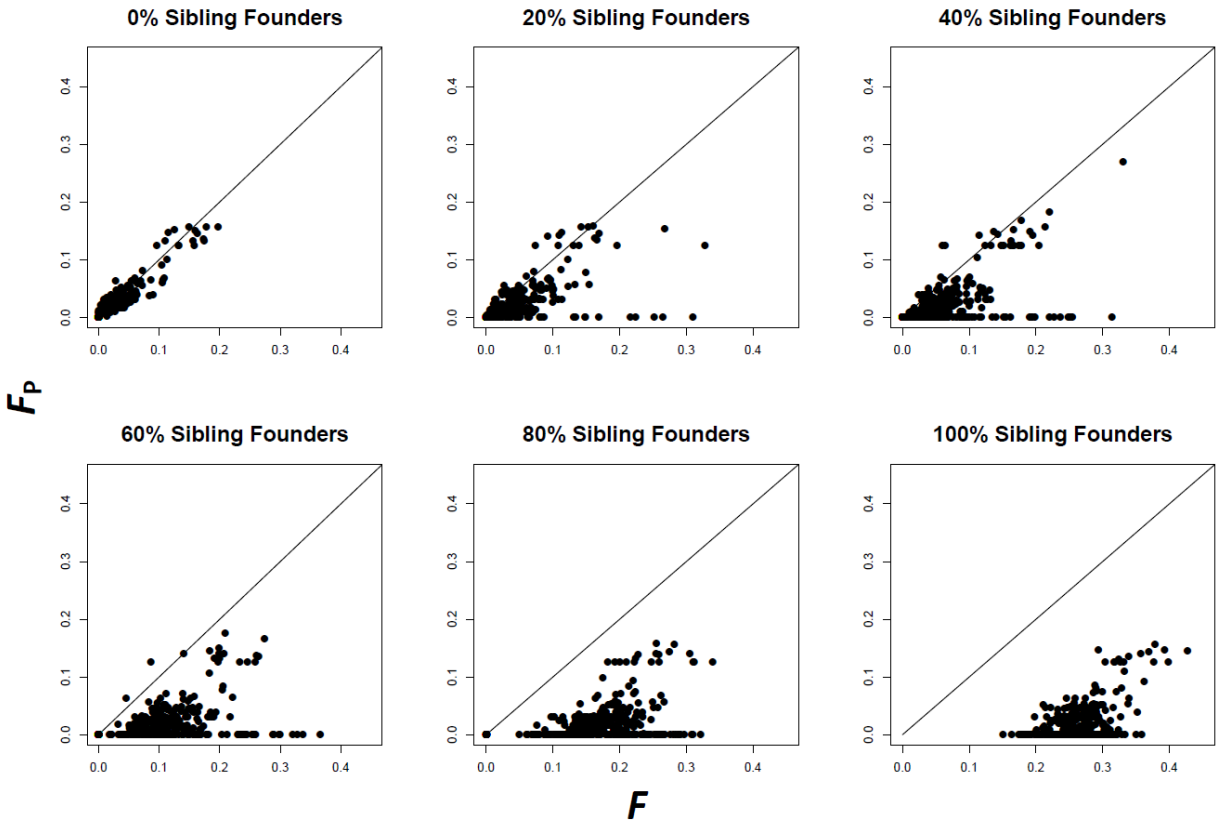

Figure 4-1. The estimated inbreeding coefficient (F_P) versus the identical by descent (IBD) proportion of the genome (F). The results shown are from simulations of six generation pedigrees with 0-100% sibling founders (from the same pair of parents). The solid diagonal lines have intercept of zero and a slope of 1. Points below the line represent underestimates of F.

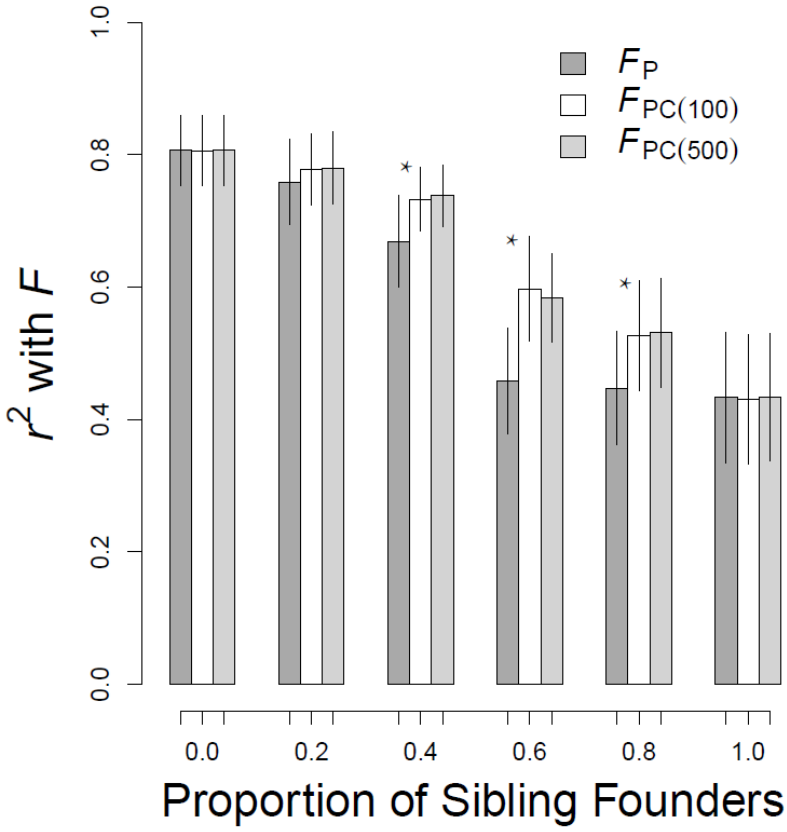

Figure 4-2. r^2 from the regressions of the pedigree inbreeding coefficient (F_P) and the corrected coefficient F_{PC} (using molecular markers) versus the true inbreeding (F; identitity by descent) plotted against the proportion of pedigree founders that were full siblings. 100 and 500 SNPs (mean $H_e \approx 0.3$) were used to estimate founder kinship coefficients which were used to estimate F_{PC}. Error bars represent the standard deviation of r^2 among twenty replicate simulated populations.

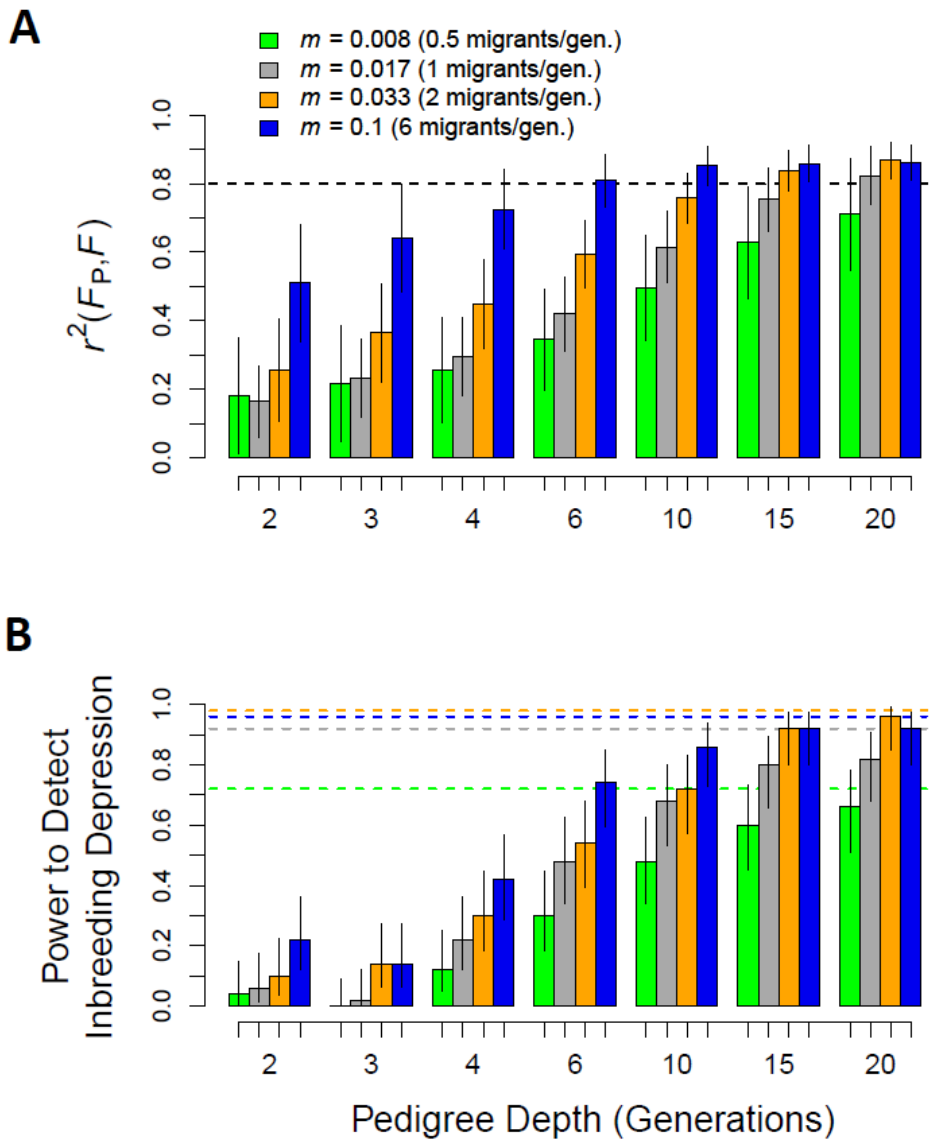

Figure 4-3. The r^2 between the estimated inbreeding coefficient (F_P) and the IBD proportion of the genome (F) ($r^2(F_E, F_T)$) versus pedigree depth. The data are from the final generation (60 individuals) of each of 50 independent simulations that were run for 70 generations with migration rates of 0.008 to 0.1 (0.5 - 6.0 migrants/generation). Error bars represent the standard deviation of r^2 among the 50 replicate simulations. Horizontal dashed lines in panel **B** represent the power to detect inbreeding depression using the true F. The vertical distance between the top of a bar and the dashed line of the same color represents the loss of power associated with imprecision of F_P.

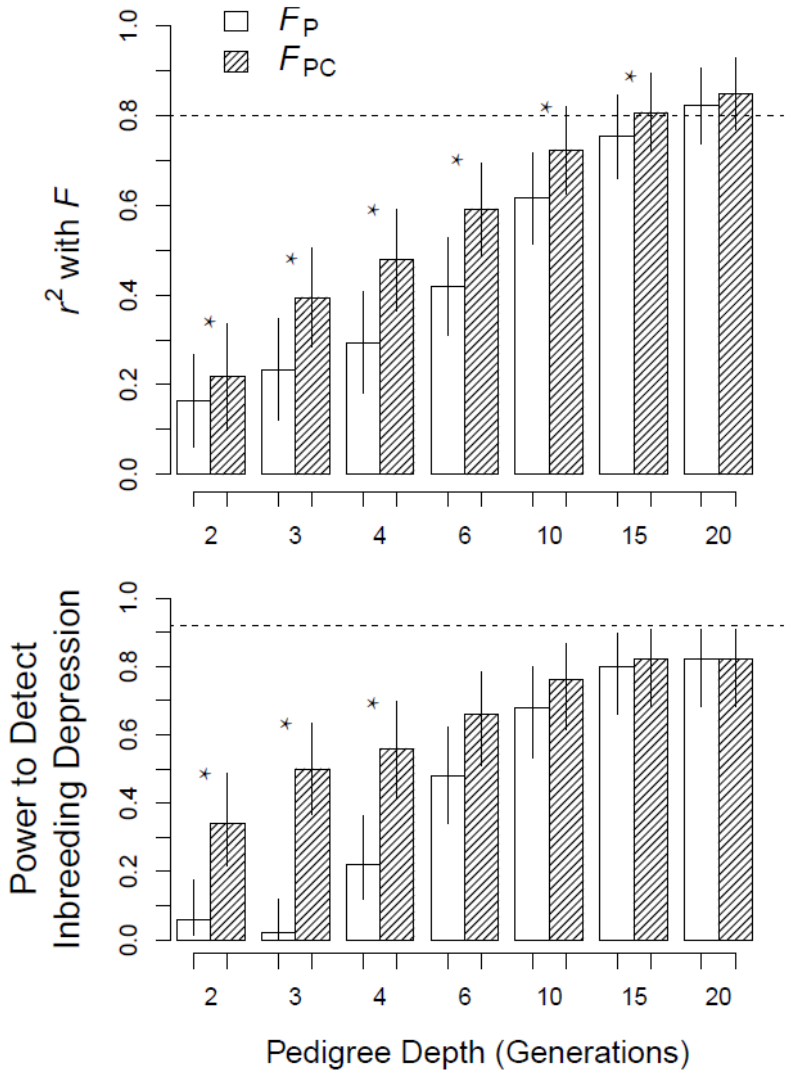

Figure 4-4. (**A**) The r^2 between the pedigree inbreeding coefficient (F_P) and the identical by descent proportion of the genome (F) [r^2 (F_P, F), open bars] and between the pedigree inbreeding coefficient corrected for founder kinship (F_{PC}) and F [r^2 (F_{PC}, F), hatched bars] versus the pedigree depth. Error bars in panel **A** are the standard deviation or r^2 among 50 replicate simulations. (**B**) The statistical power of tests for inbreeding depression using F_P and F_{PC}. Error bars in panel **B** are 95% confidence intervals for statistical power (the proportion of 50 replicate simulations with statistically significant tests [$\alpha = 0.05$] for inbreeding depression). The dashed horizontal line in panel **B** represents the power to detect inbreeding depression when using the true individual inbreeding coefficient (F; genome identity by descent).The data shown are from simulated populations with a migration rate of 0.017 (1 migrant/generation on average). Stars indicate a statistically significant difference between the mean r^2 (F_P, F) and mean r^2 (F_{PC}, F) among 50 replicate simulations. Results from simulations of other demographic scenarios are shown in Figure 4-S1.

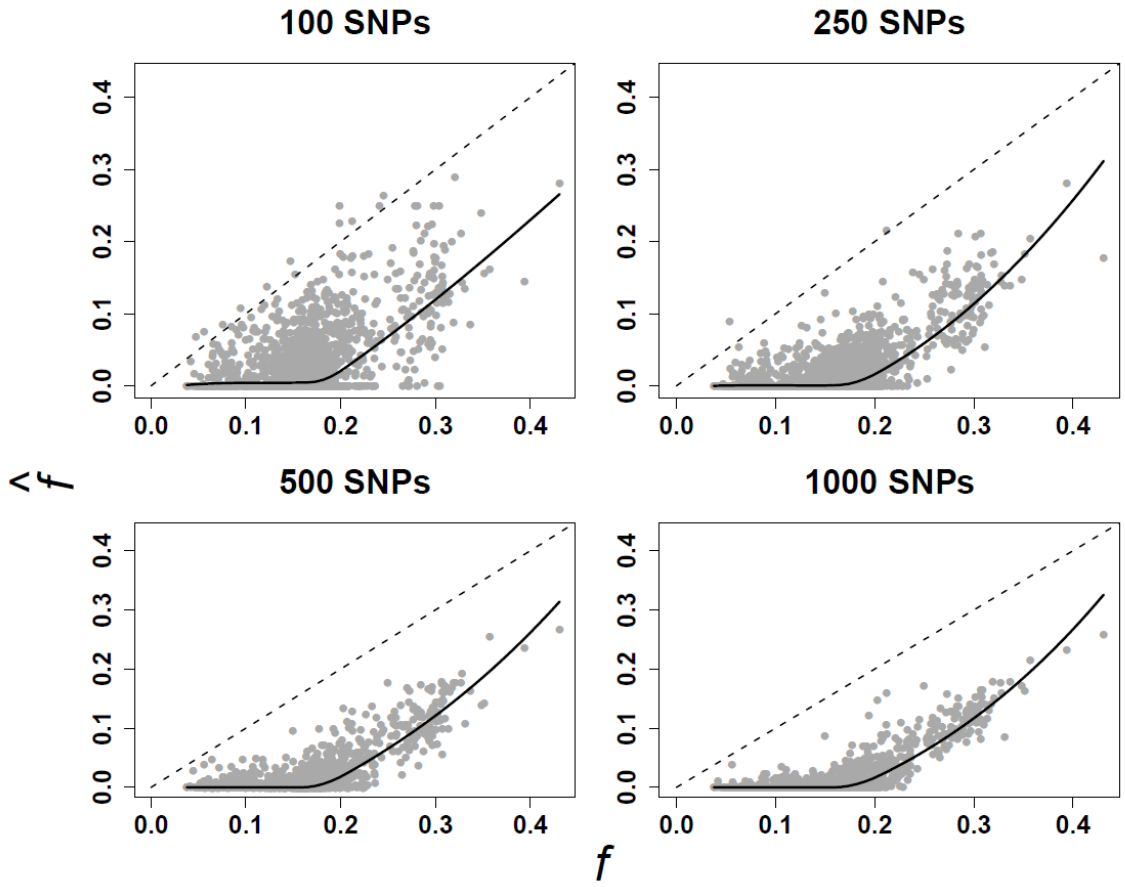

Figure 4-5. The estimated coefficient of kinship ($\hat{f}$) versus the true coefficient of kinship (f) for all pairs of 60 simulated individuals. The data are from the final generation of a simualtions with m = 0.017 that was run for 70 generations. The diagonal dashed lines have an intercept of 0 and a slope of 1. Points above the line represent over estimates and points below the line represent under estimates. The solid lines are fitted 2nd degree loess functions.

CHAPTER 5: Whole genome sequencing identifies candidate adaptive genes in wild bighorn sheep

Abstract

Understanding the genetic basis of fitness and adaptation is a central goal in evolutionary genetics, agriculture, and conservation biology. However, identifying adaptive genes is challenging in non-model species lacking whole genome reference sequences. Our objective was to identify candidate adaptive genes in wild bighorn sheep. We aligned whole genome sequences of pooled DNA from four bighorn sheep populations (> 30x coverage) to the domestic sheep reference genome assembly. We then scanned the genome for evidence of selective sweeps. We identified 2.57 M single nucleotide polymorphisms and a number of putative selective sweeps. For example, there was strong evidence for a selective sweep shared across all our study populations at a gene known to affect horn development in domestic sheep (*RXFP2*). Additionally, selection signatures were found at the growth hormone receptor (*GHR*) and insulin-like growth factor 2 (*IGF2R*) genes which are both involved with growth early in life. We have identified strong candidate genes for two phenotypic traits known to strongly affect fitness in bighorn sheep – horn development and body growth. Our results also show that whole genome reference assemblies from agricultural species can be used to identify molecular signatures of selection in wild populations of related taxa.

Key words: fitness, local adaptation, selection signature, selective sweep, population genomics, computer simulations

Introduction

Understanding the genetic basis of fitness and adaptation is a central goal in evolutionary and conservation biology. An efficient way to detect candidate adaptive loci is to scan whole genome sequences of pooled DNA and test for chromosomal segments bearing molecular signatures of natural selection. This approach has recently been used to identify signatures of selection associated with domestication in chickens (Qanbari *et al.* 2012; Rubin *et al.* 2010) and dogs (Axelsson *et al.* 2013), and to identify loci involved with local adaptation in *Arabidopsis thaliana* (Turner *et al.* 2010; Fischer *et al*. 2013). Such an approach holds great promise for identifying candidate genes and physiological pathways for adaptation in natural populations. However, using a genome-wide scanning approach to detect signatures of selection is not possible in most species because of the lack of high quality whole genome sequence assemblies.

The possibility of transferring high quality genomic resources from model or agricultural species to closely related taxa in the wild is particularly alluring (Haussler *et al.* 2009). Doing so might eliminate the need to develop de novo whole genome assemblies and annotations for many species with closely related species for which high quality whole genomes are available. The feasibility of transferring genomic resources within-genera has been demonstrated in agricultural species and in primates. For example, Cosart *et al*. 2011 used an exon capture array designed from the domestic cow (*Bos taurus*) reference genome to successfully capture and sequence thousands of exons in a wild bison (*Bos bison*). Additionally, a microarray designed from human sequence was used to capture Neanderthal exons (Burbano *et al.* 2010). These examples suggest

that genomic resources developed for other agricultural species, such as domestic sheep (*Ovis aries*) could be used to advance our understanding of the genetic basis of adaptation in wild sheep.

Bighorn sheep (*Ovis canadensis*) are probably the best studied of all wild sheep. Several important sources of natural selection in bighorn have been identified by previous research. Genome scans for selection signatures in bighorn could identify candidate genes for phenotypic traits known to affect fitness. Additionally, outlier loci could uncover unexpected genetic and physiological pathways that are important to fitness and adaptation in bighorn. Bighorn often occupy extreme environments including steep terrain, high elevation, and low temperatures and deep snow during winter (Geist 1971). Bighorn sheep have a polygynous mating system with intense male-male competition for access to mates (Geist 1971; Hogg 1984). Male reproductive success is strongly affected by the size of horns (Coltman *et al.* 2002) which are used in physical combat and in dominance displays (Geist 1966). Additionally, bighorn are often exposed to diseases originating in domestic livestock which can cause massive population die offs (Buechner 1960; Monello *et al.* 2001). Lastly, body mass is thought to strongly affect juvenile survival and subsequent reproduction in adult bighorn (Festa-Bianchet *et al.* 1997; Festa-Bianchet *et al.* 2000). Thus, genomic regions bearing genes involved with adaptation to extreme environments, sexual selection (e.g., horn development), disease resistance, and growth are likely to show signatures of selection in bighorn (e.g., very low heterozygosity).

Distinguishing the genomic signatures of selection and genetic drift could be challenging in many bighorn sheep populations. Bighorn sheep have a naturally fragmented distribution, and often occur in small isolated populations (Geist 1971; Valdez & Krausman 1999). Therefore, large chromosomal segments could potentially drift to high frequency, closely mimicking the molecular genetic patterns caused by a response to strong directional selection (i.e., a selective sweep). Thus, the genomic signatures of selection could be difficult to distinguish from the effects of genetic drift due to the naturally fragmented distribution and small local effective population size (N_e) typical of bighorn sheep.

In this study we addresses the following questions: What genomic regions bear signatures of directional selection in bighorn sheep? Do the genomic regions showing signatures of selection vary among populations with differences in elevation and recent exposure to disease?

Materials and Methods

Study populations

Our study included four native bighorn sheep populations. First, we sampled bighorn sheep from the Teton Range in Grand Teton National Park in Northwestern Wyoming. Bighorn sheep in the Teton Range are split into genetically differentiated Northern (NT) and Southern (ST) subpopulations ($F_{ST} = 0.1$, based on analysis of whole genome sequence data, Table 5-S3). NT and ST have no known history of disease-related population declines and have tested negative for antibodies to many common pathogens in 2008 (data not shown). However, domestic sheep grazed in the Tetons in the early 1900s (Whitfield 1983). Therefore it is possible that bighorn in the Tetons may have been affected by diseases originating in domestic sheep at some time in the

early post-settlement era (Whitfield 1983). The Teton bighorn populations occupy high elevation ranges above 2700 m in both Summer and Winter (Whitfield 1983). However, sheep in the Teton Range occupied lower elevation Winter ranges before approximately 1940 (Whitfield 1983). Thus the Teton populations provide the opportunity to detect candidate regions/loci for recent adaptation to high elevation. We created a DNA pool from 9 individuals from ST and a DNA pool composed of 10 individuals from NT.

Our 'Sun River' (SR) samples are from the Sun River Game Preserve in North-Western Montana. We sampled two groups of bighorn sheep with the SR (Gibson Dam, and 'other'). There have been several documented population die-offs associated with pneumonia at SR in 1925, 1927, 1932, and 1983 (Rush 1927; Andryk & Irby 1984). SR bighorn occupy relatively low elevation habitats of appx. 1500 m for at least part of the Winter, and higher ranges during the Summer. Our SR samples were collected in 1990. We created and DNA pool of 10 individuals from the 'other' group of sheep and a pool of 8 individuals from the Gibson Dam group of sheep.

Our Whiskey Basin (WB) study population is located in the Wind River Range of central Wyoming. WB also has a documented history of population die offs associated with pneumonia (e.g., 1991). Whiskey basin bighorn are mostly migratory, occupying distinct high elevation Summer ranges and lower elevation Winter ranges. We created one DNA pool from 10 individuals sampled in 1989, and second DNA pool from 11 additional individuals sampled in 2012.

Sequencing and genome alignment

The quality of DNA from each individual was assessed by separation on agarose gels before sample concentration was measured using the Nanodrop ND-1000 spectrophotometer (Wilmington, DE). Individual samples were normalized to 50 ng/ul, before equivalent amounts were mixed to form pools. To commence library construction, 2.5 ug of pooled DNA was sheared to an average insert length of 300 – 400 bp using a Covaris S220 (Woburn, MA, USA). Short insert libraries were prepared as described by the Illumina TruSeq sample preparation guide v2 (Illumina Inc, San Diego, CA). Sequencing was performed on an Illumina HiScan machine to generate 100 bp paired end (PE) reads.

Quality trimming was performed on raw sequence reads in three steps using Trimmomatic (Lohse *et al.* 2012). First, leading and trailing bases with quality score (QS) < 5 were removed. Secondly, average QS was calculated in 4 bp sliding windows. Bases were trimmed from the point in the read where average QS dropped below 15. Finally, reads were excluded that had < 50 bp following trimming. Quality trimmed reads were mapped against the domestic sheep reference genome assembly v3.1 (OARv3.1) using BWA (Li and Durban, 2009). OARv3.1 was obtained at www.livestockgenomics.csiro.au. The BWA-backtrack algorithm was used with default settings for the maximum allowable mismatches, mismatch and gap penalties. Averaged across the six libraries, 78% of raw reads from bighorn were successfully mapped to the domestic sheep reference assembly (Table 5-S2). Mapping raw reads from domestic sheep against OARv3.1 using the same method resulted in successful alignment of approximately 85%

of reads (Kijas *et al*, unpublished). This suggests around 5 – 6% of Bighorn reads failed to align due to sequence divergence between the species.

Using a reference genome from a different species is a potential source of error in our data. However we believe that this is unlikely to be a substantial source of error because the other species is closely related (approximately 2 MY divergent) (Forbes *et al.* 1995). Poissant *et al*. (2010) showed that synteny was very similar between bighorn and domestic sheep. For example, among 250 microsatellites, Poissant *et al*. (2010) found only three markers that mapped to different relative genomic positions in bighorn compared to domestic sheep. This suggests that genomic architecture is highly conserved between bighorn and domestic sheep. Additionally a large fraction of sequence reads mapped to the domestic sheep genome with high confidence. Thus we believe genomic differences between bighorn and the domestic reference genome are unlikely to have dramatically affected our results.

Variant calling

We merged the mapped sequence reads from the SR Gibson Dam, SR 'other', WB new, and WB old to create a single pool of mapped sequence reads (SR/WB pool). We also merged the mapped sequence reads from NT and ST to create a second pool of mapped sequence reads (Teton pool). We then conducted variant calling on the SR/WB and Teton Pools separately.

Variants in NGS sequence are typically detected as alternate bases during the comparison of reads from a given individual against a reference genome. In this study, reads from pooled samples were used for variant detection against the reference genome from a related species. Variant detection was performed using the SNVer program which was designed for use with pooled NGS reads (Wang *et al.* 2012). Particular attention was given to the consequence of performing variant detection in bighorn reads using the domestic sheep reference genome. Testing revealed modification was needed to the mapping quality (mq) and base quality thresholds (bq) to accurately call and remove fixed differences between species that are monomorphic within bighorn (Figures 5-S1 & 5-S2). Following optimization, variants were called using mq = 40 (default mq = 20), bq = 2 (default bq = 17) and default settings for all other variables.

We summed the allele counts from the SR/WB and Teton pools to create an 'all populations' pool. From these summed allele counts we were able to estimate allele frequencies and test for selective sweeps across all study populations.

SNP filtering

Post-processing filters were applied to putative variants called by SNVer. First, to ensure sufficient reads were available to estimate allele frequencies, positions covered by < 19 reads (summed across the SR/WB and Teton pools) were excluded. To exclude variants likely to be located within structural variation (e.g., CNV), positions covered by > 100 reads (summed across the SR/WB and Teton pools) were excluded. Variants with a minor allele frequency < 0.05 were excluded. We identified 2.57 million SNPs after all filtering steps (Table 5-S1). Variant calling in 68 domestic sheep genomes returned an average of 7 M SNP per individual (Kijas *et al*,

unpublished). The finding that only 2.57 million SNP were called in our study suggests the variant calling and SNP filtering applied was highly stringent.

Identification of selective sweeps

We tested for selective sweep signatures in sequence data pooled from multiple populations with similar selection histories. The chance fixation of the same chromosome segment due to genetic drift is unlikely to occur in multiple populations. Thus analyzing sequence data from multiple populations pooled with similar selection backgrounds should reduce the likelihood of detecting selection signatures that are really caused by drift. The Teton pool provides the opportunity to test for selective sweeps associated with recent adaptation to high elevation. The SR/WB pool provides the opportunity to test for selective sweeps associated with recent exposure to disease. We analyzed our 'all populations' pool in order to detect genomic regions that have responded similarly to directional selection in all or most study populations.

We used two approaches to identity genomic signatures of selection. First, we used sliding window estimates of expected heterozygosity (H_P) (Axelsson *et al.* 2013; Rubin *et al.* 2010) to identify genomic regions with very low heterozygosity relative to the genome-wide average (putative selective sweeps) within each DNA pool. We used 100 Kb sliding windows with a 50 Kb step size. Extremely low heterozygosity is indicative of a response to directional selection at one or more loci within a region. Estimates of allele frequencies are more precise among SNPs with higher read depth. Thus SNPs with higher read depth are weighted so that they have a larger effect on H_P than SNPs with lower read depth. H_P was calculated within each individual pool, independent of data from the other pool. Therefore analyses based on H_P are likely to detect selection signatures that may be unrelated to the biological contrasts (e.g., elevation and disease) between pools.

We also estimated F_{ST} (Weir & Cockerham 1984) for each window in order to identify genomic regions with very different allele frequencies between the two pools. F_{ST} measures genetic differentiation between pools. Therefore selection signatures identified on the basis of F_{ST} outliers are likely to be enriched for signatures present as a result of biological contrasts (e.g., elevation and disease). F_{ST} was statistically significantly negatively correlated with H_P in both the Teton and SR/WB pools ($P < 0.001$ linear regression, Figure 5-S3). Windows containing few SNPs could have high false positive rates for selection signatures because the probability of all SNPs being homozygous in small windows can be high in non-selected genomic regions. Therefore we restricted our analyses to windows containing 10 or more SNPs.

We wanted to identify windows with values of H_P or F_{ST} that deviated substantially from the genome-wide average. Thus, we Z-transformed estimates of H_P and F_{ST} (ZH_P and ZF_{ST}) (Axelsson *et al.* 2013; Qanbari *et al.* 2012; Rubin *et al.* 2010). We focus on windows with H_P or F_{ST} at least five standard deviations away from the genome-wide average ($ZH_P \leq -5$ or $ZF_{ST} \geq 5$) as this represents the extreme ends of the distributions of ZH_P and ZF_{ST}.

We tested if any particular classes of genes (e.g., immune function genes) were represented within outlying windows more often than expected by chance. Therefore we tested the gene content of outlier windows for gene ontology (GO) term enrichment using the software

GOrilla as described by Eden *et al*. (2009). A background set of 15,186 genes with associated GO terms were used to evaluate genes identified by either ZH_P or ZF_{ST}.

Differentiating genomic signatures of selection, drift, and low mutation rate

It is difficult to define threshold values of ZH_P and ZF_{ST} that would confidently exclude false selection signatures that are really caused by drift (see Discussion). Separating the effects of drift and selection can be particularly difficult when N_e is small and when there is a very limited information on historical population dynamics.

We used simulations to qualitatively evaluate the possibility that putative selective sweeps were caused by genetic drift rather than selection. Our simulation model is described in detail elsewhere (Kardos *et al.* 2013). Briefly, we used the program R to simulate populations with approximately similar N_e and levels of population connectivity as the Teton subpopulations. We chose to use simulations of the Teton populations because they are the smallest of our study populations, and presumably have very recent common ancestry. Therefore, false positive selection signatures were more likely to occur in the Teton pool than in the SR/WB pool. Estimates of N_e (Kardos *et al*. unpublished data) for NT ($N_e = 15$), and ST ($N_e = 20$) were based on 22 microsatellite loci and a linkage disequilibrium (LD) estimator of N_e (Hill 1981; Waples 2006). We chose to simulate N_e of 20 and 30 for the Teton populations because the LD estimator of N_e is likely to be downwardly biased when multiple cohorts are included in a sample (Luikart *et al.* 2010).

We simulated the approximate genomic characteristics of bighorn sheep, and assumed a 'best guess' demographic scenarios for bighorn in the Teton Range. The assumptions of the simulations include 1) a homogeneous recombination rate of 1.24 cM/Mb (Poissant *et al.* 2010); 2) moderate genetic differentiation between the North and South Teton populations ($F_{ST} \approx 0.1$, based on out whole genome sequence data, Table 5-S3); 3) a small amount of immigration to the Teton populations from a large source population (0.5 immigrants/generation on average); 4) a similar density of SNPs as in our empirical data (1.05 SNP/Kb); and 5) an effective population size of N_e = 20-30 for each of the Teton subpopulations. We ran twenty replicate simulations for $N_e = 20$ and $N_e = 20$ separately. We pooled the simulated sequence data and tested for ZH_P outlier windows in the same way as we did with our empirical data. Our model of recombination is based on Fisher's theory of junctions (Fisher 1965), and assumes no interference. For computational efficiency we simulated genomes with two 123 Mb chromosomes (total map length = 305.1 cM), which is 10% the sex averaged map length estimated for bighorn (Poissant *et al.* 2010). We wanted to simulate a similar density of SNPs as was observed in our empirical sequence data from bighorn. We simulated 257,000 SNPs that were randomly distributed throughout the genome using a random number generator. We tested for selective sweeps in the simulated data using the same methods as for our empirical sequence data.

We used the simulations only to qualitatively evaluate whether some of the selection signatures we detected were likely to be due to drift. We believe it would be a mistake to use the simulations to assign strict significance threshold values of ZH_P and ZF_{ST} because we have very little information on the demographic history of the study populations.

A locally low mutation rate could cause a chromosome segment to have very low heterozygosity relative to the genome-wide average heterozygosity – creating a false positive signature of selection. Therefore we wanted to reduce the chances of identifying putative selective sweeps that were actually caused by a locally low mutation rate. To identify windows that are likely to have a locally low mutation rate, we calculated the number of fixed differences between bighorn sheep in the Teton populations and the domestic sheep reference genome for each 100 Kb window in the genome (Figures 5-S4 & 5-S5). We excluded windows with fewer than 200 fixed differences between bighorn and domestic sheep from consideration as a putatively selected region.

Results

Illumina sequencing produced a total of 256,408 Mb of 100 bp paired end sequence reads. 78% of the sequence reads aligned with high confidence to the domestic sheep reference genome. We identified 2.57 million SNPs after quality control filtering. Mean sequence read depth across the 2.57 M SNPs was 32.6 for the Teton Pool and 34.7 for the SR/Wb Pool. Mean H_P for the Teton and SR/WB pools was 0.31 and 0.34, respectively.

Identification of putative selection signatures

We identified a total of 83 genomic regions showing signatures of selection (Figure 1, Tables 5-S5, 5-S6, 5-S7). 56 of the regions with putative signatures of selection contained genes (Table 5-S4). There were six genomic regions with $ZH_P \leq -5$ that contained at least one gene in the SR/WB pool. There were 28 genomic regions with $ZH_P \leq -5$ that contained at least one gene in the Teton pool. 22 putatively selected genomic regions that contained at least on gene were identified on the basis of $ZF_{ST} \geq 5$ (Table 5-S4). We identified one genomic region in our 'all populations' pool with $ZH_P \leq -5$ (Figure 1). Summary statistics for each putatively selected window, including gene content, are provided in Tables 5-S4 – 5-S7. We focus on selection signatures found in three genomic regions that contain genes, and have clear biological interpretations given what is known about our study system and bighorn life history.

Three putatively selected windows contained genes related to horn development or early body growth. A region on chromosome 10 with very low heterozygosity in the Teton pool ($ZH_P = -6.05$), and in the 'all populations' pool ($ZH_P = -7.4$) was located over the relaxin/insulin-like family peptide receptor 2 gene (*RXFP2*) (Figures 5-1 & 5-2). ZH_P did not reach the threshold of $ZH_P \leq -5$ in the window containing *RXFP2* in the SR/WB pool. However it is notable that ZH_P was -3.65 for this window over *RXFP2* in the SR/WB pool, which represents the bottom 0.18% of all ZH_P windows across the genome. It is notable that the only outlying windows identified in out 'all populations' were over and adjacent to *RXFP2* on chromosome 10 (Figure 5-1). *RXFP2* has been found to control the presence and size of horns in domesticated sheep (Johnston *et al.* 2010; Johnston *et al.* 2013; Kijas *et al.* 2012).

We also detected selection signatures in two genomic regions harboring genes involved with body growth. A putatively selected region on chromosome 16 in the Teton pool contained the growth hormone receptor gene (*GHR*, Figures 5-1 & 5-2, Table 5-S4). The region around *GHR* contained windows with both very high F_{ST} ($ZF_{ST} = 5.01$) and low H_P ($ZH_P = -5.27$) in the

Teton pool (Figure 5-2, Table 5-S4). *GHR* plays an important role in mediating the physiological effects of growth hormone (Argetsinger & Carter-Su 1996). Variation at *GHR* has been found to affect the expression of insulin-like growth factor I in cattle (Ge *et al.* 2003), and milk and fat production in cattle (Blott *et al.* 2003; Viitala *et al.* 2006). We additionally detected a putative selective sweep over the insulin-like growth factor 2 receptor on chromosome 8 (*IGF2R*, Figure 5-2). *IGFR2* was in a window with ZH_P = -5.62 (Table 5-S4). *IGF2R* is thought to regulate fetal growth by binding and degrading the growth hormone *IGF2* (Monk & Moore 2004). Additionally, *IGF2R* genotypes have been associated with growth traits in cattle (Berkowicz *et al.* 2012)

The lengths of the genomic regions encompassing outlier windows was highly variable. Several of the outlier regions included a single 100Kb window while others included several 100 Kb windows. For example, there was a 4 Mb outlier genomic region on chromosome 2 that contained six genes (Figures 5-1 & 5-2).

Several gene ontology terms among the putatively selected windows were statistically significantly enriched (Table 5-S5). However all of the statistically significant terms were located in a single region on chromosome 5. Therefore we believe that this apparent enrichment is spurious.

Differentiating signatures of selection from signatures of drift or low mutation rate

We detected one genomic region with $ZH_P < -5$ across 20 replicate simulations (40 chromosomes) of neutrally evolving populations with $N_e = 30$. We detected three genomic regions with $ZH_P < -5$ in our neutral simulations of populations with local $N_e = 20$. We excluded two ZH_P outlier windows (on chromosomes 4 and 22) in the Teton pool that had < 200 fixed differences between bighorn and domestic sheep (Figure 5-S4).

Discussion

We identified several candidate selected genomic regions in bighorn sheep. In particular, we identified putative selection signatures in genomic regions harboring genes related to body and horn growth.

Selection signatures at RXFP2

We identified a selection signature (very low ZH_P) around the *RXFP2* gene in the Teton pool and in our 'all populations' pool. We believe this is strong evidence that *RXFP2* was subjected to strong directional selection in our study populations. *RXFP2* has been found to strongly affect horn development in domesticated sheep (Johnston *et al.* 2013; Johnston *et al.* 2011; Kijas *et al.* 2012). For example, *RXFP2* was a highly statistically significant F_{ST} outlier between horned and hornless domestic sheep (Kijas *et al.* 2012). Additionally, both horn length and size have mapped to the genomic region containing *RXFP2* in Soay sheep – an ancient, free-ranging domesticated breed (Johnston *et al.* 2010). *RXFP2* is thought to be under strong balancing selection in Soay sheep (Johnston *et al.* 2013). In particular Johnston *et al*. (2013) found a reproductive advantage of an *RXFP2* allele associated with large horns, and a survival advantage of an allele associated

with small horns. The net effect of directional selection operating in opposing directions for different alleles has apparently maintained heritable genetic variation for horn size in Soay sheep (Johnston *et al.* 2013).

Bighorn sheep have extremely large, energetically costly horns which can comprise up to 8-12% of body mass in older rams (Geist 1966). These massive horns play a crucial role in dominance interactions and physical combat during competition for mates (Geist 1966; Geist 1971). Horn size strongly affects male fitness and has been found to be heritable and to respond to directional selection in bighorn (Coltman *et al.* 2002; Coltman *et al.* 2003). Similar selective pressures to those observed in Soay sheep could explain the maintenance of heritability for horn size in bighorn sheep (Coltman *et al.* 2002; Coltman *et al.* 2003). Bighorn rams must make heavy energetic investments in the development of very large horns. Thus there could be substantial survival costs to having large horns due to the high energetic costs related to horn development. However, there is obviously strong sexual selection favoring males with very large horns in bighorn sheep.

In contrast to Soay sheep, our results (i.e., a putative selective sweep over *RXFP2*) suggest that directional selection is the dominant evolutionary force affecting *RXFP2*, and presumably horn development, in our bighorn populations. F_{ST} was lower than the genome-wide average in both of the sliding windows that overlapped *RXFP2* (ZF_{ST} = -0.72 and -1.13). Low genetic differentiation in the region of *RXFP2* suggests that the same alleles were favored in populations comprising the Teton and SR/WB pools.

An obvious potential explanation for strong directional selection on horn development is selective human harvest. It has been shown that selective harvest of large-horned rams can result in evolution toward smaller horn size in bighorn sheep (Coltman *et al.* 2003; Hedrick 2011). We have no data on horn size in any of our study populations, so it is impossible for us to explicitly test this idea. *RXFP2* should be considered as a strong candidate gene in future studies of the genetic basis of horn development, male dominance, and reproductive success in mountain sheep.

Selection signatures at growth genes

We identified putative selection signatures at two genes involved with body growth: the growth hormone receptor (*GHR*) and insulin-like growth factor 2 receptor (*IGF2R*). *GHR* is a receptor for growth hormone (*GH*). The binding of *GH* to *GHR* initiates a metabolic cascade that triggers and regulates the expression insulin-like growth factor I (*IGF-I*) (Frago & Chowen 2005). The bulk of the effects of *GH* and GHR on the growth via the production of *IGF-I* are thought to occur during postnatal development (Frago & Chowen 2005). Variation at *GHR* has been found to affect the expression *IGF-I*, and milk fat percentage in cattle (Aggrey *et al.* 1999; Ge *et al.* 2003). The genomic region around *GHR* had both very low heterozygosity (ZH_P = -5.27) in the Teton pool and high F_{ST} (ZF_{ST} = 5.01) (Table 5-S4). Thus, it appears that an allele(s) at *GHR* was strongly favored in the Teton subpopulations but not in the SR and WB populations. This suggests that strong selection on GHR in the Tetons could be due to biological contrasts between the Teton and SRWB pools. For example, it is likely that lambs with very fast body growth have

a large selective advantage in high elevation habitats like to the Teton Range with a short growing season and extreme Winter conditions.

The selection signature around *IGF2R* could be caused by its effects on prenatal growth. *IGF2R* is thought to regulate fetal growth by binding and degrading the growth hormone *IGF2* (Monk & Moore 2004), which modulates fetal and placental size (Constância *et al.* 2002). Survival of bighorn lambs and subsequent reproduction of adults has been shown to be related to growth early in life (Festa-Bianchet *et al.* 1997; Festa-Bianchet *et al.* 2000). Thus, *IGF2R* and *GHR* should be considered as a strong candidate gene for body mass of bighorn lambs, and subsequent reproduction of adults.

Differentiating the genomic signatures of selection and drift

Differentiating signatures of selection from similar molecular patterns caused be genetic drift is challenging (Jensen *et al.* 2007; Pavlidis *et al.* 2012). Some approaches have been devised to limit or quantify the likelihood of false positive signatures of selection (Hohenlohe *et al.* 2010; Nielsen *et al.* 2005; Pavlidis *et al.* 2012; Qanbari *et al.* 2012). However, these approaches either require knowledge of historical population dynamics (e.g., Neilsen *et al*. 2005) or assume that the minor allele frequencies at closely linked loci are independent (e.g., Hohenlohe *et al*. 2010; Qanbari *et al*. 2012). The approach of Neilsen *et al*. 2005 compares a test statistic (e.g., the composite likelihood ratio) against its null distribution taken from simulated population genomic data with no selection. In order to implement this approach, the researcher must have reasonable knowledge of the population history, including historical colonization and migration events, and fluctuations in population size. Unfortunately, such detailed historic demographic information is rarely available for natural populations.

There is very limited information on historical population dynamics in our study populations. Thus, we believe it would have been inappropriate to use simulations to define strict threshold values of ZH_P and ZF_{ST} beyond which a particular putative selection signature was considered 'significant'. Nevertheless, our simulations suggest that it is likely that some of the putative selection signatures we detected in the Teton pool were false positives. However, the identification of selection signatures around genes related to functions known to strongly affect fitness in bighorn and domestic sheep gives us confidence that many of the putative selection signatures in the Teton pool were real. Given the strong possibility of false positives, the outliers we detected only in the Teton pool (e.g., *GHR* and *IGF2R*) should be considered only as candidate adaptive genes. The functional and fitness roles of these candidate genes should be confirmed by future genomic studies of bighorn and other mountain sheep.

The results of our simulations are also informative of the potential usefulness of randomization procedures to evaluate the likelihood that putative selection signatures were indeed caused by selection. Hohenlohe *et al*. 2010 and Qanbari *et al*. 2012 proposed similar randomization-based tests to statistically differentiate true selection signatures from those caused by genetic drift. This approach first involves simulating the null distribution of a statistic (e.g., H_P of F_{ST}) by estimating the statistic on thousands of sets of randomly chosen SNPs from across the genome. An empirical *P*-value for an observed estimate in a particular genomic window is determined as the proportion of the randomized estimates that are at least as extreme as the

observed value. This approach implicitly assumes that heterozygosity at closely linked loci is uncorrelated (as noted by Hohenlohe *et al*. 2010). However, as shown by our simulations (Figure 5-3), heterozygosity at closely linked loci is not independent. We believe that the non-independence of heterozygosity at closely linked markers should be carefully considered in future studies of genomic signatures of selection. Specifically, the scenarios under which the null distribution of selection test statistics can be approximated via randomization of loci from across the genome may be limited to populations with extremely large N_e (e.g., $N_e >> 1000$). Additionally, every effort should be made to include realistic values of N_e and population admixture when using simulations to approximate the null distribution of selection signature test statistics, or evaluating the performance of a particular test for selection.

Conclusions

We have identified candidate adaptive genes associated with body growth and horn development in wild bighorn sheep. We found evidence of strong directional selection on *RXFP2*, which has been shown to affect horn development in domestic sheep. Thus *RXFP2* has likely been strongly influenced by sexual selection via high increased reproductive success of rams with large horns, or alternatively by selective human harvest of large-horned rams. Selection signatures at the *GHR* and *IGF2R* further support the idea that growth is under strong selection in bighorn. Additionally, finding selection signatures at *GHR* and *IGF2R* suggests that these genes are responsible for heritable variation in growth. The best way to validate the fitness effects of these genes would be to focus on them in future studies of body and horn growth, and scans for selection signatures in other populations of mountain sheep. Lastly, this study demonstrates the power of whole genome sequencing when a high quality reference sequence is available, e.g., from a related model or agricultural species. Whole genome sequencing with a reference sequence has enormous potential to facilitate identification of selected genomic regions in natural populations.

Acknowledgements

This material is based on work supported by the National Science Foundation through the Montana Institute on Ecosystems (EPS-1101342) and the Montana Ecology of Infectious Diseases IGERT program (DGE-0504628) (MK's support). Agencies that contributed to the project include Montana Fish, Wildlife & Parks; Montana Conservation Science Institute; Grand Teton National Park; Wyoming Game and Fish Department; National Elk Refuge; Bridger-Teton National Forest; Caribou-Targhee National Forest; Wyoming State Vet Lab; and the Wyoming Cooperative Fish and Wildlife Research Unit. Project funding was provided by Grand Teton National Park; Grand Teton National Park Foundation; Wyoming Governors Big Game Licence Coalition; Greater Yellowstone Coordinating Committee; Wyoming Chapter of Foundation for North American Wild Sheep; 1% for the Tetons; National Park Service – Rocky Mountain CESU; National Park Service – BRMD; University of Wyoming/NPS - Research Station; and the Grand Teton Association. We thank Steve Cain, Steve Kilpatrick, Doug Brimeyer, Tim Fuchs, Eric Cole, Walter Scherer, Tim Lyons, Heidi Helling, Aly Courtemanch, Ari La Porte, Leah Yandow, Steve Amish, Sally Painter, and Sam Dwinell for assistance administratively, in the field, and with lab work.

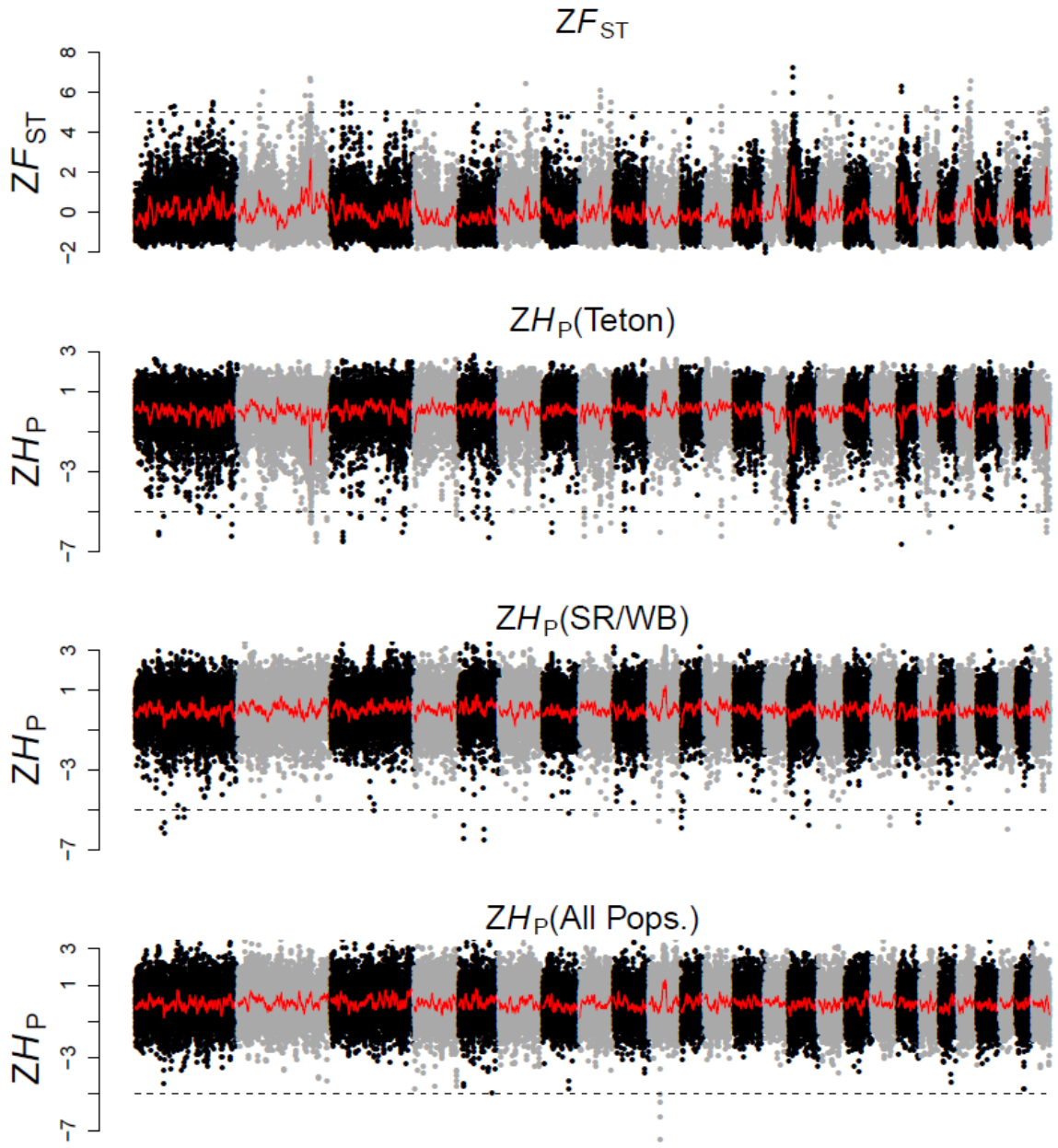

Figure 5-1. ZH_P for the Teton populations (bottom panel) and ZF_{ST} (top panel) across the bighorn sheep genome. Chromosomes are arranged 1-26 (left to right). The red line represents the rolling mean across 100 sliding windows. The dashed lines represent the threshold of significance of 5 standard deviations from the mean H_P and F_{ST} across the whole genome.

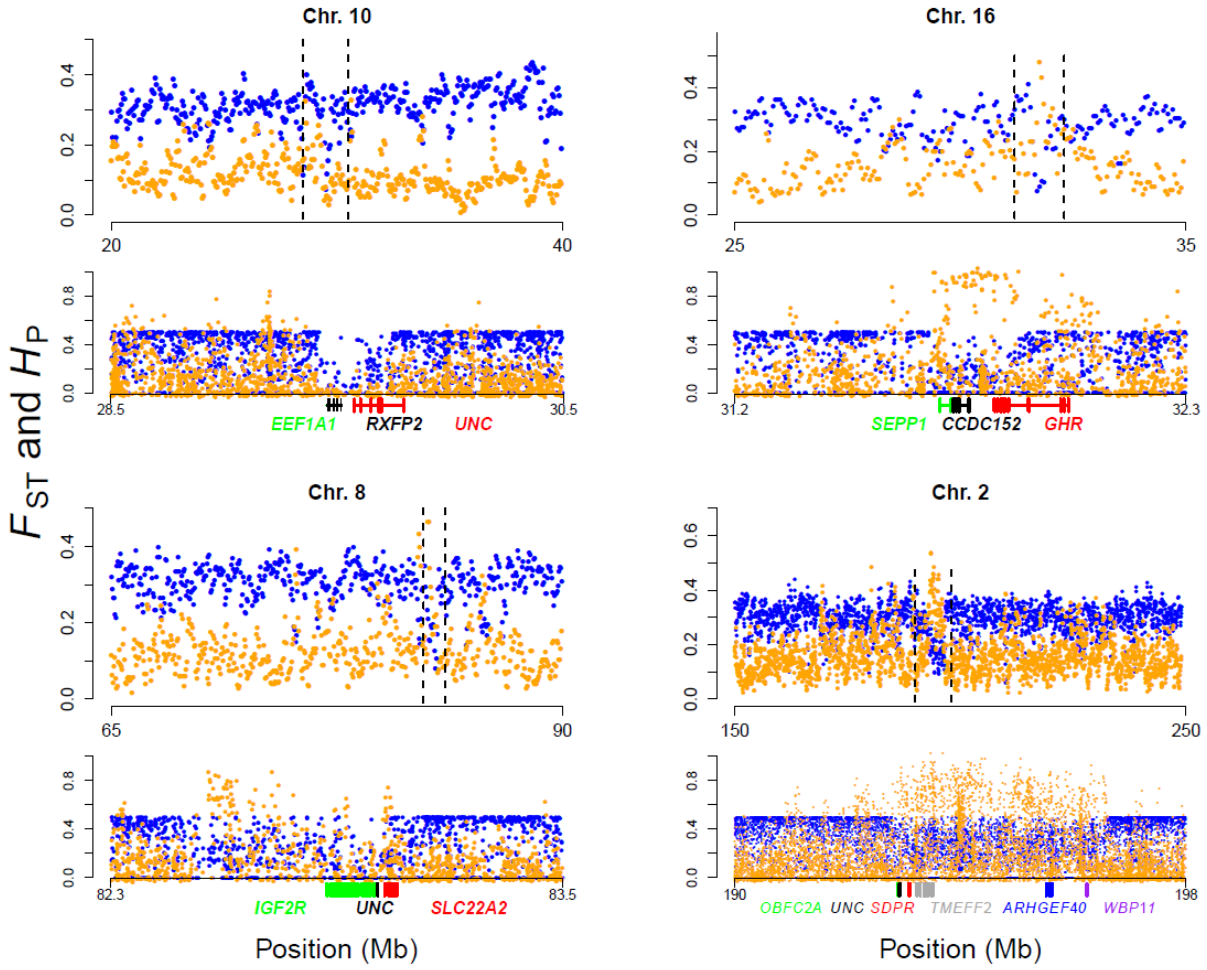

Figure 5-2. H_P in the Teton pool and F_{ST} across putatively selected regions on chromosomes 10 (**A**), 16 (**B**), 8 (**C**), and 2 (**D**). Orange points represent F_{ST} and blue points represent H_P. F_{ST} and H_P are shown in 100Kb windows in the top panels. F_{ST} and H_P at individual SNPs in the putatively selected regions are shown in the bottom panels. Genes labeled as "*UNC*" are uncharacterized.

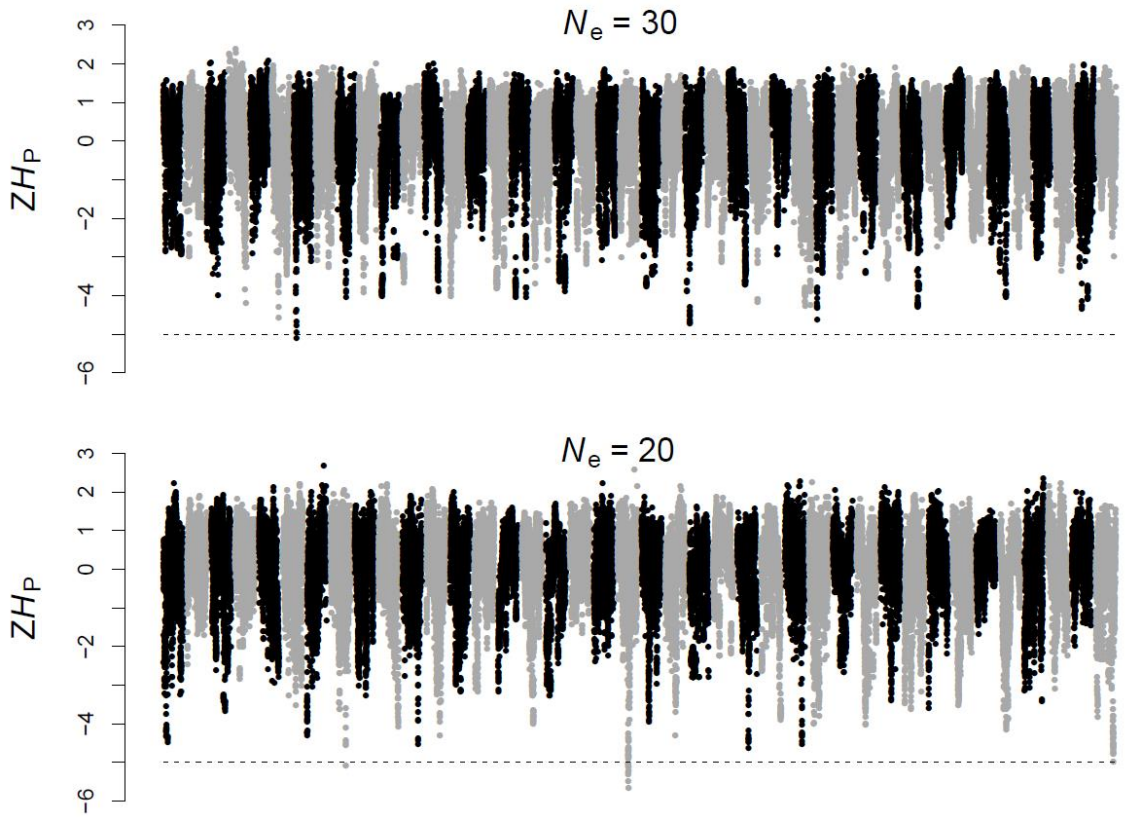

Figure 5-3. ZH_P sliding window estimates from 20 simulations of neutrally evolving populations with a demographic history approximately similar to the Teton populations. Simulations were run for 100 generations with local $N_e = 30$ (top panel) and $N_e = 20$ (bottom panel). Two chromosomes were simulated for each population. The two chromosomes from each simulated population are represented by adjacent blocks of black and gray points. The simulations with 0.5 migrants per generation on average between the two Teton subpopulations and a large source population. There was a migration rate of 0.75 individual/generation on average between the two Teton subpopulations. The mean F_{ST} was 0.09 (min. = 0.05, max. = 0.12) among the 20 simulations of populations with $N_e = 30$. The mean F_{ST} across the 20 simulations with $N_e = 20$ was 0.1 (min. = 0.06, max. = 0.15). The dashed lines represent the threshold value of ZH_P that we used for our empirical data to consider a window as being putatively selected.

LITERATURE CITED

Acevedo-Whitehouse K, Spraker TR, Lyons E, Melin SR, Gulland F, Delong RL, Amos W (2006) Contrasting effects of heterozygosity on survival and hookworm resistance in California sea lion pups. *Molecular Ecology* 15, 1973-1982.

Aggrey SE, Yao J, Sabour MP, Lin CY, Zadworny D, Hayes JF, Kuhnlein U (1999) Markers within the regulatory region of the growth hormone receptor gene and their association with milk-related traits in Holsteins. *Journal of Heredity* 90, 148-151.

Alho J, Lillandt B-G, Jaari S, Merilä J (2009) Multilocus heterozygosity and inbreeding in the Siberian jay. *Conservation Genetics* 10, 605-609.

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

Andryk, TA, and LR Irby (1986) Population characteristics and habitat use by mountain sheep prior to a pneumonia die off. *Proceedings of the Biennial Symposium of the Northern Wild Sheep and Goat Council* 5, 272-291.

Argetsinger LS, Carter-Su C (1996) Mechanism of signaling by growth hormone receptor. *Physiological Reviews* 76, 1089-1107.

Armstrong DP, Cassey P (2007) Estimating the effect of inbreeding on survival. *Animal Conservation* 10, 487-492.

Axelsson E, Ratnakumar A, Arendt M-L*, et al.* (2013) The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360-364.

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

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* 12, 767-780.

Barrett SCH, Harder LD, Worley AC (1996) The comparative biology of pollination and mating in flowering plants. *Philosophical Transactions of the Royal Society of London. Series B* 351, 1271-1280.

Berkowicz E, Magee D, Berry D, Sikora KM, Howard DJ, Mullen MP, Evans RD Spilane C, MacHugh DE (2012) Single nucleotide polymorphisms in the imprinted bovine insulin-like growth factor 2 receptor gene (IGF2R) are associated with body size traits in Irish Holstein-Friesian cattle. *Animal Genetics* 43, 81-87.

Blott S, Kim J-J, Moisio S, Schmidt-Kuntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* 163, 253-266.

Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology & Evolution* 18, 503-511.

Buechner HK (1960) The bighorn sheep in the United States, its past, present, and future. *Wildlife monographs* 4, 3-174.

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.

Chapman N, Thompson E (2003) A model for the length of tracts of identity by descent in finite random mating populations. *Theoretical Population Biology* 64, 141-150.

Charlesworth D, Charlesworth B (1987) Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology and Systematics* 18, 237-268.

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

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

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

Cosart T, Beja-Pereira A, Chen S*, et al.* (2011) Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics* 12, 347.

Constância M, Hemberger M, Hughes J, Dean W, Ferguson-Smith A, Fundele R, Stewart F, Kelsey G, Fowden A, Colin S (2002) Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature* 417, 945-948.

Crow JF, Kimura M (1970) *An introduction to population genetics theory* The Blackburn Press, Caldwell, NJ.

Darwin, C. R. 1868. *Variation of Animals and Plants under Domestication.* London: John Murray

David P (1997) Modeling the genetic basis of heterosis: tests of alternative hypotheses. *Evolution* 51, 1049-1057.

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.

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.

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

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics 10,* 48.

Efron B, Tibshirani RJ (1994) *An introduction to the bootstrap.* Chapman & Hall/CRC, Boca Raton, Florida.

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

Festa-Bianchet M, Jorgenson JT, Bérubé CH, Portier C, Wishart WD (1997) Body mass and survival of bighorn sheep. *Canadian Journal of Zoology* 75, 1372-1379.

Festa-Bianchet M, Jorgenson JT, Réale D (2000) Early development, adult mass, and reproductive success in bighorn sheep. *Behavioral Ecology* 11, 633-639.

Fischer MC, Rellstab C, Tedder A*,* Zoller S, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2013) Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. *Molecular Ecology* 22, 5594-5607.

Fisher RA (1965) *The theory of inbreeding*. Oliver and Boyd, Edinburgh.

Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. *Mol Biol Evol* 12, 1106-1113.

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.

Frago LM, Chowen JA (2005) Basic physiology of the growth hormone/insulin-like growth factor (IGF-) axis. In: *Advances in experimental medicine and biology* (eds. Valrela-Nieto I, Chowen JA). Springer, New York.

Franklin I (1977) The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theoretical Population Biology* 11, 60-80.

Gage MJ, Surridge AK, Tomkins JL, Green E, Wiskin L, Bell DJ, Hewitt GM (2006) Reduced heterozygosity depresses sperm quality in wild rabbits, *Oryctolagus cuniculus*. *Current Biology* 16, 612-617.

Ge W, Davis M, Hines H, Irvin K, Simmen R (2003) Association of single nucleotide polymorphisms in the growth hormone and growth hormone receptor genes with blood serum insulin-like growth factor I concentration and growth traits in Angus cattle. *Journal of animal science* 81, 641-648.

Geist V (1966) The evolutionary significance of mountain sheep horns. *Evolution* 20, 558-566.

Geist V (1971) *Mountain sheep: a study in behavior and evolution.* University of Chicago Press, Chicago.

Grant PR., Grant BR, Petren K (2001) A population founded by a single pair of individuals: establishment, expansion, and evolution. *Genetica* 112-113, 359-382.

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

Harrell FE (2012) Hmisc: Harrell Miscellaneous. R Package version 3.9-3. http://CRAN.R-project.org/package=Hmisc.

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 P, Fredrickson R, Ellegren H (2001) Evaluation of d^2, a microsatelline measure of inbreeding and outbreeding. *Evolution* 55, 1256-1260.

Hedrick PW (2011) Rapid Decrease in Horn Size of Bighorn Sheep: Environmental Decline, Inbreeding Depression, or Evolutionary Response to Trophy Hunting? *Journal of Heredity* 102, 770-781.

Hill WG, Robertson A (1968) The effects of inbreeding at loci with heterozygote advantage. *Genetics* 60, 615-628.

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. *Genetical Research* 38, 209-216.

Hill WG, Weir BS (2011) Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research* 93, 47-64.

Hill WG, Weir BS (2011) Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research* 93, 47-64.

Hogg J (1984) Mating in bighorn sheep: multiple creative male strategies. *Science* 225, 526-529.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics 6*(2), e1000862

Howrigan D, Simonson M, Keller M (2011) Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics* 12, 460.

Jacquard A (1975) Inbreeding: One word, several meanings. *Theoretical Population Biology* 7, 338-363.

Jensen JD, Wong A,Aquadro CF (2007) Approaches for identifying targets of positive selection. *Trends in Genetics*, *23* 568-577.

Jensen-Seaman MI, Furey TS, Payseur BA*, et al.* (2004) Comparative recombination rates in the rat, mouse, and human genomes. *Genome research* 14, 528-538.

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, 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*. doi:10.1038/nature12489

Kardos M, Allendorf FW, Luikart G (2013) Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium? *Molecular Ecology Resources*. DOI: 10.1111/1755-0998.12193

Keller LF (1998) Inbreeding and its fitness effects in an insular population of song sparrows (Melospiza melodia). *Evolution* 52, 240-250.

Kalinowski ST, Wagner AP, Taper ML (2006) ml-relate: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* 6, 576-579.

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution* 17, 230-241.

Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189, 237-249.

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.

Kirin M, McQuillan R, Franklin CS*, et al.* (2010) Genomic runs of homozygosity record population history and consanguinity. *Plos One* 5, e13996.

Kuningas M, McQuillan R, Wilson JF*, et al.* (2011) Runs of Homozygosity Do Not Influence Survival to Old Age. *Plos One* 6, e22580.

Leutenegger A-L, Labalme A, Génin E*, et al.* (2006) Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: application to Taybi-Linder syndrome. *The American Journal of Human Genetics* 79, 62-66.

Levine R (2011) i5k: The 5,000 Insect Genome Project. *American Entomologist* 57, 110-113.

Lohse M, Bolger AM, Nagel A*, et al.* (2012) RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic acids research* 40, W622-W627.

Luikart G, Cornuet J-M, Allendorf FW, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* 89, 238-247.

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics* 11, 355-373.

Luquet E., Léna L-P, David P, Prunier J, Joly P, Lengagne T, Perrin N, Plénet S (2012) Within- and among-population impact of genetic erosion on adult fitness-related traits in the European tree frog *Hyla arborea*. *Heredity* 110, 347-354.

Madsen T, Shine R, Olsson M, Wittzell H (1999) Conservation biology: Restoration of an inbred adder population. *Nature* 402, 34-35.

Manly BFJ (2007) Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall/CRC, Boca Raton.

McQuillan R, Leutenegger AL, Abdel-Rahman R*, et al.* (2008) Runs of homozygosity in European populations. *The American Journal of Human Genetics* 83, 359-372.

McQuillan R, Eklund N, Pirastu N*, et al.* (2012) Evidence of inbreeding depression on human height. *PLoS Genetics* 8, e1002655.

Mills LS, Smouse PE (1994) Demographic Consequences of Inbreeding in Remnant Populations. *The American Naturalist* 144, 412-431.

Monello RJ, Murray DL, Cassirer EF (2001) Ecological correlates of pneumonia epizootics in bighorn sheep herds. *Canadian Journal of Zoology* 79, 1423-1432.

Monk D, Moore GE (2004) Intrauterine growth restriction-genetic causes and consequences. *Seminars in Fetal and Neonatal Medicine* 9, 371-378.

Morton NE, Crow JF, Muller HJ (1956) An estimate of the mutational damage in man from data on consanguineous marriages. *Proceedings of the National Academy of Sciences USA* 42, 855-863.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C (2005) Genomic scans for selective sweeps using SNP data. *Genome research 15*, 1566-1575.

O'Grady JJ, Brook BW, Reed DH*, et al.* (2006) Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological Conservtion* 133, 42-51.

Ohta T (1971) Associative overdominance caused by linked detrimental mutations. *Genetical Research* 18, 277-286.

Ouberg NJ (2010) Integrating population genetics and conservation biology in the age of genomics. *Biology Letters* 6, 3-6.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution 29*, 3237-3248.

Pemberton J (2004) Measuring inbreeding depression in the wild: the old ways are the best. *Trends in Ecology & Evolution* 19, 613-615.

Pemberton JM (2008) Wild pedigrees: the way forward. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 275, 613-621.

Poissant J, Hogg J, Davis C*, et al.* (2010) Genetic linkage map of a wild genome: genomic structure, recombination and sexual dimorphism in bighorn sheep. *BMC Genomics* 11, 524.

Pollinger JP, Earl DA, Knowles JC*, et al.* (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome research* 21, 1294-1305.

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

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.

Qanbari S, Strom TM, Haberer G, Weigend S, Gheyas AA, Turner F, Simianer H (2012) A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens. *PloS one 7*(11), e49525.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* 94, 9197-9201.

Reid JM, Arcese P, Keller LF, Elliott KH, Sampson L, Hasselquist D (2007) Inbreeding effects on immune response in free-living song sparrows (*Melospiza melodia*). *Proceedings of the Royal Society of London. Series B: Biological Sciences* 274, 697-706.

Robinson SP, Simmons LW, Kennington WJ (2013) Estimating relatedness and inbreeding using molecular markers and pedigrees: the effect of demographic history. *Molecular Ecology*: doi: 10.1111/mec.12529

Rubin C-J, Zody MC, Eriksson J*, et al.* (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 464, 587-591.

Rudnick JA, Lacy RC (2008) The impact of assumptions about founder relationships on the effectiveness of captive breeding strategies. *Conservation Genetics* 9, 1439-1450.

Ruiz-López MJ, Gañan N, Godoy JA, Olmo AD, Garde J, Espeso G, Vargas A, Martinez F, Roldán ERS, Gemendio M (2012) Heterozygosity-fitness correlations and inbreeding depression in two critically endangered mammals. *Conservation Biology* 26, 1121-1129.

Saccheri I, Kuussaari M, Kankare M*, et al.* (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature* 392, 491-494.

Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative

genetics: finding the genes underlying ecologically important traits. *Heredity* 100: 158-170.

Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, McEwan JC (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* 93, 255-265.

Stam P (1980) The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genetics Research* 35, 131-155.

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

Templeton AR, Read B (1994) Inbreeding: One word, several meanings, much confusion. In: *Conservation Genetics* (eds. Loeschcke V, Tomiuk J, Jain SK), pp. 91-106. Birkhäuser-Verlag, Basel.

Therneau T, Atkinson E, Sinnwell J*, et al.* (2011) kinship2: Pedigree functions. R package version 1.3.3. http://CRAN.R-project.org/package=kinship2.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14, 178-192.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. *Nature genetics* 42, 260-263.

Valdez R, Krausman P (1999) *Mountain sheep of North America*. University of Arizona Press.

Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Maki-Tanila A, Georges M, Vikki J (2006) The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. *Genetics* 173, 2151-2164.

Wagner AP, Creel S, Kalinowski S (2006) Estimating relatedness and relationship using microsatellite loci with null alleles. *Heredity* 97, 336-345.

Wang J (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genetics Research* 89, 135-153.

Wang J (2011) COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources* 11, 141-145.

Wang W, Hu W, Hou F, Hu P, Wei Z (2012) SNVerGUI: a desktop tool for variant analysis of next-generation sequencing data. *Journal of medical genetics* 49, 753-755.

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci*. *Conservation Genetics* 7, 167-184.

Weir BS, Cockerham CC (1973) Mixed self and random mating at two loci. *Genetical Research* 21, 247-262.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358-1370.

Weir BS, Anderson AD, Hepler AB (2006) Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics* 7, 771-780.

Westemeier RL, Brawn JD, Simpson SA*, et al.* (1998) Tracking the long-term decline and recovery of an isolated population. *Science* 282, 1695-1698.

Whitfield, MB (1983) Bighorn sheep history, distributions, and habitat relationships in the Teton Mountain Range, Wyoming. M.S. Thesis, Idaho State University, Pocatello.

Wright S (1922) Coefficients of inbreeding and relationship. *The American Naturalist* 56, 330-338.

Wright S (1931) Evolution in Mendelian populations. *Genetics* 16, 97-159.

Zheng B (2000) Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine* 19, 1265-1275.

Appendix 2-1

The power to detect identity disequilibrium and inbreeding depression via HFCs is expected to be affected not only by $\sigma^2(F)$, but also by the heterozygosity of the typed loci (Slate *et al.* 2004); therefore we controlled for heterozygosity in our simulations. We used the mean F of individuals in generations 49-50 of the simulated populations to determine how many equally frequent alleles in the first generation were necessary to produce expected heterozygosity (H_e) of 0.65 for loci sampled in generations 49-50.

We determined the mean F in generation 49-50 of the simulated populations as the observed mean F in generation 50 among 100 replicate preliminary simulations (parameterized as described above) using each of the eight immigration rates (m). Expected heterozygosity (H_e) can be determined as

$$H_e = H_0(1- F) \qquad \text{eq. 1}$$

where H_0 is the mean heterozygosity of non-inbred individuals and F is the mean IBD fraction of the genome among individuals in the population (Crow & Kimura 1970, p. 66). We used equation 1 to determine the expected heterozygosity of immigrants and individuals in the first simulated generation (H_0) needed to achieve $H_e \approx 0.65$ in generations 49-50. Therefore we solved for H_0 after substituting the mean F in generations 49-50 and the desired H_e of 0.65. We then determined the number of equally frequent alleles necessary to achieve H_0.
H_e can be determined as

$$H_e = 1 - \sum_{i=1}^{n} p_i^2 \qquad \text{eq. 2}$$

where p_i is the frequency of the i^{th} allele. If all alleles are equally frequent and N is the number of alleles, $1/N$ can be substituted for p in equation 2:

$$H_e = 1 - \left(\frac{1}{N}\right)^2 N \quad \text{eq. 3}$$

We substituted H_0 for H_e in equation 3 and solved for N to determine the number of equally frequent alleles to be included in the founding generation.

We used the mean F of individuals in generations 49-50 of the simulated populations (determined as explained above for microsatellite loci) to determine the SNP minor allele frequency in the first generation that was necessary to produce a mean expected heterozygosity (H_e) of 0.30 for loci sampled in the final simulated generation. We determined H_0 as above for simulations with microsatellite loci. We then solved the following equation for the minor allele frequency p:

$H_0 = 2p(1 - p)$ eq. 4

This approach is made possible by using a simulation framework where founders and their descendants are assigned genotypes after the demographic portion of the simulation is complete. The simulation model keeps track of the ancestral origin of each chromosome segment (i.e., the founder chromosome copy from which each chromosome segment in an individual originates). Founders are assigned genotypes after the last generation of the simulation. Then the descendants of founders are assigned genotypes based on the alleles found on the ancestral chromosome segments in the founders.

Supplement to Chapter 2

The randomization test for a difference between the mean estimated and true r^2 between survival and F proceeded as follows:

1. Calculate the raw difference between the mean estimated r^2 and the mean true r^2 between survival and F ($r^2(survival, F)$ and $\hat{r}^2(survival, F)$) for simulations within each category of $\sigma^2(F)$.

2. For each simulation within each category of $\sigma^2(F)$, randomly reassign the values of $r^2(survival, F)$ and $\hat{r}^2(survival, F)$ as being either $r^2(survival, F)$ or $\hat{r}^2(survival, F)$. Repeat this 10,000 times, each time recalculating the difference between the mean true r^2 and mean estimated r^2. This simulated the sampling distribution of the difference between means assuming there was no true difference.

3. The P-value for the randomization test for a difference between the mean $r^2(survival, F)$ and $\hat{r}^2(survival, F)$ for a given category of $\sigma^2(F)$ was the proportion of the randomized estimates of the difference between means that was greater than or equal to the original difference between means.

Agreement of simulation output with theoretical expectations

The slope from a regression of multiple locus heterozygosity (MLH) scaled to H_0 (MLH/H_0) versus F ($\hat{\beta}$ = -0.99) closely agreed with the theoretically expected slope (β = -1.0; Figure 2-S1). The mean F at migration-drift equilibrium among 30 replicate simulations agreed closely with theoretical predictions (Figure 2-S2). The relationship between the proportion of the genome that is identical by descent (F) and the pedigree inbreeding coefficient (F_P) agreed closely with theoretical expectations (Figure 2-S3).

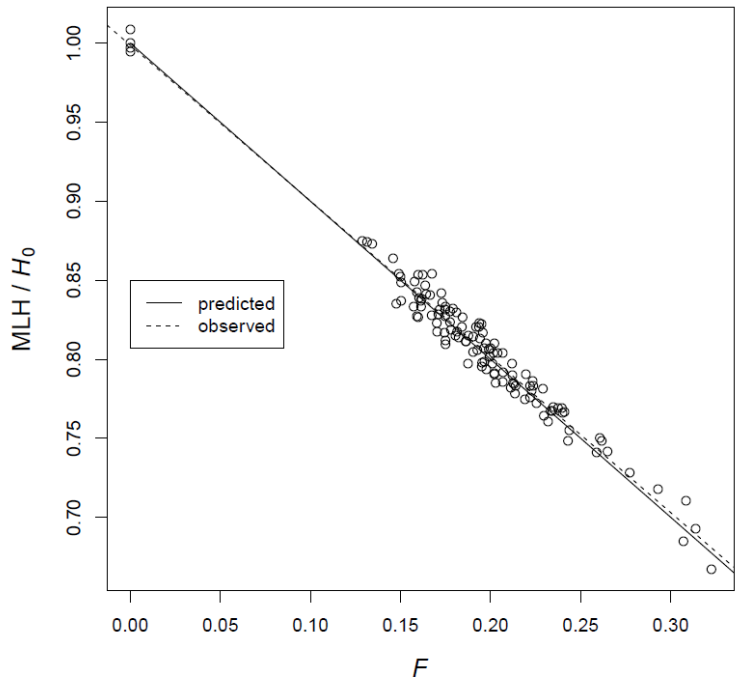

Figure 2-S1. Multiple-locus heterozygosity (scaled to the heterozygosity of non-inbred individuals (H_0)) measured at 5000 microsatellite loci plotted against F. The data shown are from a single simulated population with constant size of N = 60 individuals, and m = 0.004. The fitted line and slope ($\hat{\beta}$ = -0.99) are from a simple linear regression model.

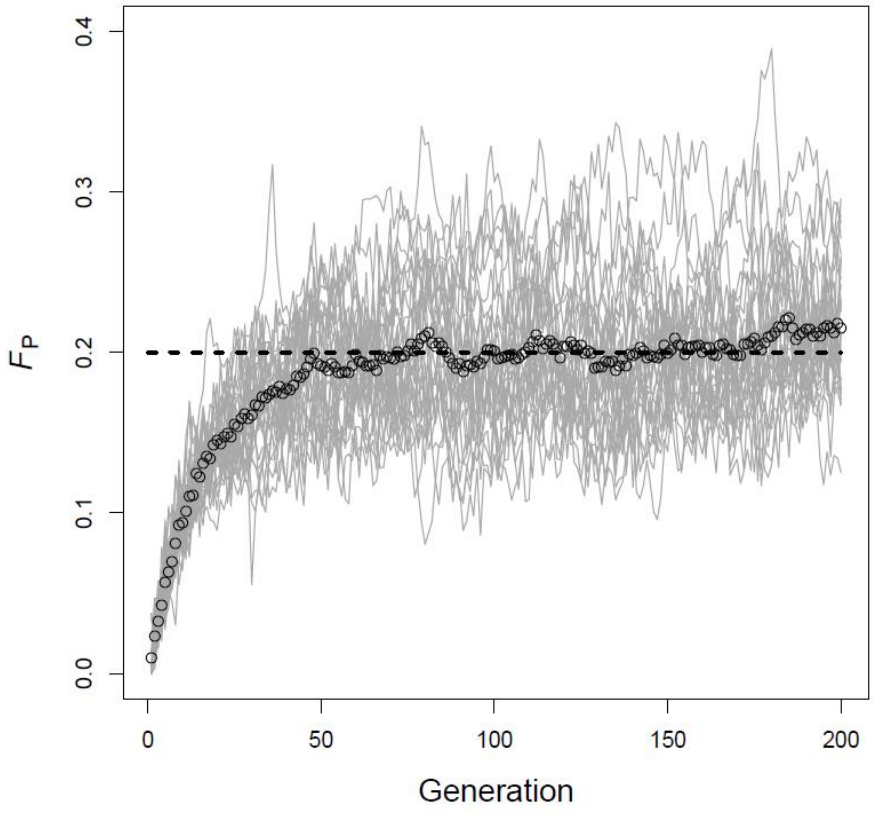

Figure 2-S2. The pedigree inbreeding coefficient (F_P) versus time in generations. Here FP was calculated using pedigrees that included the entire ancestry of all simulated individuals. The data presented here are from 30 simulated random mating populations with an effective populations size (N_e) of 40. Each simulated population received one immigrant per generation on average (m = 1/40) and was initiated with unrelated and non-inbred individuals. The dashed line represents the theoretically expected mean F_P at migration drift equilibrium which was calculated as $F_P = \frac{1}{4N_e m + 1}$. Each gray line represents the mean F_P for a single population across 200 generations. The black points represent the mean F_P across all 30 simulated populations each generation. We used F_P here instead of F because our simulation program calculates F only for individuals in the last two generations of the simulation. The results would be equivalent with F, because F_P is an unbiased estimator of F (see Figure 2-S3 below).

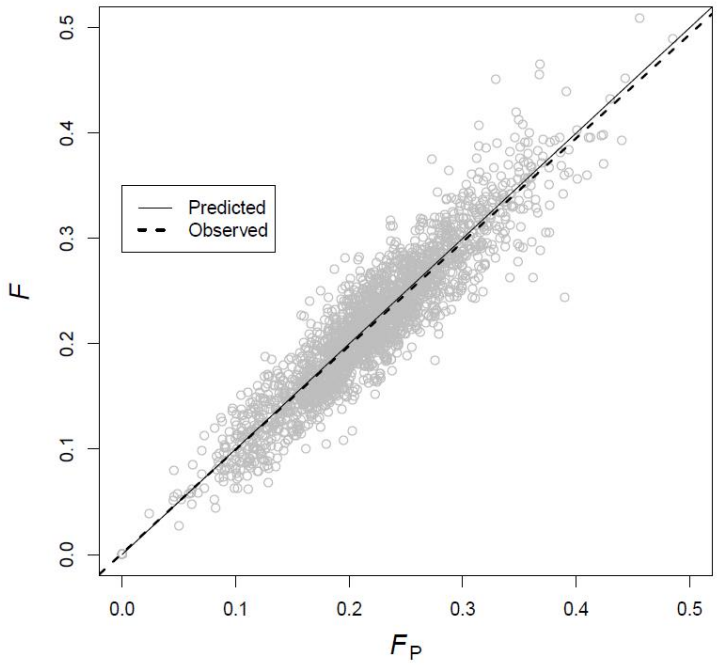

Figure 2-S3. The fraction of the genome that is identical by descent (F) versus the pedigree inbreeding coefficient F_P. Data are from individuals in the last generation of each of 30 simulated populations with $N_e = 40$. Each population received one immigrant per generation on average. The simulated populations were founded by unrelated and non-inbred individuals. F_P was calculated using pedigrees that included every simulated generation. When the founders of a pedigree are unrelated and non-inbred, F_P is expected to be an unbiased estimator of F. Thus, we expect a linear regression of F versus F_P to have an intercept of 0 and a slope of 1 (solid black line). The observed regression of F versus F_P is represented by the dashed line.

Statistical Results

Simulations with random mating and 6 diploid lethal equivalents

Comparison of the statistical power of tests for HFC and ID

The relative statistical power of HFC versus g_2 tests for ID depended strongly on the number of loci used to calculate H_s, $\sigma^2(F)$, and the strength of inbreeding depression (Figure 2-S4 below). The power to detect ID was significantly higher than the power to detect HFC in 23 out of 24 simulate combinations of $\sigma^2(F)$, and the type and number of markers. The greatest observed difference in statistical power was with 100 microsatellite loci and $\sigma^2(F) = 0.002$, where the power to detect ID was 0.69 greater than for HFC.

Approximately five times as many SNPs than microsatellites were necessary to achieve a given statistical power to detect ID or HFCs (Figure 2-S4). For example, the power to detect HFC or ID were very similar when using either 50 microsatellite loci or 250 SNPs. Statistical power was lower than 80% for HFCs regardless of the type of marker, $\sigma^2(F)$, or the number of loci used. 25 and 100 microsatellites were necessary for statistical power of the g_2 tests to exceed 80% when $\sigma^2(F)$ was 0.005 and 0.002, respectively. Statistical power for ID tests never exceeded 80% when $\sigma^2(F)$ was 0.001 regardless of the number and type of loci used. 100 and 250 SNPs were necessary for the power of ID tests to exceed 80% when $\sigma^2(F)$ was 0.005 and 0.002, respectively.

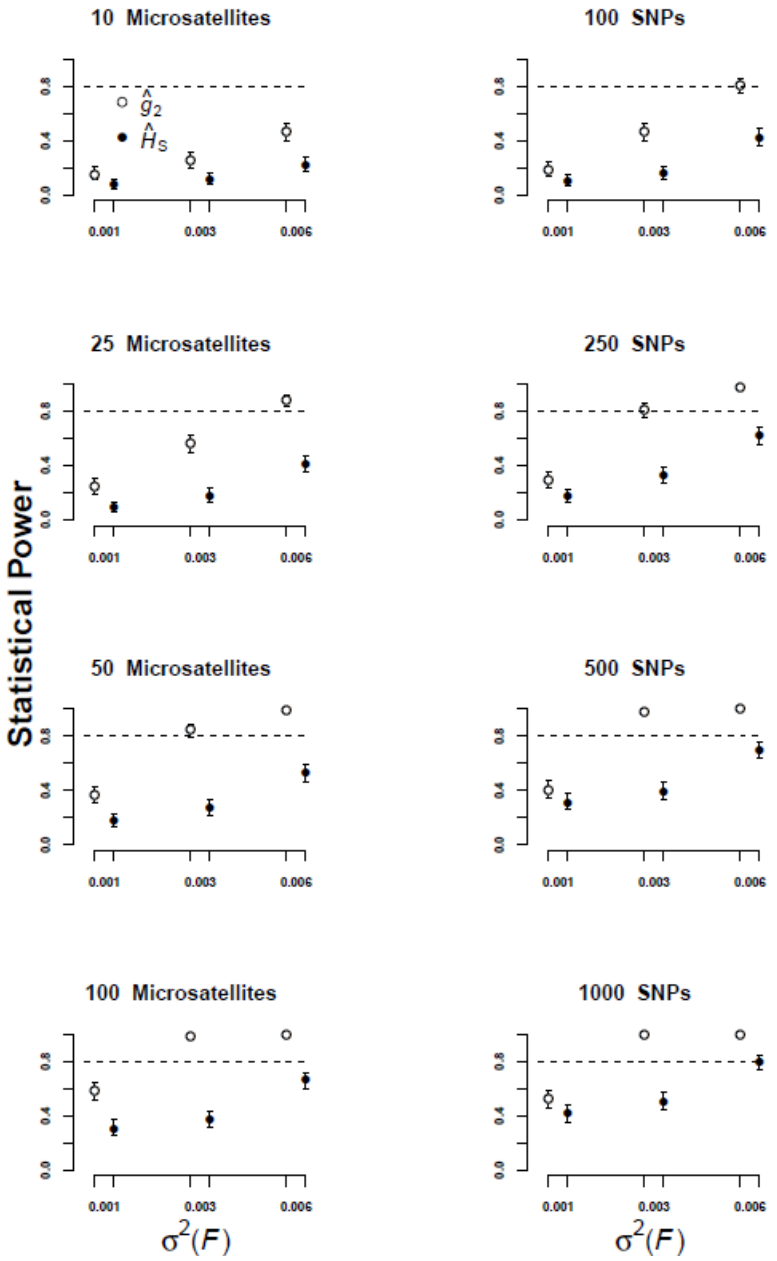

Figure 2-S4. The proportion of significant tests for g_2 (open circles) and for HFC (closed circles) versus $\sigma^2(F)$ for simulations with random mating, six diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars are 95% confidence intervals for the proportion of significant tests. Statistically significant differences between the statistical power of HFC and g_2 tests among simulations within each category of $\sigma^2(F)$ are labeled with asterisks.

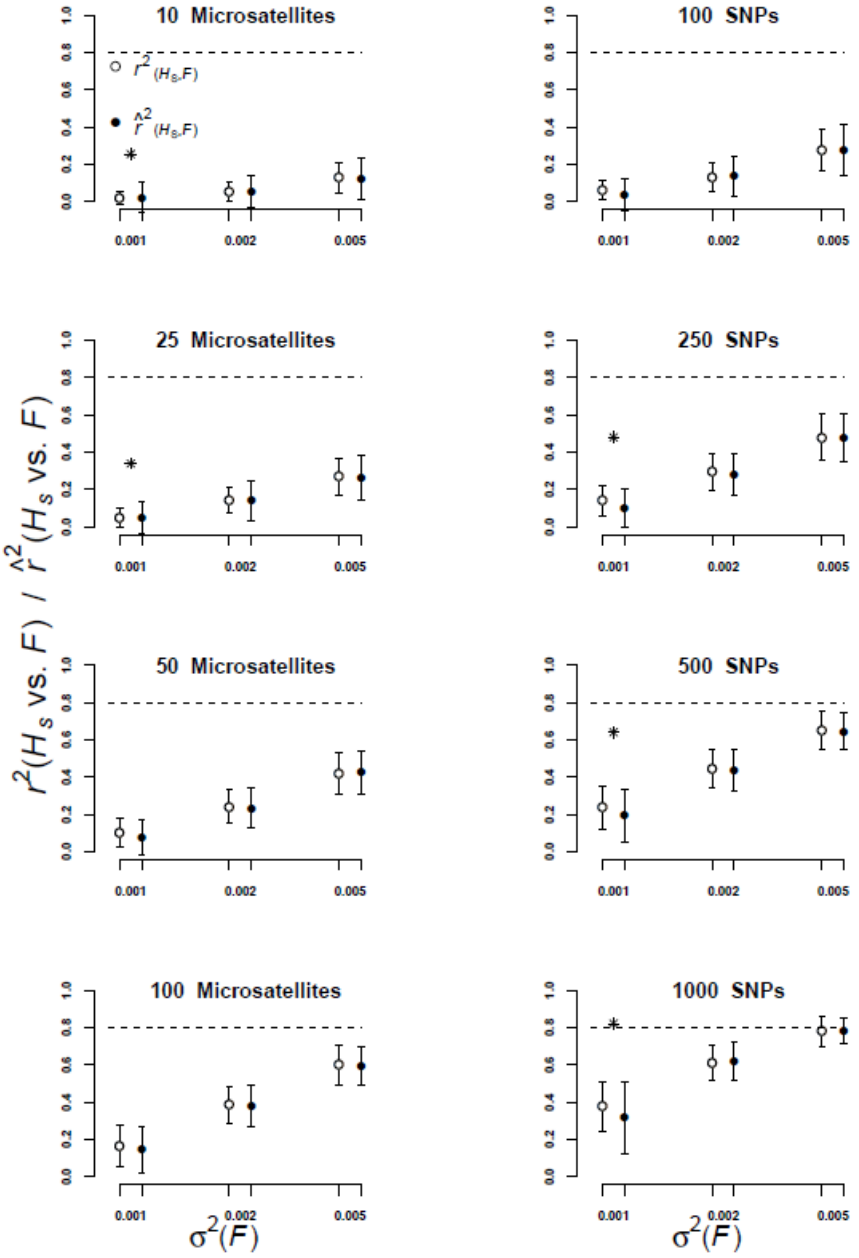

Figure 2-S5. The true (open circles) and estimated (closed circles) proportion of variance in H_S explained by F ($r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$) versus the variance of F ($\sigma^2(F)$). The data shown are from simulations of random mating populations with 6 diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars represent one standard deviation. Asterisks indicate statistically significant differences between the means of $r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$ among the simulations with each category of $\sigma^2(F)$.

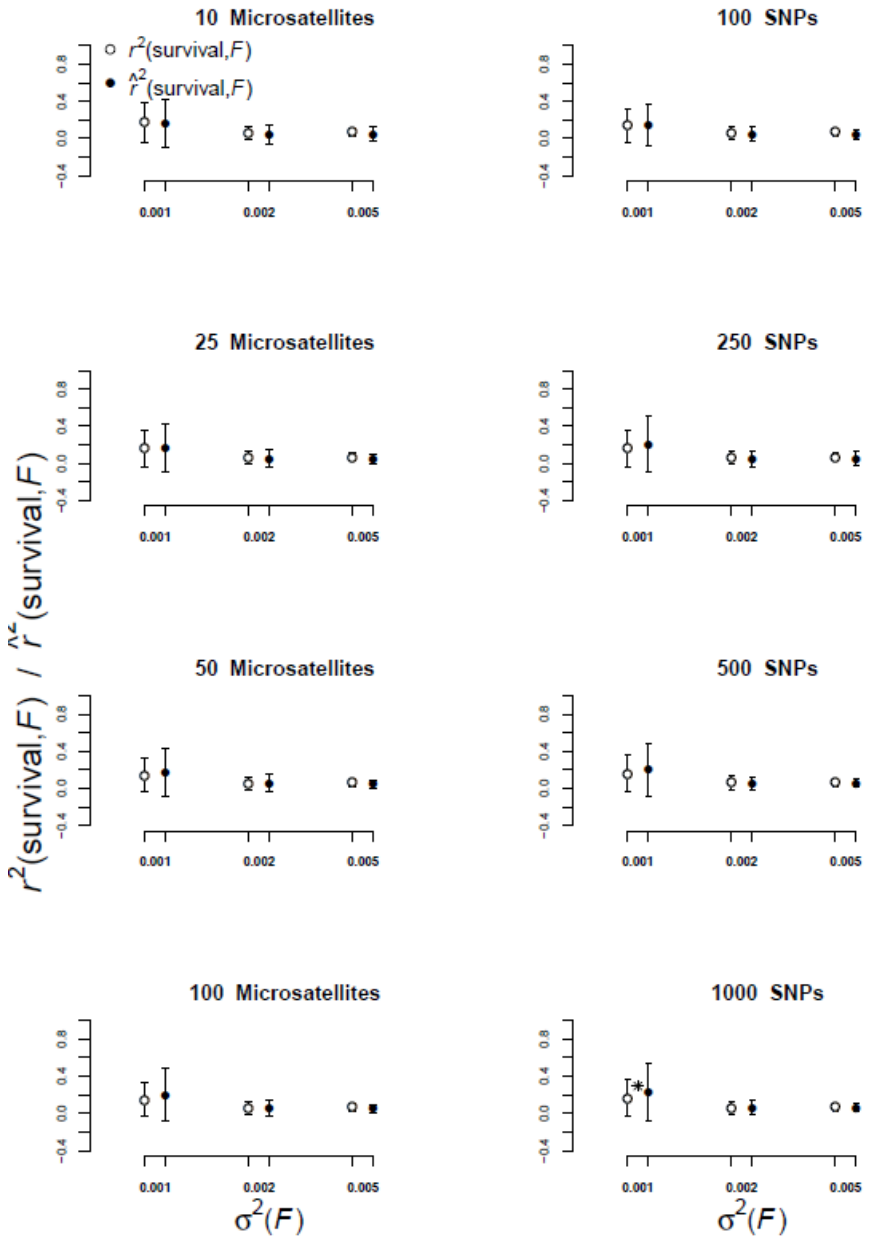

Figure 2-S6. The true (open circles) and estimated (closed circles) proportion of variation in survival explained by F ($r^2(survival, F)$ and $\hat{r}^2(survival, F)$) versus the variance of F (σ^2(F)). The data shown are from simulations of random mating populations with 6 diploid lethal equivalents, and using 10-100 microsatellites (left column) and 100-1000 SNPs (right column). Asterisks indicate statistically significant differences between the mean $r^2(survival, F)$ and $\hat{r}^2(survival, F)$ among simulations within each category of σ^2(F).

Simulations with random mating and 12 diploid lethal equivalents

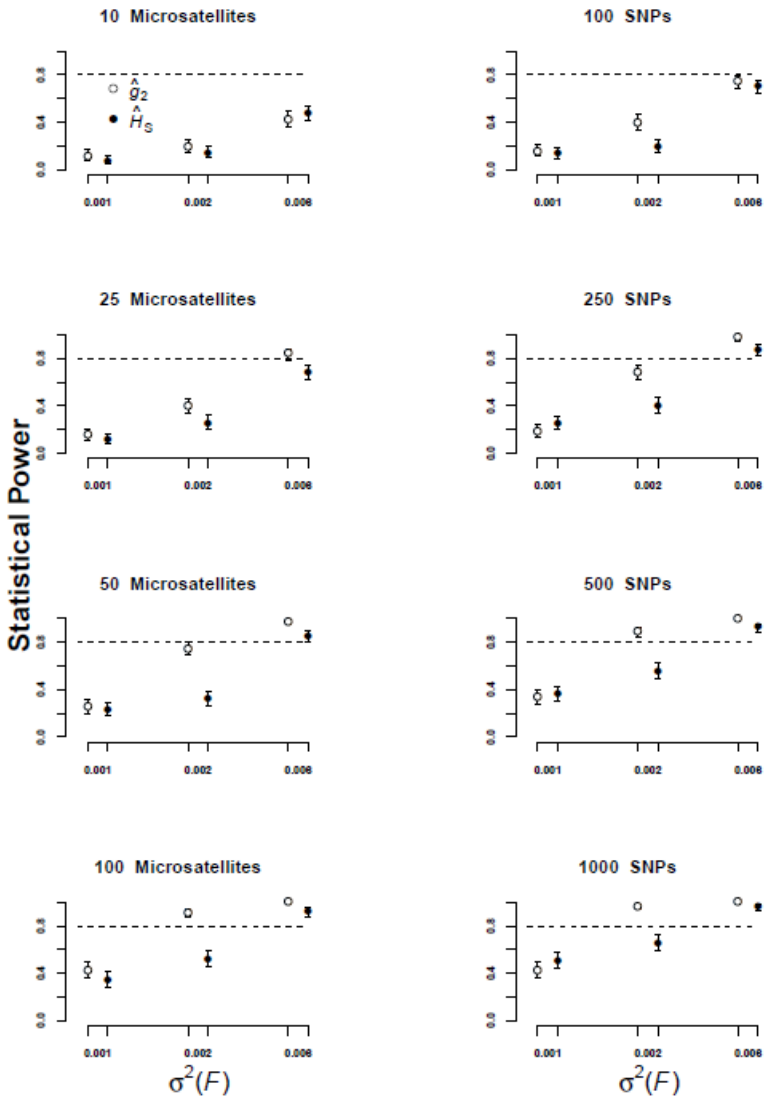

Figure 2-S7. The proportion of significant tests for g_2 (open circles) and for HFC (closed circles) versus $\sigma^2(F)$ for simulations with random mating, 12 diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars are 95% confidence intervals for the proportion of significant tests. Statistically significant differences between the statistical power of HFC and g_2 tests among simulations within each category of $\sigma^2(F)$ are labeled with asterisks.

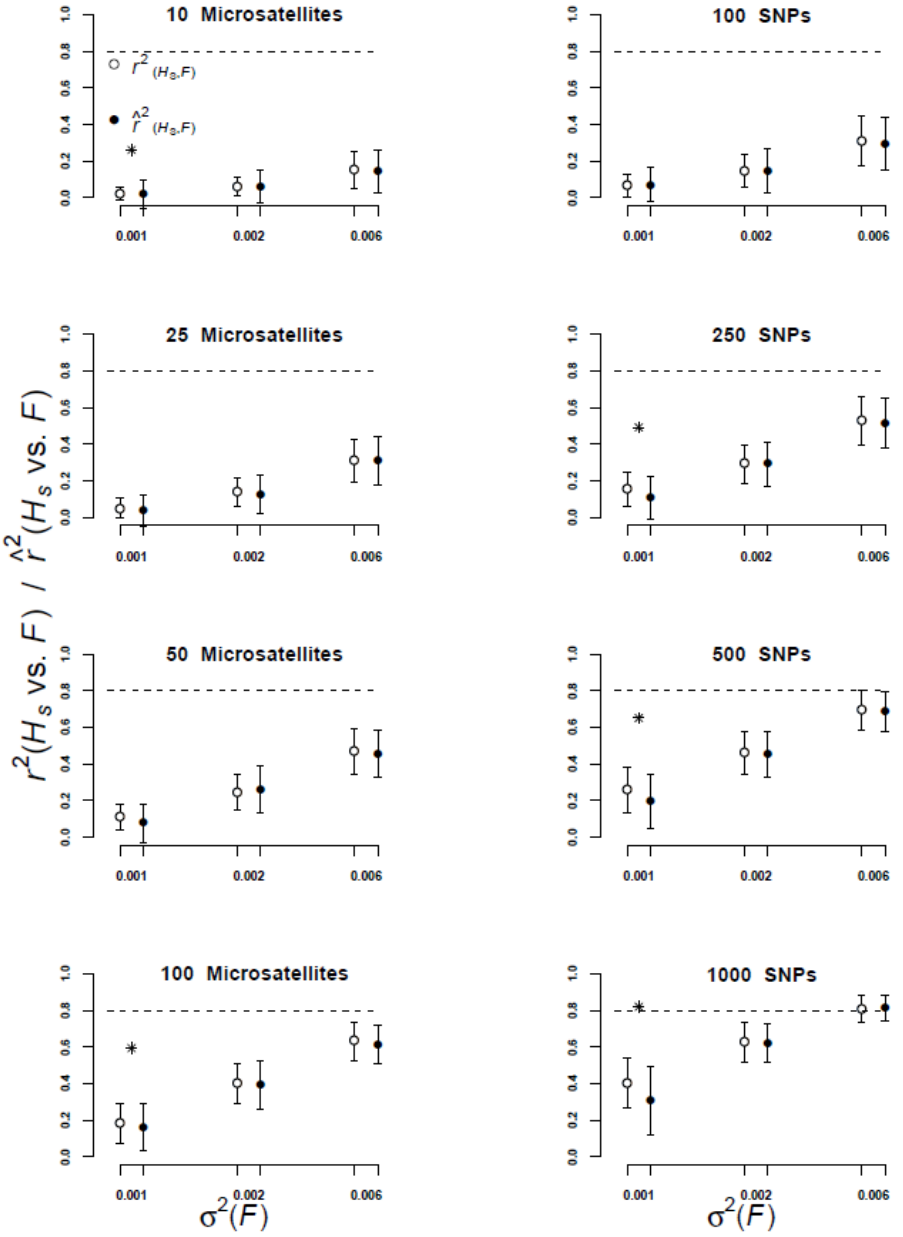

Figure 2-S8. The true (open circles) and estimated (closed circles) proportion of variance in H_S explained by F ($r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$) versus the variance of F ($\sigma^2(F)$). The data shown are from simulations of random mating populations with 12 diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars represent one standard deviation. Asterisks indicate statistically significant differences between the means of $r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$ among the simulations with each category of $\sigma^2(F)$.

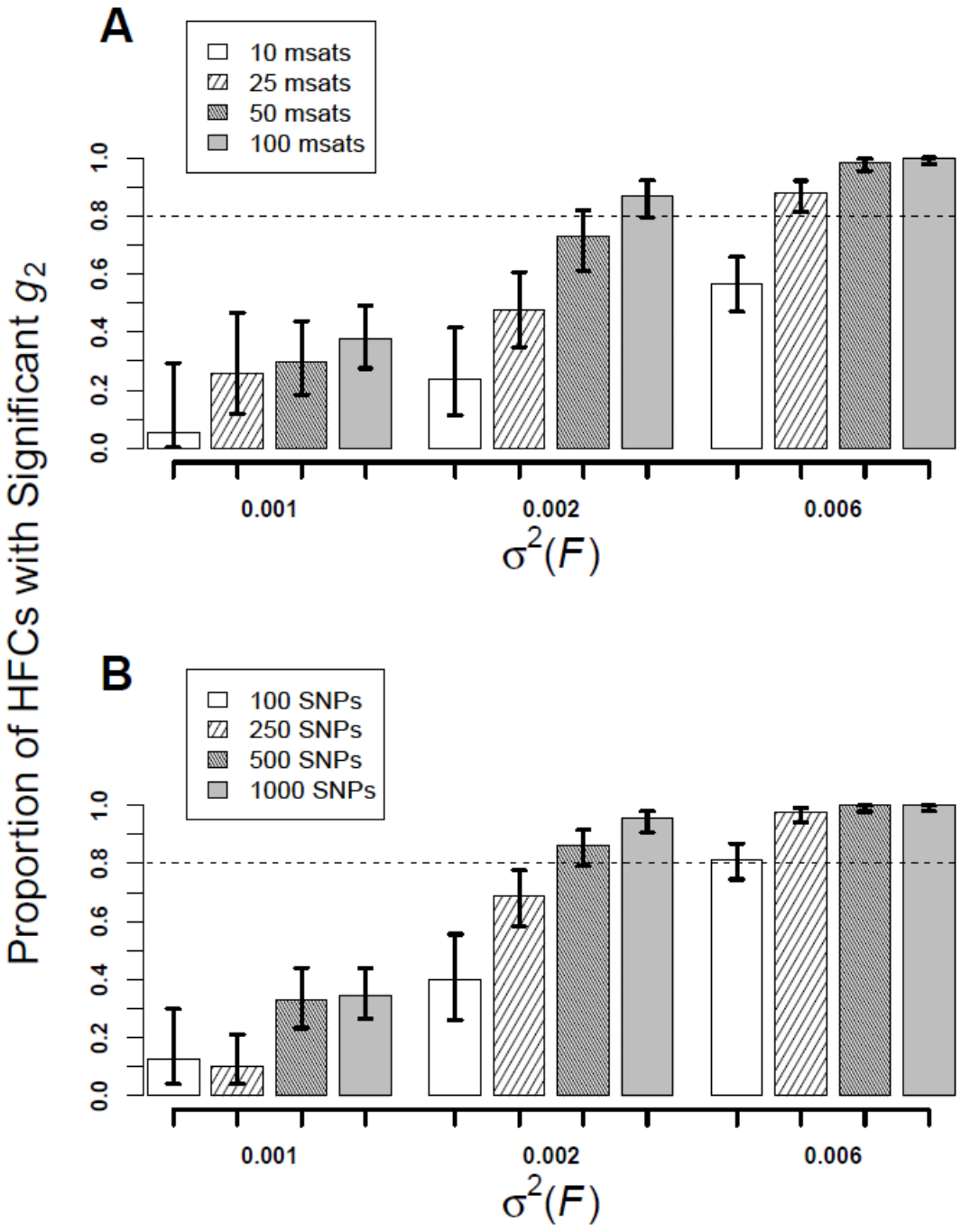

Figure 2-S9. Barplot of the proporion of simulations with significant HFCs that were also significant for g_2 with 95% confidence itnervals. The data are from simulations of random mating populations with 12 diploid lethal equivalents, microsatellite loci (upper panel) and SNPs (lower panel).

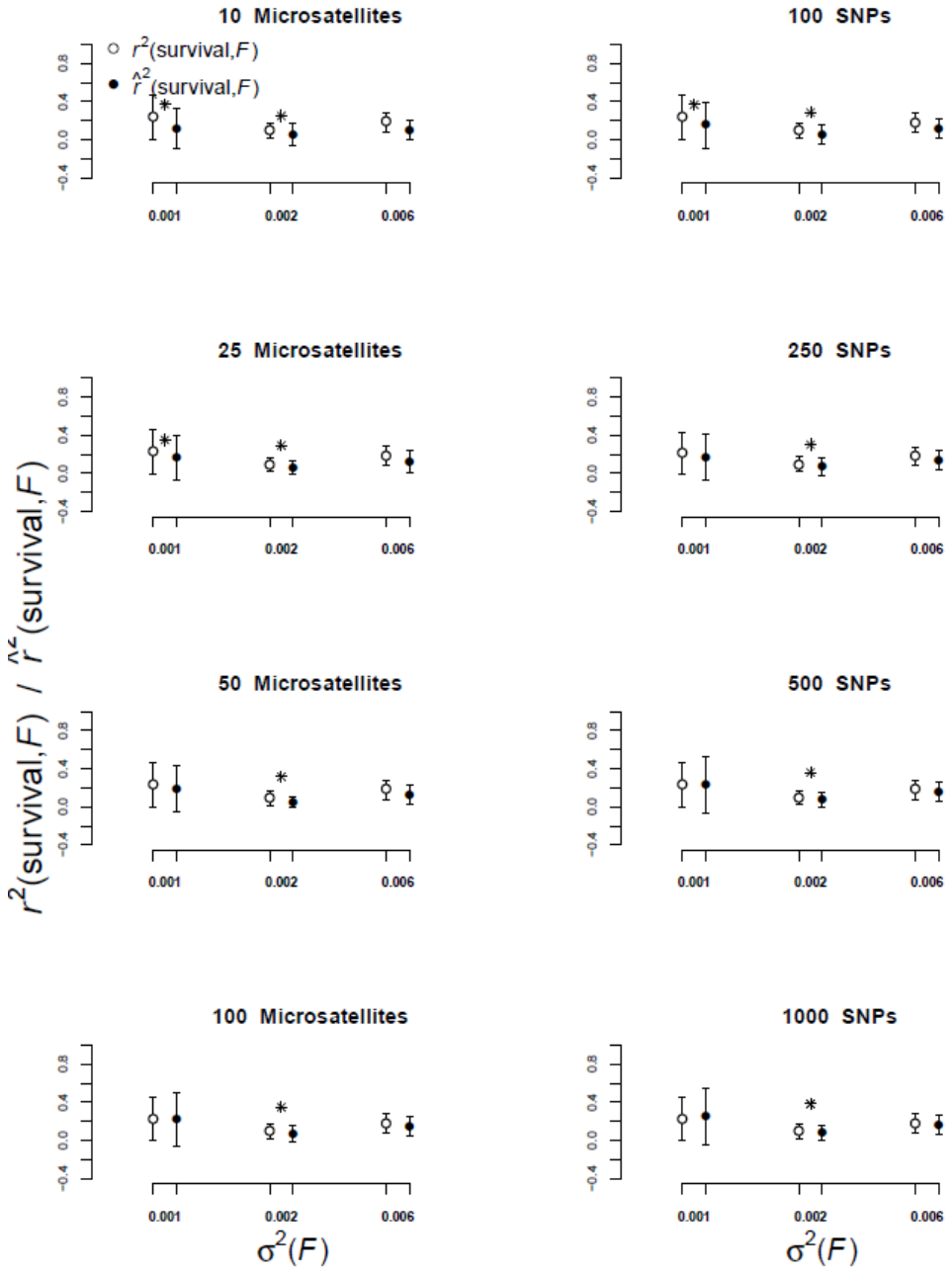

Figure 2-S10. The true (open circles) and estimated (closed circles) proportion of variation in survival explained by F ($r^2(survival, F)$ and $\hat{r}^2(survival, F)$) versus the variance of F (σ^2(F)). The data shown are from simulations of random mating populations with 12 diploid lethal equivalents, and using 10-100 microsatellites (left column) and 100-1000 SNPs (right column). Asterisks indicate statistically significant differences between the mean $r^2(survival, F)$ and $\hat{r}^2(survival, F)$ among simulations within each category of σ^2(F).

Simulations with partial selfing and 6 diploid lethal equivalents

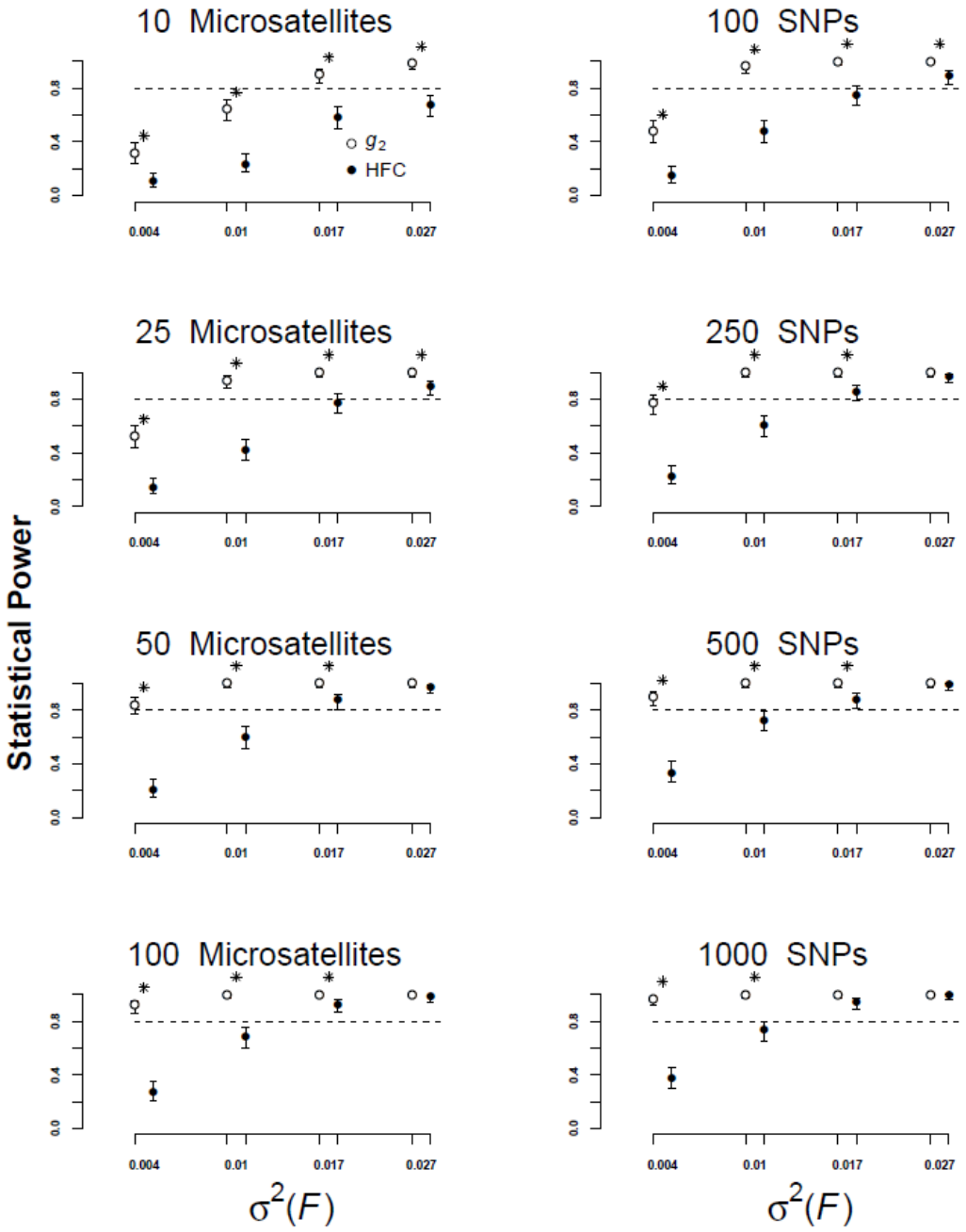

Figure 2-S11. The proportion of significant tests for g_2 (open circles) and for HFC (closed circles) versus $\sigma^2(F)$ for simulations with partial selfing and six diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars are 95% confidence intervals for the proportion of significant tests. Statistically significant differences between the statistical power of HFC and g_2 tests among simulations within each category of $\sigma^2(F)$ are labeled with asterisks.

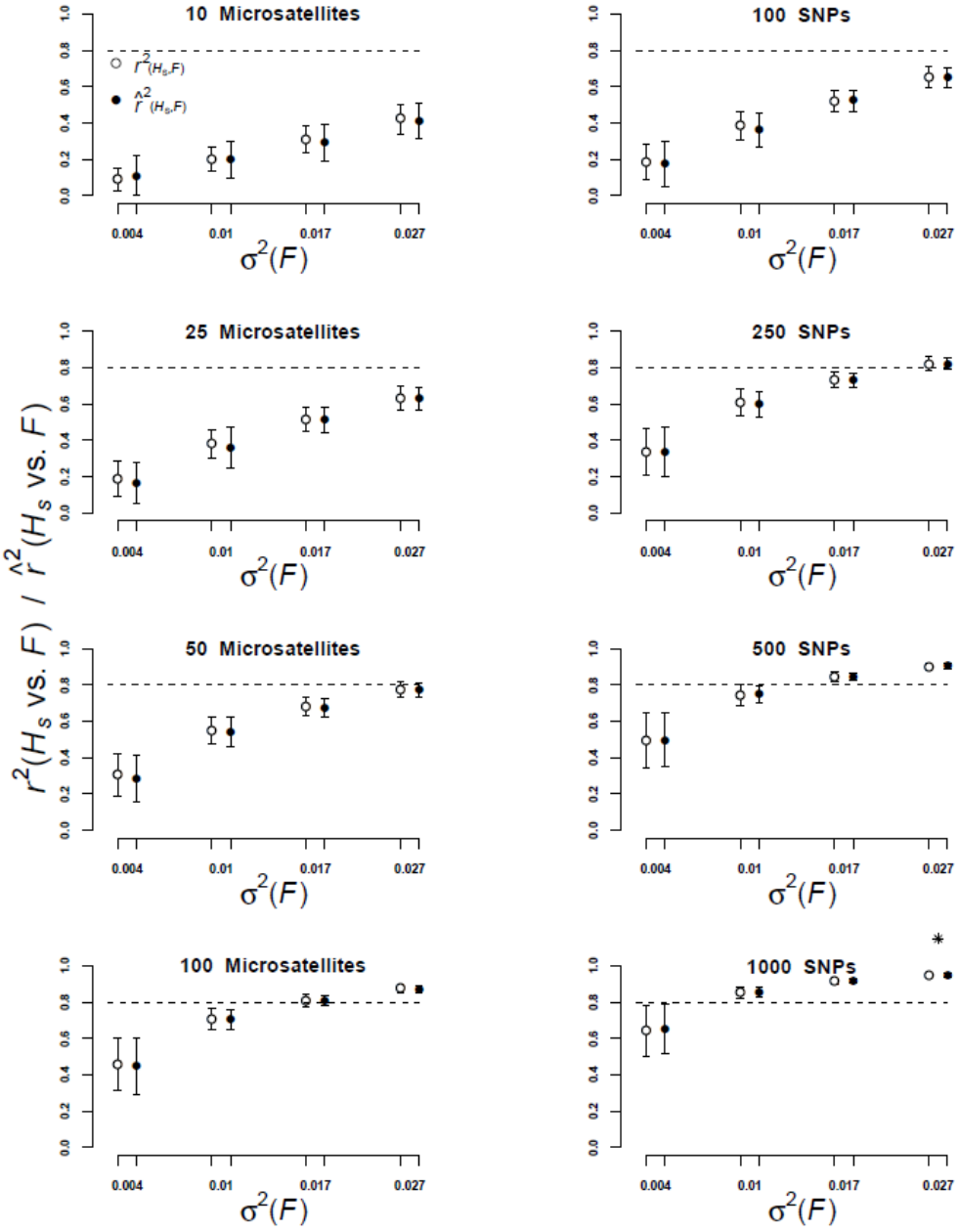

Figure 2-S12. The true (open circles) and estimated (closed circles) proportion of variance in H_S explained by F ($r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$) versus the variance of F ($\sigma^2(F)$). The data shown are from simulations of partially selfing populations with 6 diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars represent one standard deviation. Asterisks indicate statistically significant differences between the means of $r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$ among the simulations with each category of $\sigma^2(F)$.

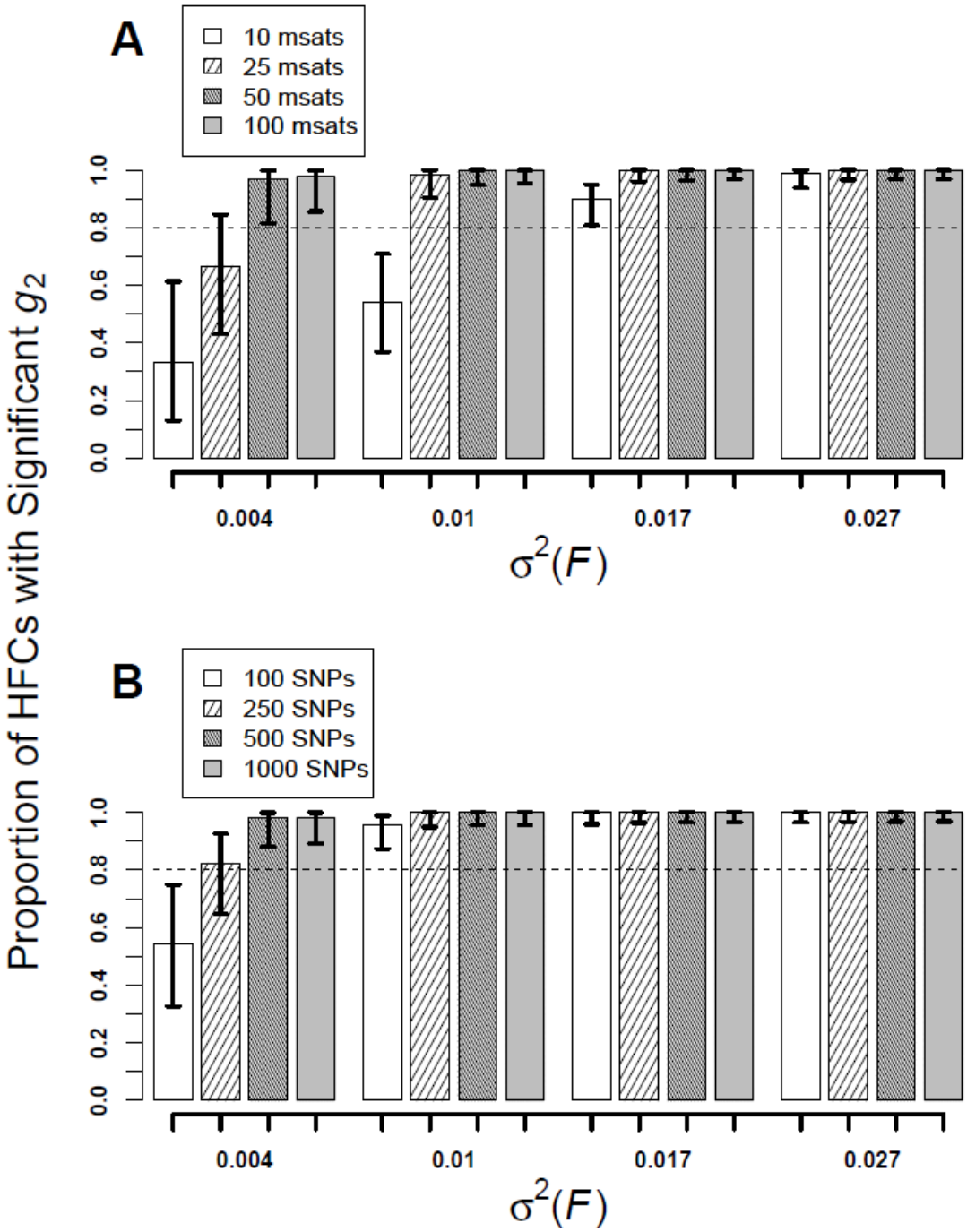

Figure 2-S13. Barplot of the proporion of simulations with significant HFCs that were also significant for g_2 with 95% confidence itnervals. The data are from simulations of partial selfing populations with 6 diploid lethal equivalents, microsatellite loci (A) and SNPs (B).

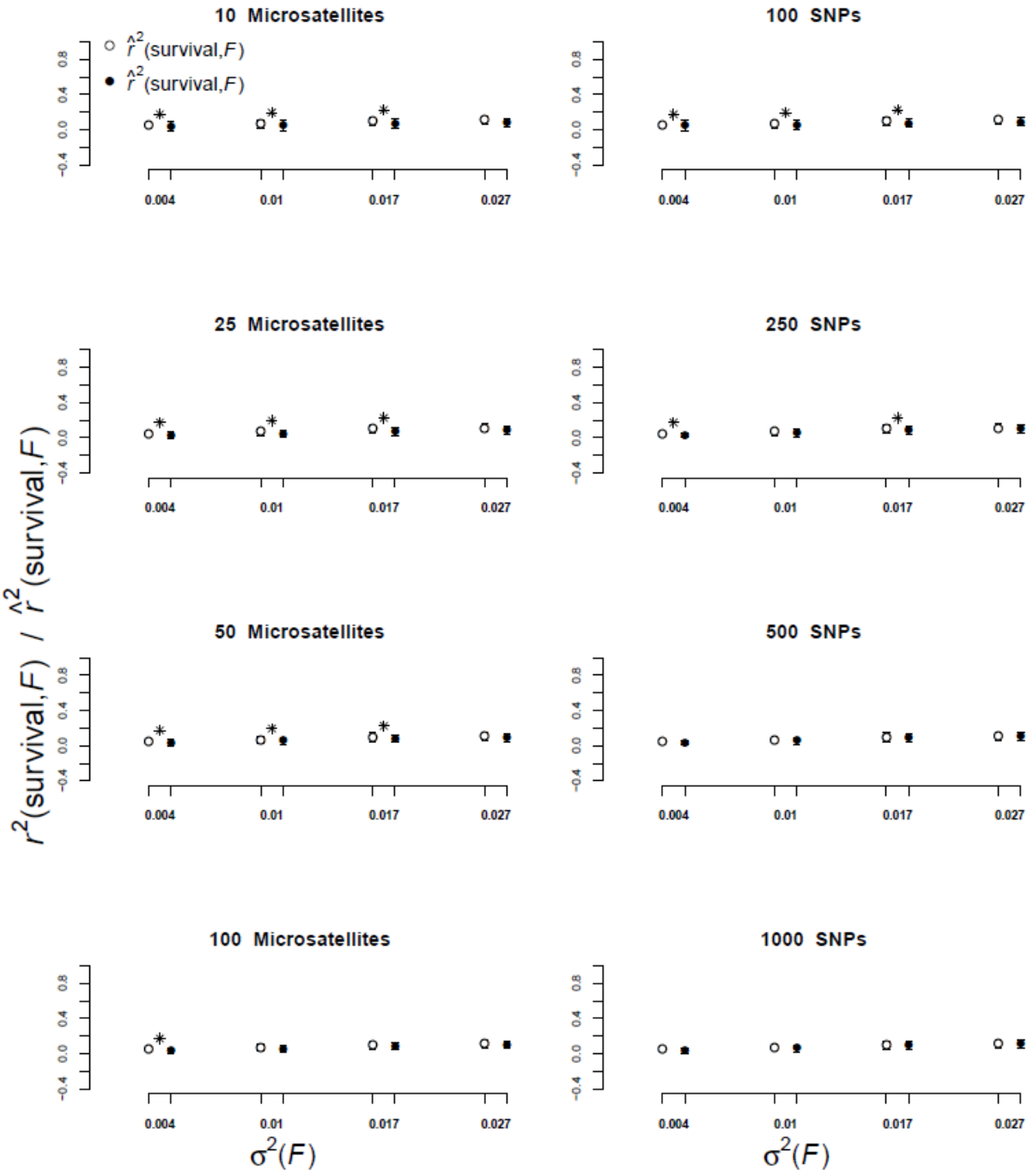

Figure 2-S14. The true (open circles) and estimated (closed circles) proportion of variation in survival explained by F ($r^2(survival, F)$ and $\hat{r}^2(survival, F)$) versus the variance of F (σ^2(F)). The data shown are from simulations of partial selfing populations with 6 diploid lethal equivalents, and using 10-100 microsatellites (left column) and 100-1000 SNPs (right column). Asterisks indicate statistically significant differences between the mean $r^2(survival, F)$ and $\hat{r}^2(survival, F)$ among simulations within each category of σ^2(F).

Simulations with partial selfing and 12 diploid lethal equivalents

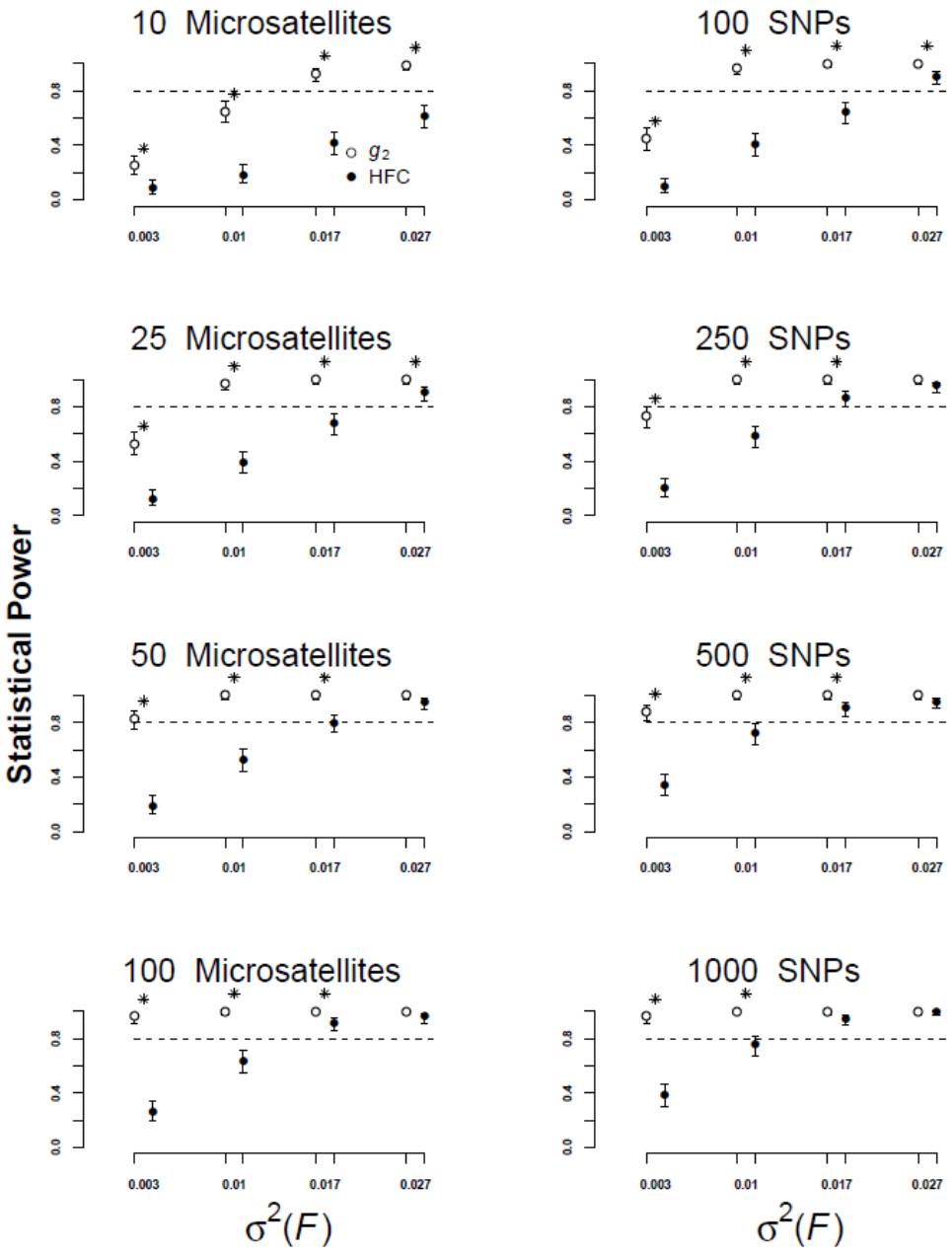

Figure 2-S15. The proportion of significant tests for g_2 (open circles) and for HFC (closed circles) versus $\sigma^2(F)$ for simulations with partial selfing and 12 diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars are 95% confidence intervals for the proportion of significant tests. Statistically significant differences between the statistical power of HFC and g_2 tests among simulations within each category of $\sigma^2(F)$ are labeled with asterisks.

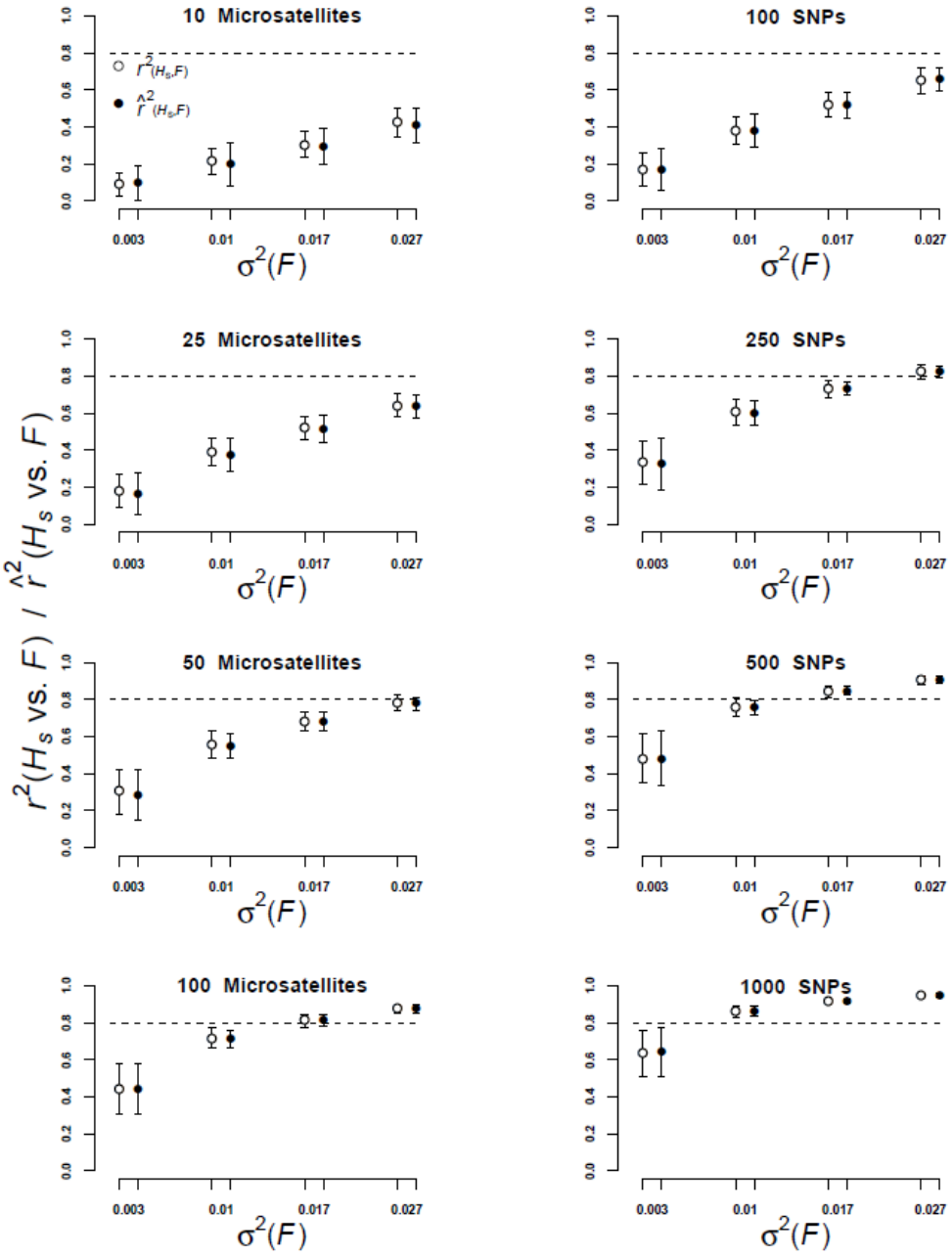

Figure 2-S16. The true (open circles) and estimated (closed circles) proportion of variance in H_S explained by F ($r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$) versus the variance of F ($\sigma^2(F)$). The data shown are from simulations of partial selfing populations with diploid lethal equivalents. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars represent one standard deviation. Asterisks indicate statistically significant differences between the means of $r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$ among the simulations with each category of $\sigma^2(F)$.

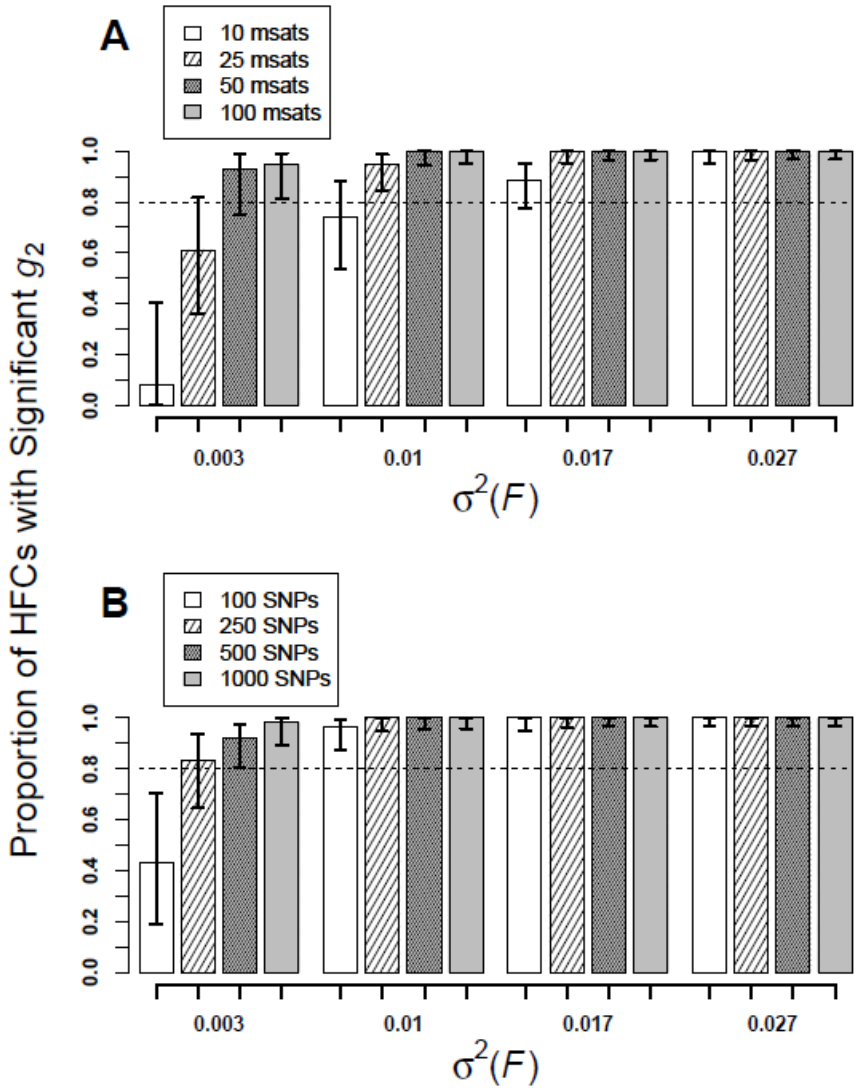

Figure 2-S17. Barplot of the proporion of simulations with significant HFCs that were also significant for g_2 with 95% confidence itnervals. The data are from simulations of partial selfing populations with 12 diploid lethal equivalents, microsatellite loci (A) and SNPs (B).

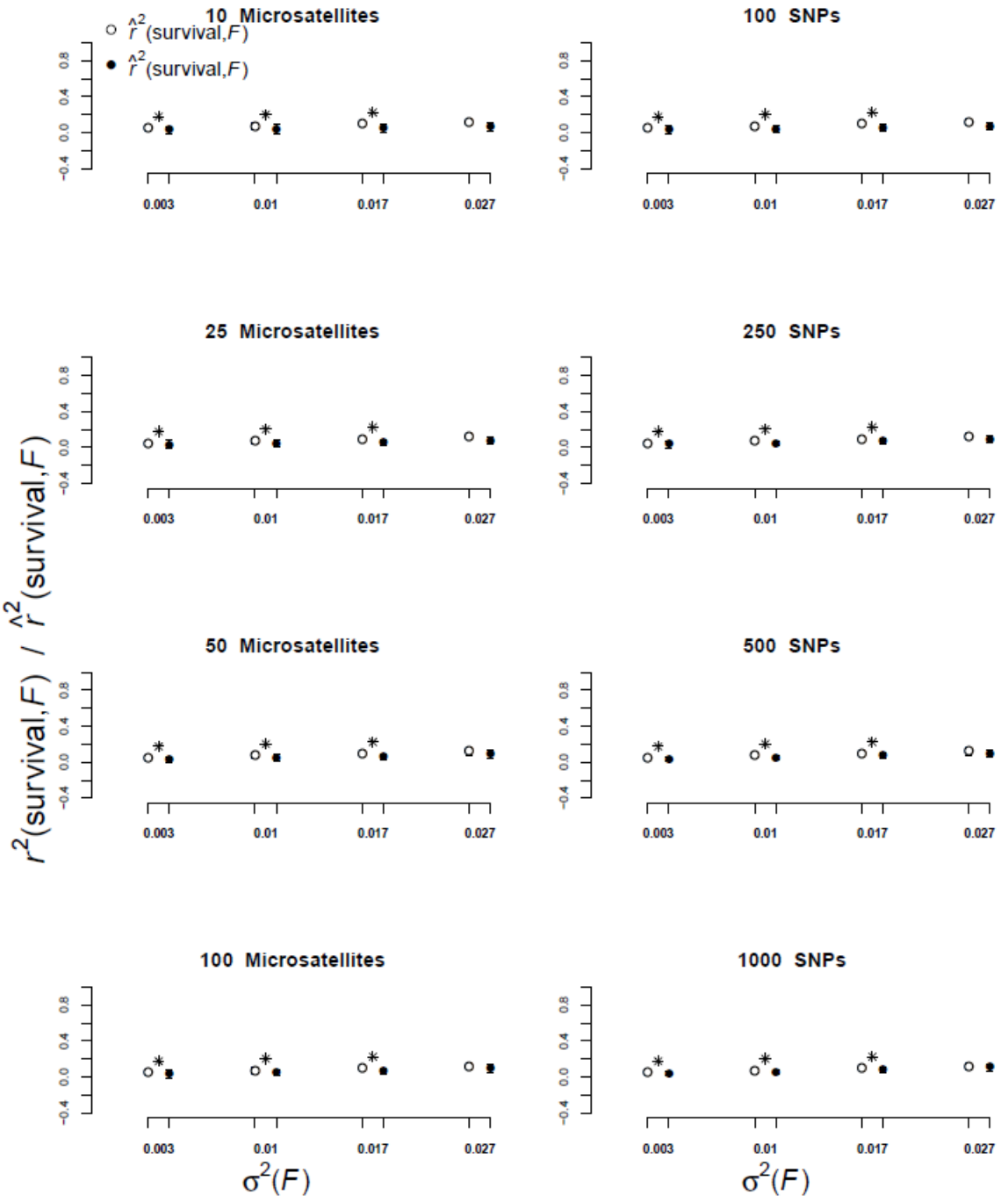

Figure 2-S18. The true (open circles) and estimated (closed circles) proportion of variation in survival explained by F ($r^2(survival, F)$ and $\hat{r}^2(survival, F)$) versus the variance of F (σ^2(F)). The data shown are from simulations of partial selfing populations with 12 diploid lethal equivalents, and using 10-100 microsatellites (left column) and 100-1000 SNPs (right column). Asterisks indicate statistically significant differences between the mean $r^2(survival, F)$ and $\hat{r}^2(survival, F)$ among simulations within each category of σ^2(F).

Simulations with random mating, 6 diploid lethal equivalents, and 1000 cM genomes with 10 chromosomes

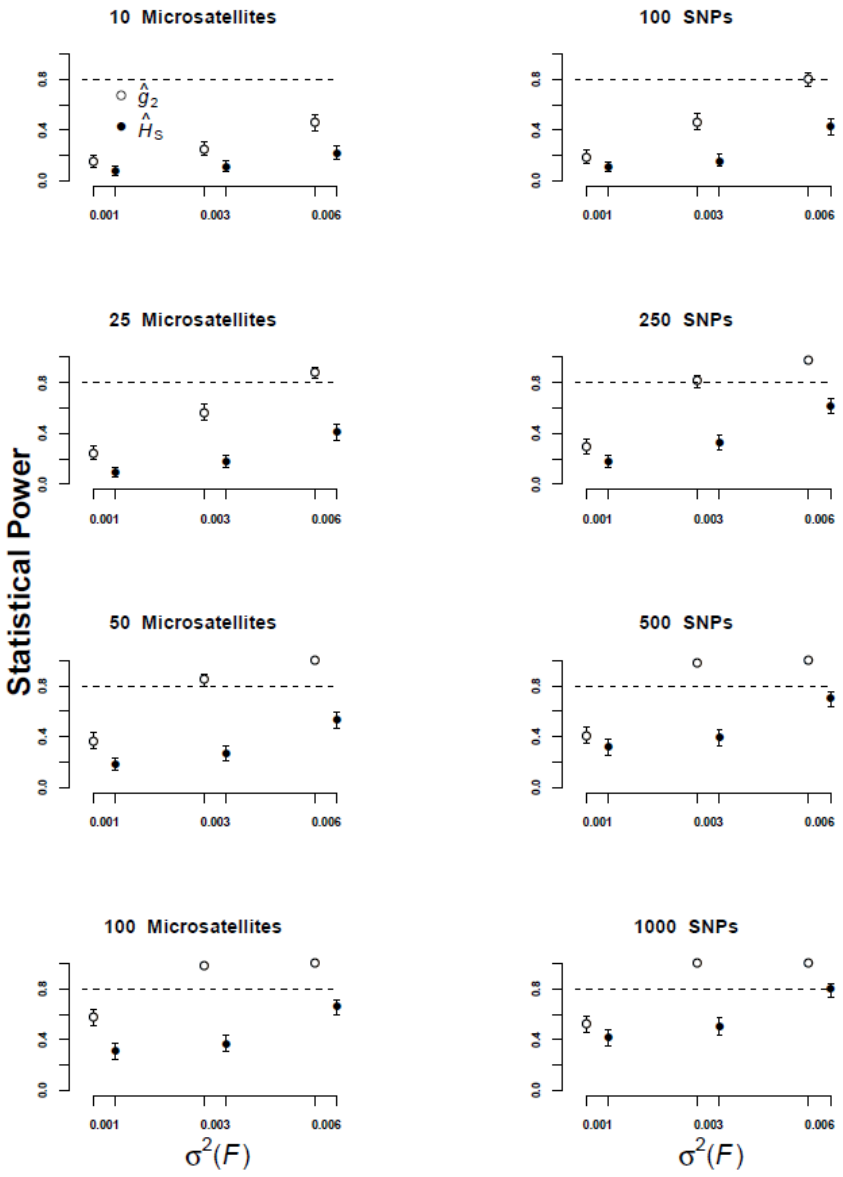

Figure 2-S19. The proportion of significant tests for g_2 (open circles) and for HFC (closed circles) versus $\sigma^2(F)$ for simulations with random mating, six diploid lethal equivalents and 1000cM genomes with 10 chromosomes. Results are shown from analyses with H_s estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars are 95% confidence intervals for the proportion of significant tests. Statistically significant differences between the statistical power of HFC and g_2 tests among simulations within each category of $\sigma^2(F)$ are labeled with asterisks.

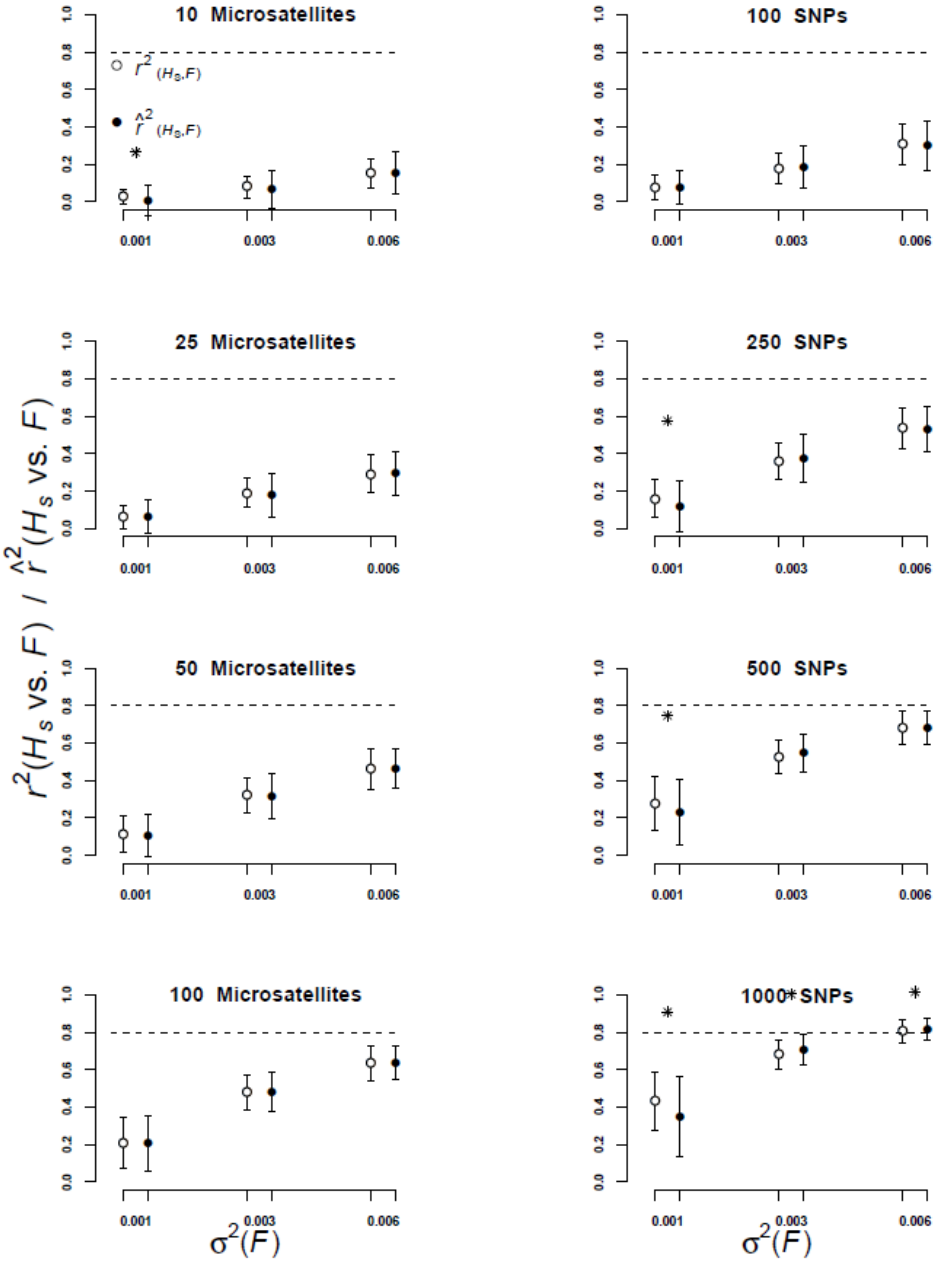

Figure 2-S20. The true (open circles) and estimated (closed circles) proportion of variance in H_S explained by F ($r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$) versus the variance of F ($\sigma^2(F)$). The data shown are from simulations of random mating populations with 6 diploid lethal equivalents, 1000cM genomes with 10 chromosomes. Results are shown from analyses with H_S estimated using 10, 25, 50, and 100 microsatellites (left column), and 100, 250, 500, and 1000 SNPs (right column). Error bars represent one standard deviation. Asterisks indicate statistically significant differences between the means of $r^2(H_S, F)$ and $\hat{r}^2(H_S, F)$ among the simulations with each category of $\sigma^2(F)$.

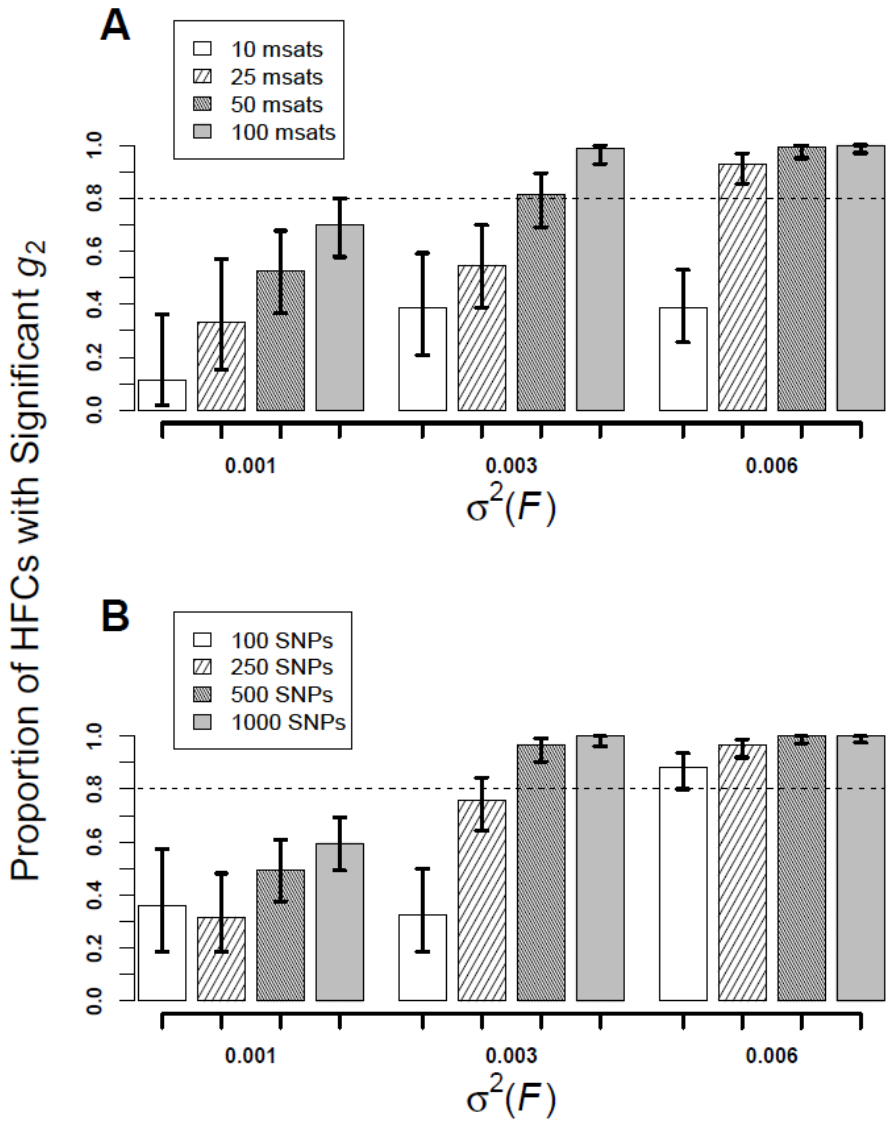

Figure 2-S21. The proportion of statistically significant HFCs that also had statistically significant tests for identity disequilibrium. Data are from random mating populations with 6 diploid lethal equivalents and 1000cM genomes with 10 chromosomes. Error bars are 95% confidence intervals.

Barplot of the proporion of simulations with significant HFCs that were also significant for g_2 with 95% confidence itnervals. The data are from simulations of random mating populations with 6 diploid lethal equivalents and 1000cM genomes with 10 chromosomes. HS was estimated with microsatellite loci (A) and SNPs (B).

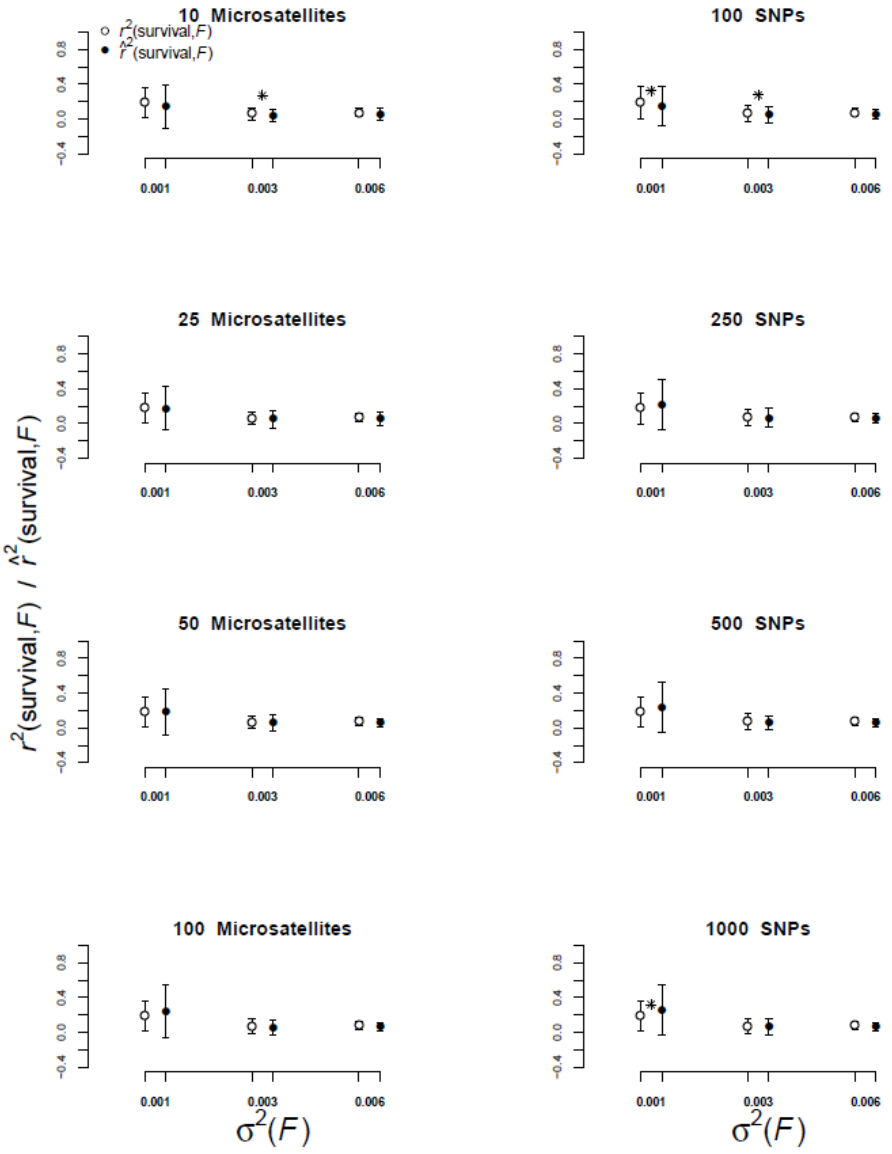

Figure 2-S22 The true (open circles) and estimated (closed circles) proportion of variation in survival explained by F ($r^2(survival, F)$ and $\hat{r}^2(survival, F)$) versus the variance of F (σ^2(F)). The data shown are from simulations of random mating populations with 6 diploid lethal equivalents, and 1000cM genomes with 10 chromosomes. H_S was estimated using 10-100 microsatellites (left column) and 100-1000 SNPs (right column). Asterisks indicate statistically significant differences between the mean $r^2(survival, F)$ and $\hat{r}^2(survival, F)$ among simulations within each category of σ^2(F).

Appendix 3-1

Here we use basic population genetics theory to demonstrate why F_H was a biased measure of PGIBD in our study. As mentioned in the main text, F_H measures the reduction in multiple-locus heterozygosity relative to the Hardy-Weinberg expected heterozygosity (H_e). The formulation of F_H used in this paper, and implemented in the program PLINK (Purcell *et al.* 2007) is

$$F_{Hi} = \frac{(O_i - E_i)}{(L_i - E_i)} \qquad \textit{Equation 1}$$

where O_i is the observed number of homozygous SNPs for individual i, E_i is the Hardy-Weinberg expected number of homozygous SNPs for individual i, and L_i is the number of typed SNPs for individuals i. Equation 1 is equivalent to

$$F_{Hi} = \frac{(1 - MLH_i) - (1 - H_e)}{H_e} = \frac{H_e - MLH_i}{H_e} \qquad \textit{Equation 2}$$

where MLH_i is the multiple-locus heterozygosity of individual i and H_e is the Hardy-Weinberg expected heterozygosity of the typed SNPs. Equation two can be rearranged to yield

$$MLH_i = H_e(1 - F_{Hi}), \qquad \textit{Equation 3}$$

which is equivalent to the classical description of the relationship between individual heterozygosity and individual inbreeding:

$$H_i = H_0(1 - F_i) \qquad \textit{Equation 4}$$

where H_i is the genome wide heterozygosity of individual i, F_i is the PGIBD of individual i, and H_0 is the genome-wide heterozygosity of a non-inbred individual (Crow & Kimura 1970, p. 66). It is obvious from a comparison of Equations 3 & 4 that H_e must be equal to H_0 for F_H to be an unbiased estimator of PGIBD.

H_e will not equal H_0 when allele frequencies are measured in a population with $N_e << \infty$, because mating between relatives occurs in any finite population. From classical population genetics (Wright 1931), we expect a proportional increase in the population mean F, and a proportional decrease in H_e of $1/2N_e$ each generation as a result of genetic drift. Thus, allele frequencies must be measured from an historical population where the assumption of He = H0 is reasonable for F_H to be an unbiased measure of PGIBD. Alternatively, allele frequencies could be measured across many different extant populations where the collective N_e is very large and the assumption that H_e would approximately equal H_0.

Supplement to Chapter 3

Table 3-S1. PLINK ROH detection settings for analyses based on different numbers of SNPs and minimum ROH lengths. The PLINK settings listed for > 100K SNPs were used in our preliminary simulations. The preliminary simulations showed that using more than 100K SNPs did not increase the correlation of PGIBD with F_H or with F_{ROH}.

Number of SNPs	Minimum ROH length	SNP Density (kb/SNP)	Maximum Gap between Adjacent SNPs (Kb)	Minimum number of SNPs/ROH
20-100K	2 Mb	100	250	30
200-300K	2 Mb	75	175	70
≥ 400K	2 Mb	50	100	100
20-100K	8 Mb	200	500	60
200-300K	8 Mb	150	350	140
≥ 400K	8 Mb	100	200	200

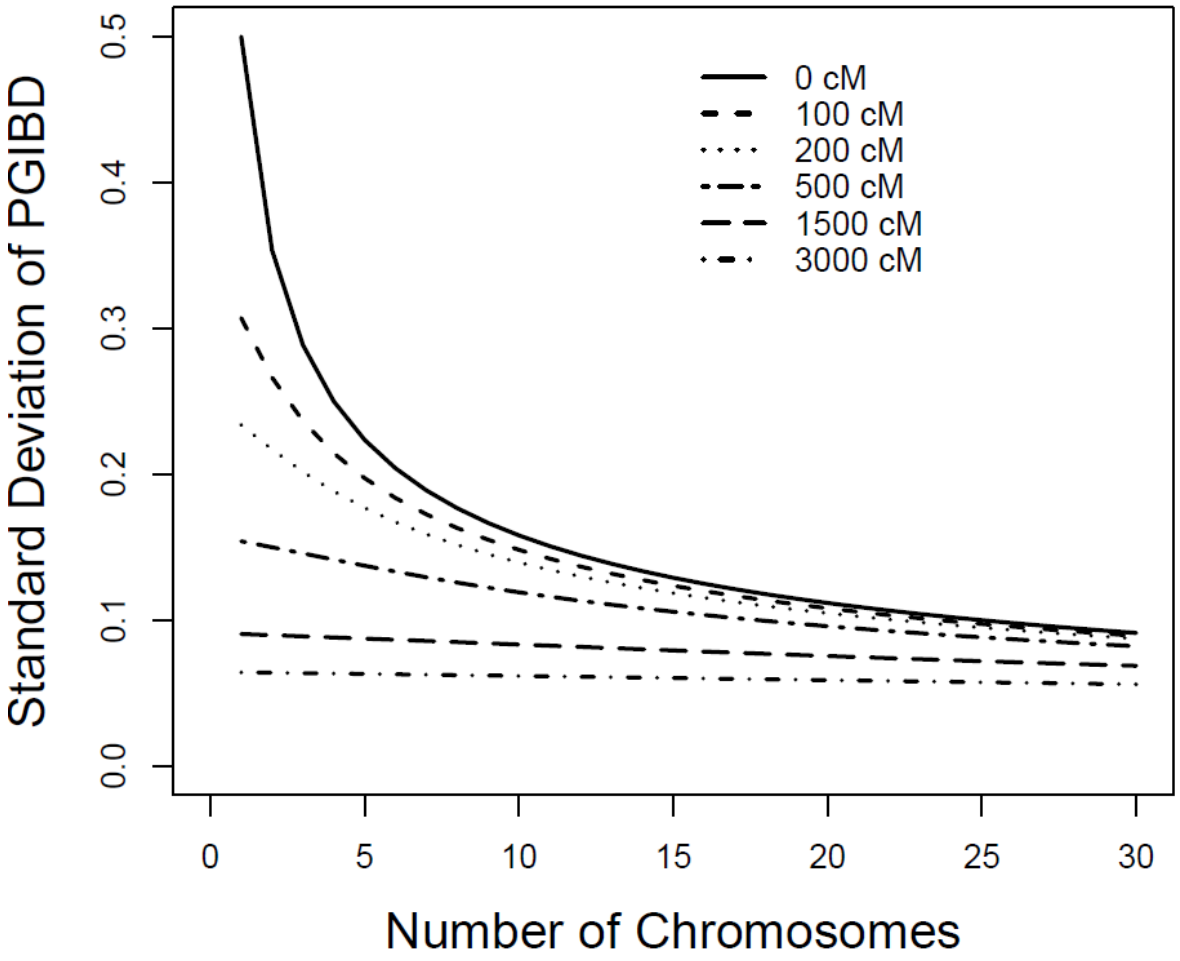

Figure 3-S1. The predicted standard deviation of PGIBD after one generation of selfing versus the number of chromosomes (Franklin 1977). Predictions from genomes with different genetic map lengths are represented by different line types.

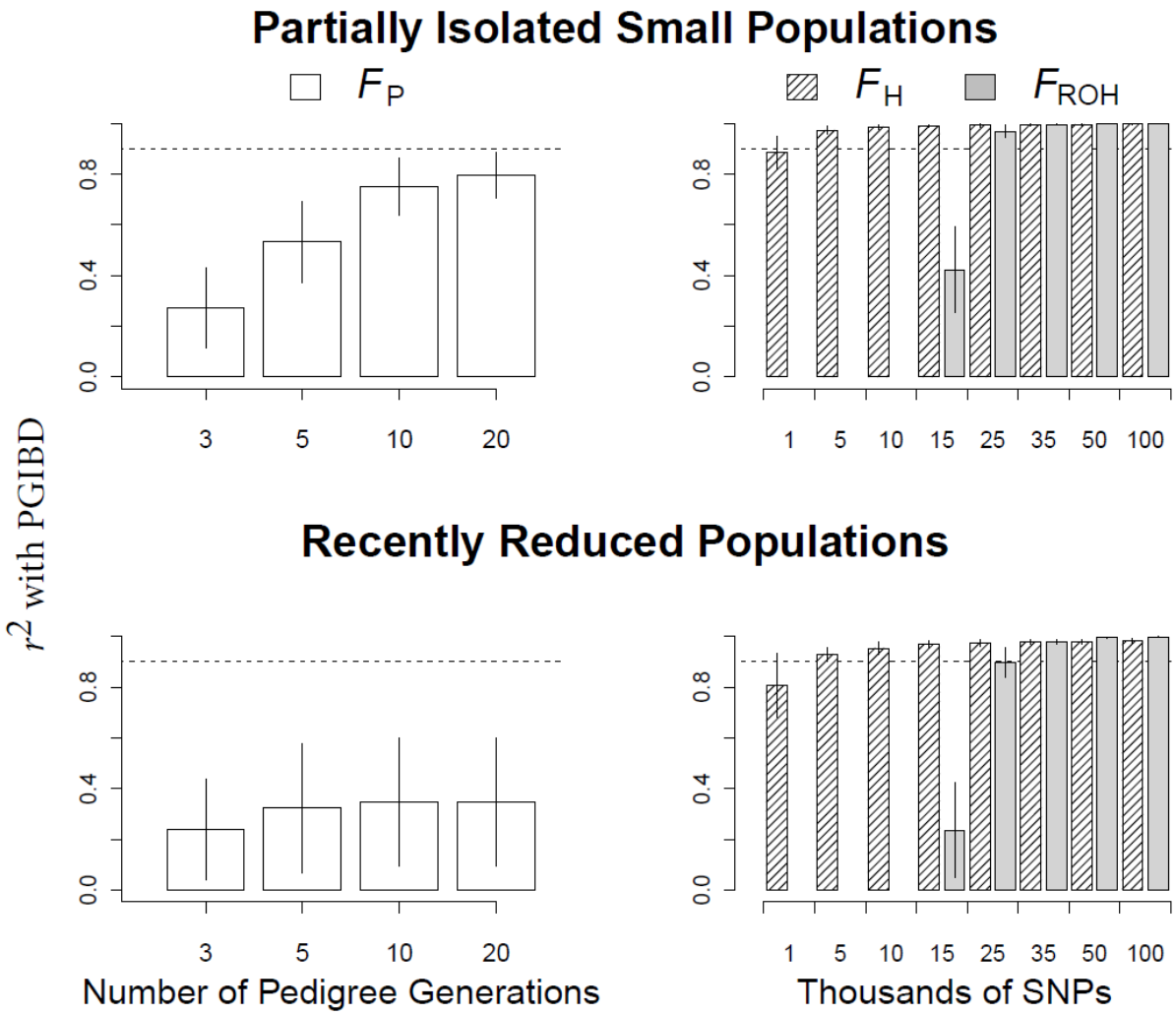

Figure 3-S2. Barplots of the mean r^2 (+/- 1 s.d. across 20 simulated populations) from regressions of F_P, F_H and F_{ROH} versus PGIBD. Results shown here are from simulations that used a genetic map length of 800 cM. Results are shown from 20 partially isolated (m = 0.05) small populations (local N_e = 20) in the top row. The data shown in the bottom row are from 20 populations with recently reduced N_e (from N_e = 500 to N_e = 20). Horizontal dotted lines are placed at r^2 = 0.9 to aid comparison of r^2across F_P, F_H and F_{ROH}.

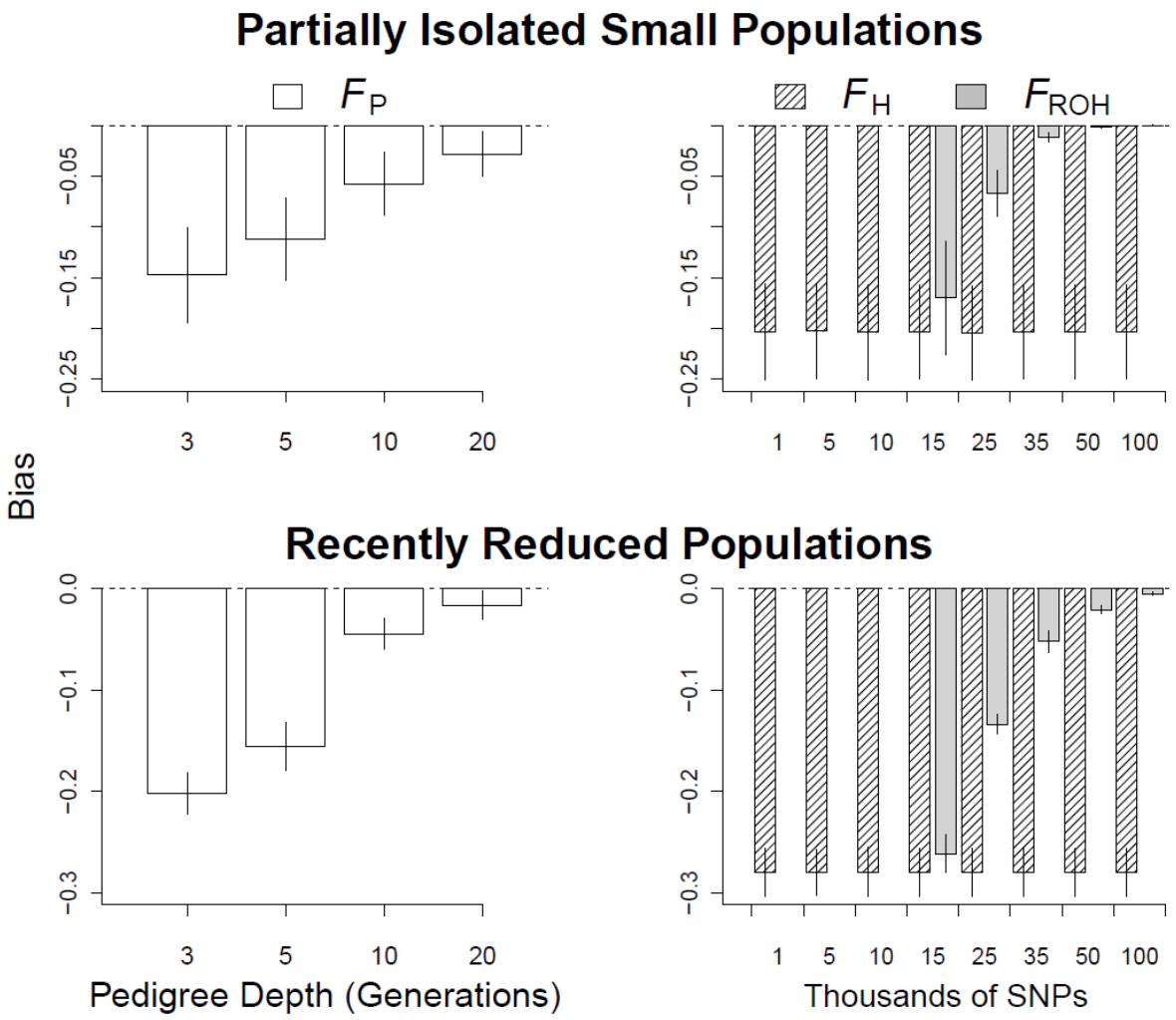

Figure 3-S3. The bias of F_P, F_H, and F_{ROH} among simulations that used a genetic map length of 800 cM. Results from 20 simulations of partially isolated small populations are shown in the top row. Results from 20 simulations of populations with a recent reduction in N_e are shown in the bottom row.

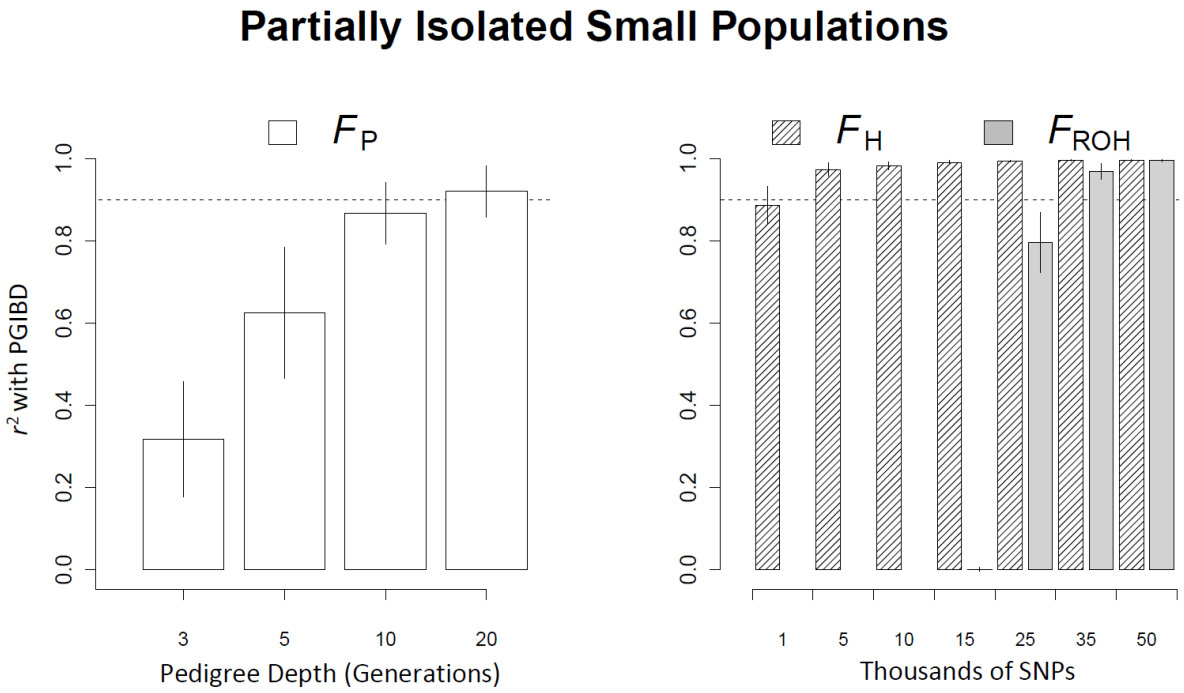

Figure 3-S4. Barplots of mean r^2 from regressions of PGIBD versus F_P, F_H, and F_{ROH} across 20 simulations of partially isolated small populations (N_e = 20). The results presented here are from the same simulated partially isolated populations shown in Figure 3-2 in the main text, except here the SNP genotype data were LD-pruned using the default settings in PLINK (window size = 50 SNPs, step size = 5 SNPs, minimum r^2 threshold for removal of a SNP = 0.5). We only used up to 50K SNPs to estimate F_H and F_{ROH} here because some loci were removed from the 100K simulated SNPs during LD pruning.

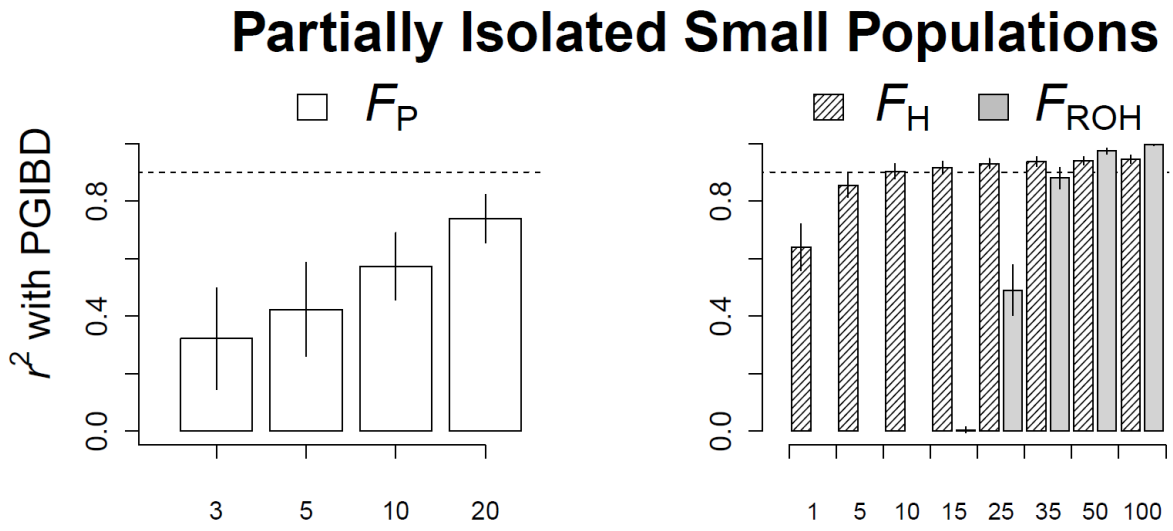

Figure 3-S5. Barplots of the mean r^2 (+/- 1 s.d. across 20 simulated populations) from regressions of F_P, F_H and F_{ROH} versus PGIBD. Results shown here are from simulations that used a genetic map length of 3600 cM, N_e = 100, and m = 0.01.

Supplement to Chapter 4

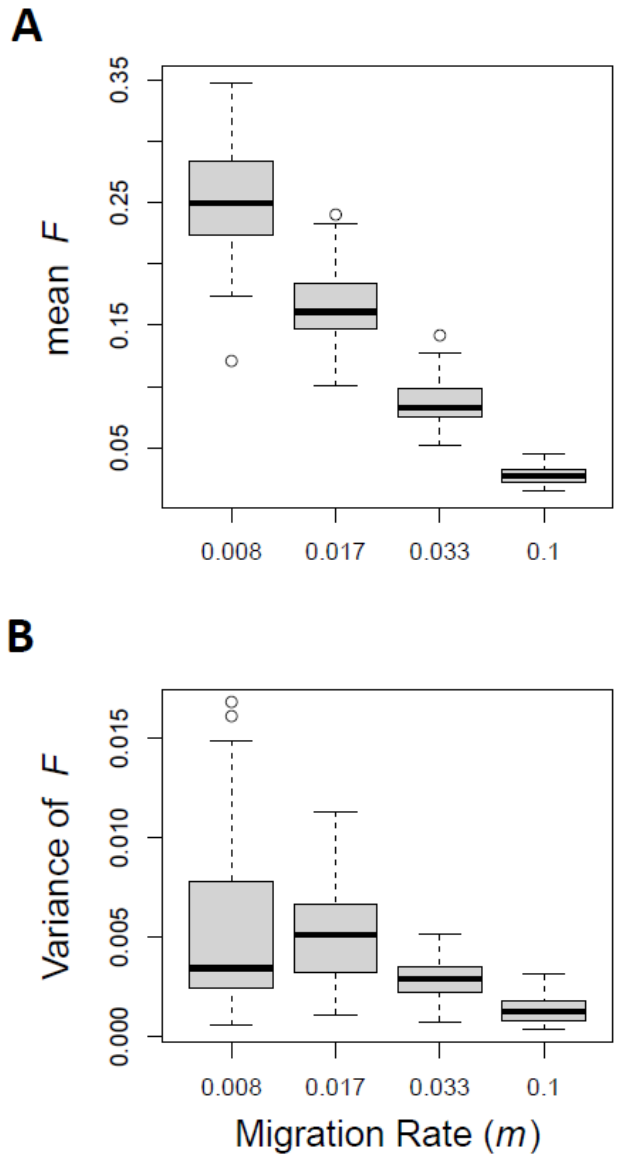

Figure 4-S1. The mean (A) and variance (B) of F versus the migration rates used in our simulations. The data shown are from 50 replicate simulations for each migration rate.

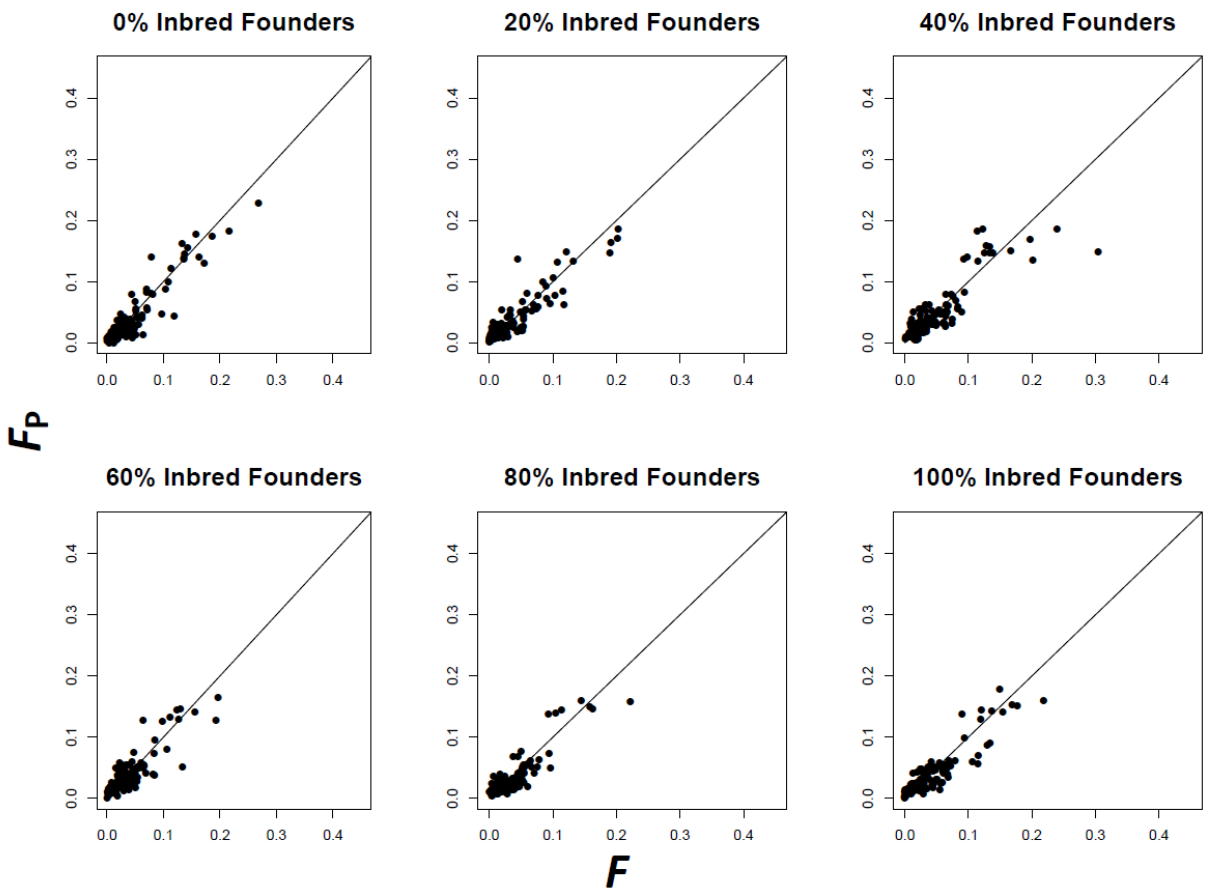

Figure 4-S2. F_P versus the true F in from pedigrees with 0-100% inbred ($F_P = 0.25$) founders. The solid line represents a slope of 1.0.

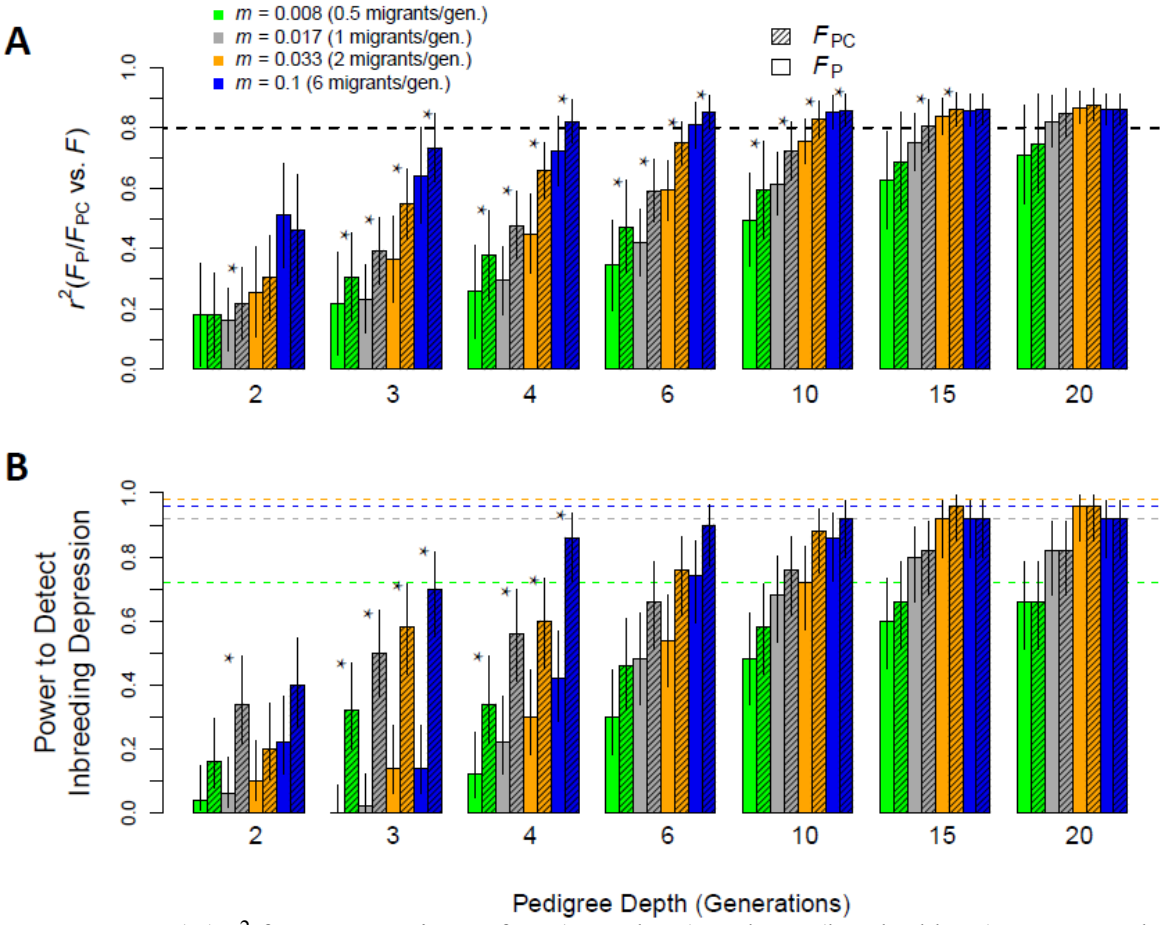

Figure 4-S3. (**A**) r^2 from regressions of F_P (open bars) and F_{PC} (hatched bars) versus F plotted against pedigree depth. (B) The power to detect inbreeding depression using F_P and F_{PC} plotted against the pedigree depth. The colored dashed lines represent the statistical power to detect inbreeding depression when using the true F in each of the simulated demographic scenarios. Results from simulations using different migration rates are represented by different color bars. Scenarios where the r^2 with F, or the power to detect inbreeding was statistically significantly different between F_P and F_{PC} are indicated with stars.

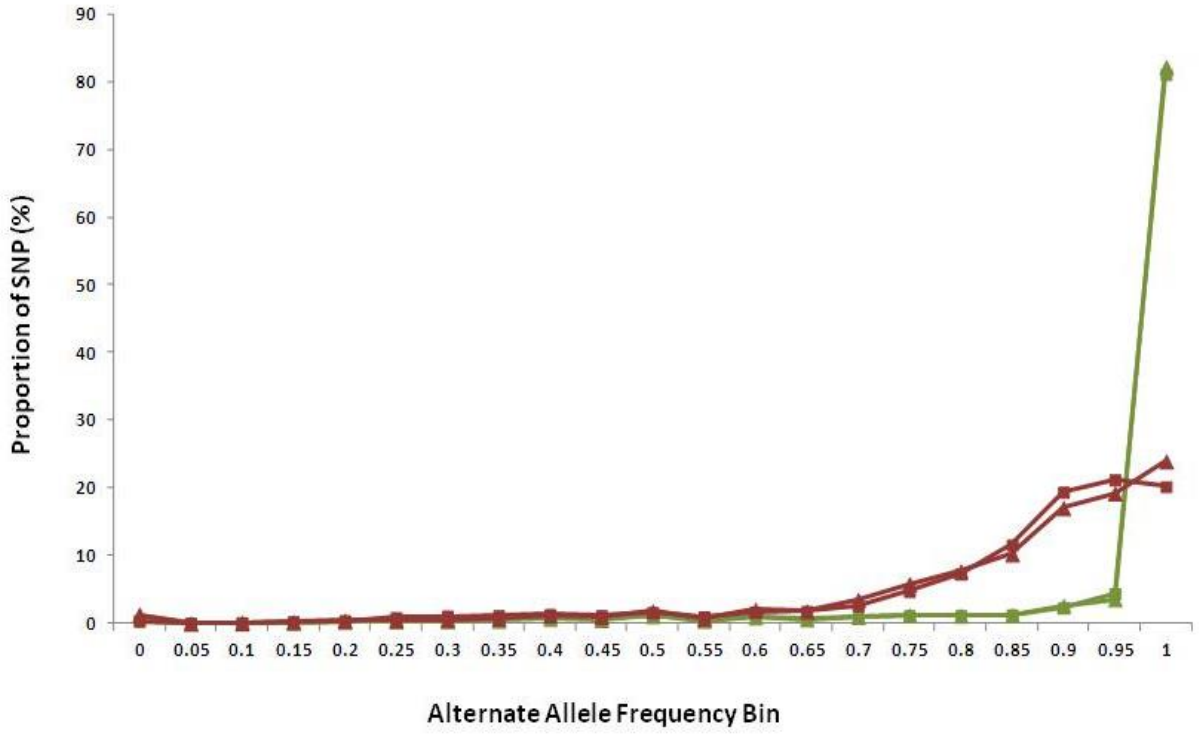

Figure 5-S1. Variants were called in pooled data from the South Teton (triangles) and North Teton (squares) populations. The proportion of total SNPs is given for the full range (0 – 1) of alternate allele (AF) frequency. The result using default SNVer settings for read mapping quality (mq = 20) and nucleotide base quality (bq = 17) is shown in red. The increase in SNP with high AF (> 0.80) prompted manual inspection of individual variants using IGV (Thorvaldsdóttir *et al.* 2013).This revealed a large number of positions were fixed between species, however were assigned an AF value < 1.0. Variant calling was subsequently performed using an elevated threshold for mapping quality (mq = 40) to exclude reads incorrectly positioned due to sequence divergence between *O. Canadensis* and *O. aries*. For reads successfully mapped, almost all positions within a read were used for variant detection by reducing the bq value to 2. The result using these modified SNVer settings (mq = 40; bq = 2) is shown for both pools (green). Greater than 80% of variants in each pool had AF = 1.0 and the distribution for the remaining variants is almost flat.

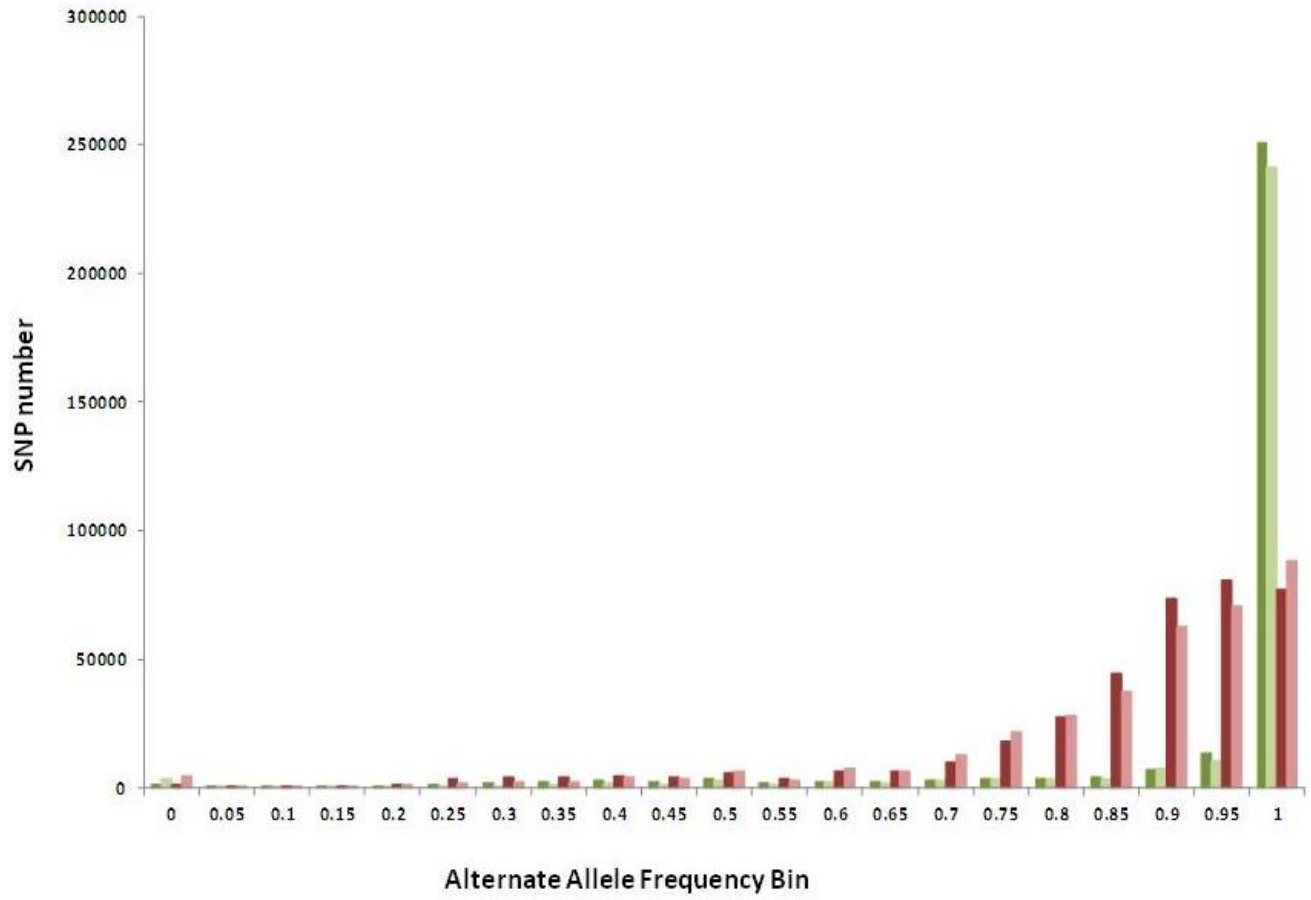

Figure 5-S2. The total number of SNP called is given for each alternate allele frequency bin (AF) following SNP calling. The data is the same as given in Figure S1 however the absolute number of SNP is given in the Y axis. SNVer was used with default (red) or modified parameters (green) on two Bighorn populations: South Teton (lighter shade) and North Teton (darker shade). The modified parameters greatly reduced the number of SNP called in AF bins 0.70 – 0.95, while dramatically increasing the number of SNP (≈ 250,000) called with AF = 1. This reflects the large number of positions that are monomorphic within pools and fixed for an allele that is different to that residing within the reference genome assembly built from a domestic sheep.

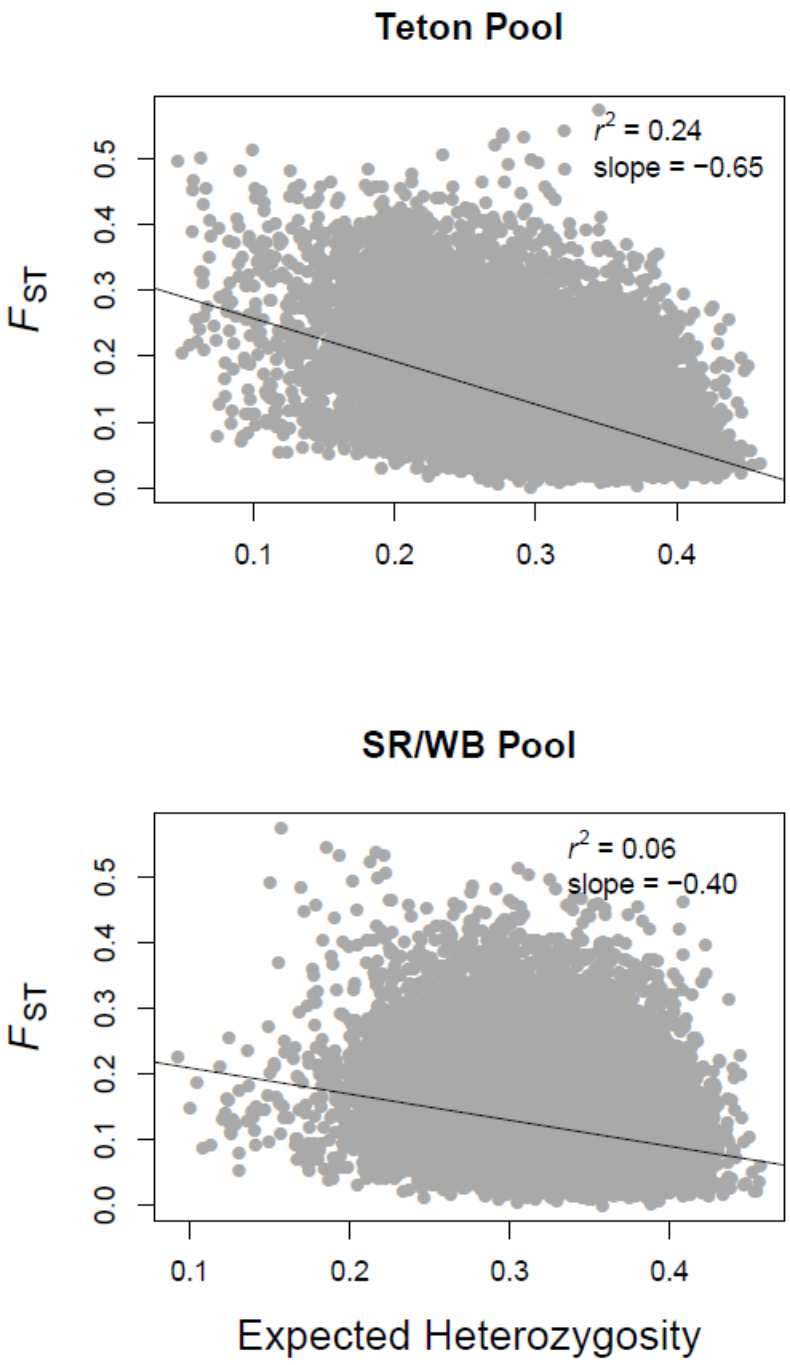

Figure 5-S3. F_{ST} versus H_P for all sliding windows across the bighorn sheep genome. Data from the Teton pool are shown in the top panel and data from the SR/WB pool are shown in the bottom panel. The fitted lines are from linear regression of F_{ST} versus expected heterozygosity. Both linear regression models were highly statistically significant ($P << 0.001$)

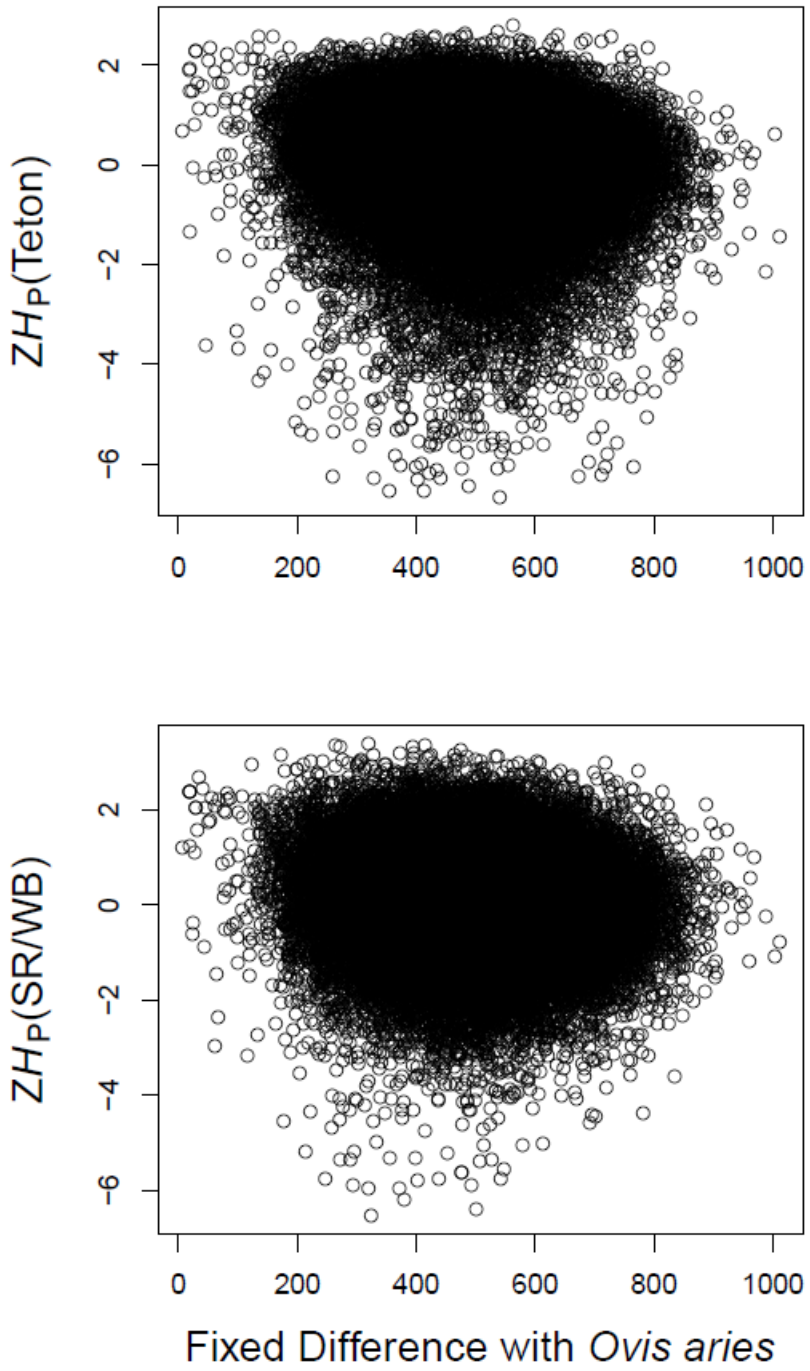

Figure 5-S4. ZH_P from the Teton pool versus the number of fixed differences between bighorn and the domestic sheep reference genome sequence. We excluded windows with fewer than 200 fixed differences from consideration as putative selective sweeps.

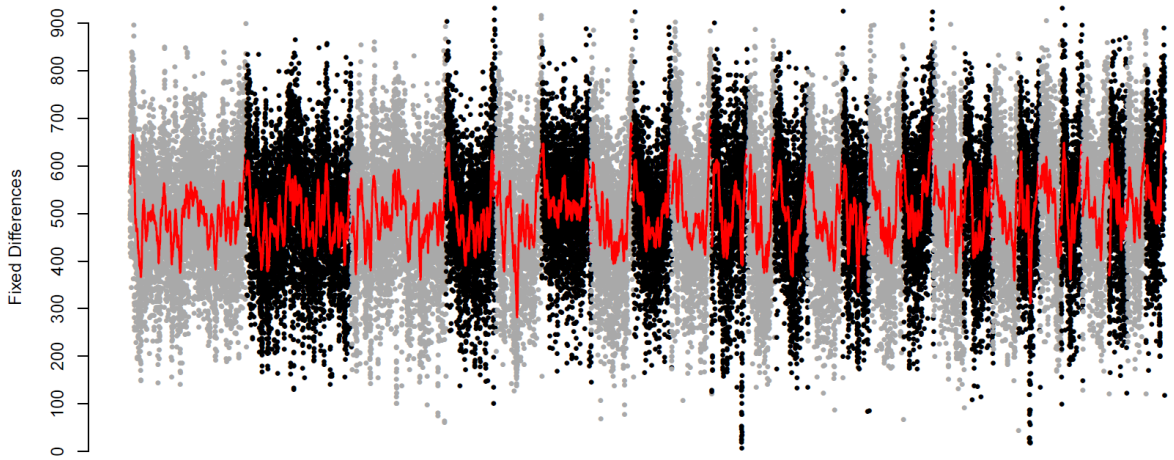

Figure 5-S5. The number of fixed differences between Teton bighorn and the domestic sheep reference genome sequence in 100 Kb sliding windows. The red line is the rolling mean over 100 windows. Chromosomes 1-26 are arranged left to right and are distinguished by bands of gray and black points.

Table 5-S1. Number of single nucleotide polymorphisms (SNPs)by chromosome number (after all filtering steps) across all 26 autosomes.

Chromosome	SNPs
1	283323
2	244915
3	216929
4	128111
5	111328
6	123827
7	103156
8	94989
9	103082
10	96742
11	65119
12	87459
13	85201
14	65171
15	88237
16	78674
17	74719
18	76991
19	63199
20	63009
21	55536
22	53288
23	68998
24	44372
25	48116
26	48373
Total:	**2572864**

Table 5-S2. Sequence summary statistics for each of the six individual DNA pools.

Population	Abbrev	Pool_N	Altitude	Disease	Range	Total Mb	Aligned Mb	Aligned (%)	SNP
Sun River Gibson Dam	SRGD	8	Low	Die-offs	Migratory	22,082	17,761	80	
Sun River Other	SRO	10	Low	Die-offs	Migratory	40,512	28,804	71	
North Teton	NT	10	High	Free	Non-migratory	66,296	52,016	78	
South Teton	ST	9	High	Free	Non-migratory	44,422	35,240	79	
Whiskey Basin New	WBN	11	Mid - High	Die-offs	Migratory	41,242	33,084	80	
Whiskey Basin Old	WBO	10	Mid - High	Die-offs	Migratory	41,854	32,433	77	
Total						**256,408**	**199,337**	**78**	**2572864**

Table 5-S3. Pairwise estimates of F_{ST} between individual DNA pools. The F_{ST} estimates presented here are from 1.72 million SNPs identified via variant calling conducted on each of the six individual pools individually.

Pool Combination	$\boldsymbol{F_{ST}}$
Tetons vs. Sun River and Whiskey Basin	0.119
Sun River Gibson vs. Sun River Other	-0.045
Whiskey Old vs. Whiskey New	-0.016
Whiskey Basin vs. Sun River	0.138
North Teton vs. South Teton	0.098

Table 5-S4. All outlier regions containing genes. Regions were identified on the basis of ZH_P in the Teton pool (**A**), ZH_P in the SR/WB pool (**B**), and ZF_{ST} between pools (**C**). The z-scores listed are the lowest (for ZH_P) or highest (for ZF_{ST}) among the 100 Kb windows making up an outlier region. Outlier regions are defined as contiguous blocks of outlier 100 Kb windows.

Chr	Start(Mb)	Stop(Mb)	Size(Kb)	Region_N	nSNP	Z-score	Genes
(A) Teton Pool (TET) tested using ZH_P							
1	62.15	62.30	150	1	210	-6.188	COL24A1
1	258.25	258.35	100	5	188	-6.228	DSCAM
2	162.00	162.10	100	7	113	-5.019	SLC25A23
2	238.55	238.70	150	12	52	-5.435	AHDC1, FGR
3	32.10	32.35	250	13	79	-6.528	ADCY3,CENPO,NCOA1,DNAJC27
3	79.95	80.25	300	15	140	-5.314	PPM1B,SLC3A1,RPL7,LRPPRC
3	141.70	141.80	100	16	398	-5.414	NELL2
3	191.10	191.25	150	17	97	-6.040	SOX5
3	197.65	198.00	350	18	99	-5.629	PGRMC2
4	112.45	112.55	100	19	225	-5.166	A4IFP6, GIMAP1
5	9.10	9.20	100	20	242	-5.396	ZNF333
5	49.55	49.65	100	22	71	-5.346	PCDHB17,PCDHB4,PCDHB5,PCDHB6
5	79.50	79.65	150	24	252	-6.324	ATP6AP1L,RPS23,GPR98
5	88.00	88.10	100	25	63	-5.570	GPR98
7	24.35	24.50	150	26	481	-6.053	OR4N2
8	10.85	11.10	250	27	84	-6.254	C6ORF174,SNAP91,ECHDC1
8	53.55	53.70	150	29	67	-6.009	PTPRK
8	82.90	83.00	100	30	93	-5.622	IGF2R
9	27.45	27.55	100	31	134	-5.636	E5KBL6
9	28.95	29.05	100	32	97	-5.978	ANXA13,FAM91A1
10	29.45	29.55	100	33	24	-6.055	RXFP2,EEF1A1
12	44.55	44.70	150	35	59	-6.248	CAMTA1
14	52.55	52.65	100	38	63	-5.032	HIF3A,CCDC61,NOVA2,PGLYRP1,IGFL1
15	13.80	13.95	150	40	101	-5.552	MTMR2,CEP57
15	16.40	16.50	100	41	61	-5.223	ALKBH8
16	31.65	31.90	250	42	159	-5.652	CCDC152,SEPP1,GHR
21	31.50	31.60	100	47	67	-5.783	ARHGAP32,KCNJ5
26	36.05	36.20	150	48	315	-6.056	HOOK3,THAP1,RNF170
(B) Sun River Whiskey Basin Pool (SRWB) tested using ZH_P							
1	69.15	69.25	100	1	154	-5.911	EVI5, RPL5, FAM69A
1	77.50	77.65	150	2	127	-6.187	DPH5
7	69.95	70.05	100	9	121	-5.179	MNAT1, TRMT5, SLC38A6
10	25.80	25.90	100	11	220	-5.569	DCLK1
15	54.45	54.55	100	15	309	-5.759	ACER3

24	18.60	18.70	100	19	51	-5.951	DCUN1D3, F1RPB4, TIMM44, LYRM1

(C) TET and SRWB Pools compared using ZF_{ST}

1	104.65	104.75	100	2	58	5.305	MEX3A,LMNA,SEMA4A,UBQLN4, LAMTOR2,RAB25
1	208.15	208.30	150	4	61	5.496	TBL1XR1
2	65.75	65.85	100	5	33	6.038	SMC5,KLF9,F1MR21
2	193.30	193.60	300	7	86	6.687	TMEFF2
3	32.10	32.30	200	9	61	5.540	ADCY3,F1SDL0,CENPO,NCOA1, DNAJC27,F1AXJ8
5	49.65	49.75	100	12	45	5.398	PCDHB18,RPSA,PCDHB15,PCDHB7, PCDHB13,PCDHB16,PCDHB17
6	71.30	71.45	150	13	52	6.416	CEP135,ZC3H14,EXOC1
8	53.60	53.75	150	14	77	6.110	PTPRK
8	55.10	55.25	150	15	44	5.780	C6ORF191
8	82.00	82.65	650	16	70	5.516	SOD2,TAGAP
12	44.60	44.70	100	17	52	5.332	CAMTA1
14	23.90	24.00	100	18	41	5.944	GNA0
16	31.70	31.85	150	20	98	5.783	CCDC152,SEPP1,GHR
20	16.70	16.80	100	22	56	5.098	KLC4,MRPL2,MEA1,KLHDC3,CUL7, PPP2R5D
20	17.20	17.30	100	23	53	5.261	MRPS18A,DTD1,RSPH9,GTPBP2, MAD2L1BP,POLH
20	44.50	44.60	100	24	69	5.059	GCNT2
21	44.35	44.50	150	25	51	5.706	KDM2A,PC,ADRBK1
22	31.45	31.60	150	27	75	6.165	POL
22	33.20	33.35	150	28	127	6.599	NHLRC2,DCLRE1A
26	18.00	18.10	100	29	48	5.008	FGL1,PCM1
26	34.90	35.00	100	30	49	5.147	GINS4,GOLGA7
26	36.20	36.30	100	31	152	5.127	SGK196,HGSNAT,FNTA

Table 5-S5. All outlier regions identified on the basis of ZH_P in the Teton pool.

Chr	Start(Mb)	Stop(Mb)	Size(Kb)	nSNP	ZH_P_Teton	Genes
1	62.15	62.30	150	210	-6.188	COL24A1
1	75.30	75.40	100	142	-5.285	
1	173.80	173.90	100	63	-5.073	
1	223.35	223.45	100	277	-5.265	
1	258.25	258.35	100	188	-6.228	DSCAM
1	259.10	259.20	100	196	-5.663	
2	162.00	162.10	100	113	-5.019	SLC25A23
2	194.15	194.30	150	133	-5.492	
2	195.00	195.25	250	101	-5.577	
2	196.40	196.50	100	92	-5.284	
2	209.95	210.15	200	61	-6.552	
2	238.55	238.70	150	52	-5.435	AHDC1, FGR
3	32.10	32.35	250	79	-6.528	ADCY3, CENPO, NCOA1, DNAJC27
3	51.95	52.05	100	129	-5.096	
3	79.95	80.25	300	140	-5.314	PPM1B,SLC3A1,RPL7,LRPPRC
3	141.70	141.80	100	398	-5.414	NELL2
3	191.10	191.25	150	97	-6.040	SOX5
3	197.65	198.00	350	99	-5.629	PGRMC2
4	112.45	112.55	100	225	-5.166	A4IFP6, GIMAP1
5	9.10	9.20	100	242	-5.396	ZNF333
5	39.55	39.65	100	384	-5.029	
5	49.55	49.65	100	71	-5.346	PCDHB17,PCDHB4,PCDHB5,PCDHB6
5	75.30	75.40	100	77	-5.361	
5	79.50	79.65	150	252	-6.324	ATP6AP1L,RPS23,GPR98
5	88.00	88.10	100	63	-5.570	GPR98
7	24.35	24.50	150	481	-6.053	OR4N2
8	10.85	11.10	250	84	-6.254	C6ORF174,SNAP91,ECHDC1
8	42.25	42.35	100	251	-5.244	
8	53.55	53.70	150	67	-6.009	PTPRK
8	82.90	83.00	100	93	-5.622	IGF2R
9	27.45	27.55	100	134	-5.636	E5KBL6
9	28.95	29.05	100	97	-5.978	ANXA13,FAM91A1
10	29.45	29.55	100	24	-6.055	RXFP2,EEF1A1
12	6.50	6.60	100	317	-5.268	
12	44.55	44.70	150	59	-6.248	CAMTA1
14	29.45	29.55	100	155	-5.617	
14	32.45	32.55	100	233	-5.428	
14	52.55	52.65	100	63	-5.032	HIF3A,CCDC61,NOVA2,PGLYRP1,IGFL1
15	12.50	12.75	250	92	-5.454	

15	13.80	13.95	150	101	-5.552	MTMR2,CEP57
15	16.40	16.50	100	61	-5.223	ALKBH8
16	41.50	41.60	100	192	-5.760	
16	52.10	52.20	100	408	-5.259	
19	9.70	9.80	100	89	-6.666	
20	22.90	23.05	150	111	-6.081	
21	31.50	31.60	100	67	-5.783	ARHGAP32,KCNJ5
26	36.05	36.20	150	315	-6.056	HOOK3,THAP1,RNF170
26	36.45	36.55	100	68	-5.799	

Table 5-S6. Outlier regions identified on the basis of ZH_P in the SR/WB pool.

Chr	Start(Mb)	Stop(Mb)	Size(Kb)	nSNPHp	ZH_P_SRWB	Genes
1	69.15	69.25	100	154	-5.911	EVI5, RPL5, FAM69A
1	77.50	77.65	150	127	-6.187	DPH5
1	113.35	113.45	100	238	-5.341	
1	130.45	130.55	100	236	-5.371	
3	114.95	115.05	100	225	-5.062	
4	54.70	54.80	100	195	-5.318	
5	11.65	11.80	150	268	-6.404	
5	66.85	67.00	150	137	-6.525	
7	69.95	70.05	100	121	-5.179	MNAT1, TRMT5, SLC38A6
8	42.45	42.55	100	320	-5.207	
10	25.80	25.90	100	220	-5.569	DCLK1
10	62.35	62.45	100	234	-5.017	
11	0.05	0.25	200	123	-5.908	
15	11.40	11.50	100	567	-5.350	
15	54.45	54.55	100	309	-5.759	ACER3
16	53.30	53.40	100	204	-5.810	
18	47.40	47.55	150	219	-5.758	
19	54.60	54.75	150	201	-5.622	
24	18.60	18.70	100	51	-5.951	DCUN1D3, F1RPB4, TIMM44, LYRM1

Table 5-S7. Outlier regions identified on the basis of ZF_{ST} estimated between the Teton and SR/WB pools.

Chr	Start(Mb)	Stop(Mb)	Size(Kb)	nSNP	ZF_{ST}	Genes
1	94.95	95.05	100	40	5.250	
1	104.65	104.75	100	58	5.305	MEX3A,LMNA,SEMA4A,UBQLN4,LAMTOR2,RAB25
1	204.15	204.25	100	68	5.119	
1	208.15	208.30	150	61	5.496	TBL1XR1
2	57.70	57.80	100	93	5.373	
2	65.75	65.85	100	33	6.038	SMC5,KLF9,F1MR21
2	180.30	180.40	100	86	5.824	
2	193.30	193.60	300	86	6.687	TMEFF2
2	193.90	195.30	1400	77	5.830	
3	32.10	32.30	200	61	5.540	ADCY3,F1SDL0,CENPO,NCOA1,DNAJC27,F1AXJ8
3	52.05	52.15	100	95	5.423	
4	8.90	9.00	100	48	5.062	
5	49.65	49.75	100	45	5.398	PCDHB18,RPSA,PCDHB15,PCDHB7,PCDHB13, PCDHB16,PCDHB17
6	71.30	71.45	150	52	6.416	CEP135,ZC3H14,EXOC1
8	53.60	53.75	150	77	6.110	PTPRK
8	55.10	55.25	150	44	5.780	C6ORF191
8	82.00	82.65	650	70	5.516	SOD2,TAGAP
12	44.60	44.70	100	52	5.332	CAMTA1
14	23.90	24.00	100	41	5.944	GNA0
15	11.35	11.55	200	466	7.258	
16	31.70	31.85	150	98	5.783	CCDC152,SEPP1,GHR
19	9.70	9.85	150	62	6.285	
20	16.70	16.80	100	56	5.098	KLC4,MRPL2,MEA1,KLHDC3,CUL7,PPP2R5D
20	17.20	17.30	100	53	5.261	MRPS18A,DTD1,RSPH9,GTPBP2,MAD2L1BP,POLH
20	44.50	44.60	100	69	5.059	GCNT2
21	44.35	44.50	150	51	5.706	KDM2A,PC,ADRBK1
22	27.50	27.60	100	52	5.397	
22	31.45	31.60	150	75	6.165	POL
22	33.20	33.35	150	127	6.599	NHLRC2,DCLRE1A
26	18.00	18.10	100	48	5.008	FGL1,PCM1
26	34.90	35.00	100	49	5.147	GINS4,GOLGA7
26	36.20	36.30	100	152	5.127	SGK196,HGSNAT,FNTA