Evaluation of inbreeding depression and the comparison of genomic inbreeding measures in domestic cattle

by

Douglas W Bjelland

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

Doctor of Philosophy (Dairy Science)

at the

University of Wisconsin-Madison

2013

Date of final oral examination: 7/15/13

The dissertation is approved by the following members of the Final Oral Exam Committee:

Kent A. Weigel, Professor, Dairy Science
Brian Kirkpatrick, Professor, Animal Science
Guilherme J. M. Rosa, Associate Professor, Animal Science
Corinne D. Engelman, Assistant Professor, Population Health Sciences
Mark W. Craven, Professor, Biostatistics and Medical Informatics

TABLE OF CONTENTS

List of tables iv

List of figures vi

CHAPTER 1

Introduction 1

CHAPTER 2

Review of Literature 5

2.1 Calculation of inbreeding from pedigrees 6

2.2 Effects of increases in pedigree inbreeding 7

2.3 Limitations of pedigree inbreeding 11

2.4 Use of genomic selection in dairy cattle and its effect on inbreeding 13

2.5 Calculating inbreeding from genomic information 17

2.6 Quantification of genomic inbreeding and genomic inbreeding depression 21

2.7 Reproduction challenges in Holstein cattle 23

References 26

CHAPTER 3

Evaluation of inbreeding depression in Holstein cattle using whole genome SNP markers and alternative measures of genomic inbreeding 35

Abstract 36

Introduction 36

Materials and Methods 40

Results 44

Discussion 47

Conclusions 52

Acknowledgements 53

References 54

CHAPTER 4

Comparison of genomic inbreeding in Holstein, Jersey, Angus, and Nelore cattle using dense and reduced SNP marker panels 66

Abstract 67

Introduction 68

Materials and Methods 70

Results and Discussion 75

Conclusions 82

Acknowledgements 83

References 84

CHAPTER 5

Comparison of genomic inbreeding within a family-based structure in Holstein cattle 95

Abstract 96

Introduction 97

Materials and Methods 99

Results and Discussion 103

Conclusions 109

Acknowledgements 110

References 111

CHAPTER 6

Overall summary 122

List of tables

Table 3.1. Estimates of inbreeding depression for production traits, expressed as change in phenotype per 1% increase in percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}). 60

Table 3.2. Estimates of inbreeding depression for reproductive traits, expressed as change in phenotype per 1% increase in percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}). 61

Table 3.3. Estimates of inbreeding depression for linear type traits on a 50-point scale, expressed as change in phenotype per 1% increase in percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}). 62

Table 3.4. Estimates of inbreeding depression for all significant traits, expressed as the difference in predicted phenotype between plus or minus 2 SD from the mean for percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}). 63

Table 4.1. Number of animals and SNP before and after data editing. 88

Table 5.1. All possible genotypes in progeny for given sire and dam genotypes, with probability the genotype is homozygous. 115

Table 5.2. F_{PHE} and F_{PH} values calculated using the simple method and F_{PHE}, F_{PH}, F_{ROHE}, and F_{ROH} values calculated using simulated progeny with zero crossover events. 116

List of figures

Figure 3.1. Description of the process for discovery of runs of homozygosity (ROH) using a sliding window of SNP markers along the chromosome, as implemented with PLINK software (Purcell et al., 2007). 64

Figure 3.2. Frequency distribution of animals in the present study, according to (**A**) percent homozygosity (F_{PH}), (**B**) runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and (**C**) inbreeding coefficient derived from a genomic relationship matrix (F_{GRM}) using the method of VanRaden et al. (2011a). 65

Figure 4.1. Description of the process for discovery of runs of homozygosity (ROH) using a sliding window of SNP markers along the chromosome, as implemented with PLINK software (Purcell et al., 2007). 89

Figure 4.2. Frequency distributions for percent homozygosity, F_{ROH4}, and F_{ROH8} derived from 50K SNP panel. 90

Figure 4.3. Frequency distributions for percent homozygosity from 50K, equally spaced 6K, and low density SNP panels. 91

Figure 4.4. Frequency distributions for F_{ROH4} calculated from 50K, equally spaced 6K, and low density SNP panels. 92

Figure 4.5. Frequency distributions for F_{ROH8} calculated from 50K, equally spaced 6K, and low density SNP panels. 93

Figure 4.6. Relationship between F_{ped} and percent homozygosity, and F_{ped} and F_{ROH4} for elite Holstein (A and B, respectively) and Jersey (C and D, respectively) cattle. 94

Figure 5.1. F_{PHE} predicted using the simple method compared to actual F_{PH} of live progeny (A) and a histogram (B) with F_{PHE} predicted using the simple method (light bars) and F_{PH} from actual live progeny (dark bars). 117

Figure 5.2. F_{PHE} predicted using simulated progeny to actual F_{PH} of live progeny (A) and a histogram (B) with F_{PHE} predicted using the simulated progeny (light bars) and F_{PH} from actual live progeny (dark bars). 118

Figure 5.3. F_{ROHE} predicted using the simulated progeny compared to actual F_{ROH} of live progeny (A) and a histogram (B) with F_{ROHE} predicted using the simulated progeny (light bars) and F_{ROH} from actual live progeny (dark bars). 119

Figure 5.4. Effects of changes in recombination rate when simulated progeny on the average predicted F_{PH} (A), average predicted F_{ROH} (B), and the correlation between actual progeny F_{PH} and F_{ROH} and average predicted F_{PH} and F_{ROH} (C). 120

Figure 5.5. Scatterplot of F_{ROH} (A) and F_{PH} (C) between sibling pairs and the absolute values of the differences of F_{ROH} (B) and F_{PH} (D) between siblings. 121

CHAPTER 1

Introduction

The research presented herein focuses on the calculation and use of genomic inbreeding measures in domestic cattle. Pedigree inbreeding has been constantly increasing in dairy cattle populations, and, along with this increase, negative effects due to increases in inbreeding have been discovered. Recently, genomic tools have been developed and have been utilized extensively in domestic animal species, especially dairy cattle, for the use of genomic selection. The calculation and use of genomic inbreeding values from this information is also possible and potentially beneficial to animal breeders.

Chapter 2 provides a review of literature relevant to the subsequent research topics. First, an overview of deriving pedigree inbreeding (F_{ped}) values is presented. The following section focuses on the effects of F_{ped} on economically important traits in Holstein cattle, as well as other domestic cattle breeds. Next, the use of genomic selection in dairy cattle is presented, followed by its possible effects on overall inbreeding within the Holstein breed. Subsequent sections discuss measures of genomic inbreeding which have been developed in both animal and human populations and the effects of increases in genomic inbreeding on quantitative traits in humans. The final section of this chapter focuses on the discovery and consequences of markers and haplotypes which negatively affect fertility in dairy cattle.

The research presented in Chapter 3 presents the results of increases of several measures of genomic inbreeding on economically important traits in Holstein cattle. A total of 5,853 Holstein cattle were genotyped for 54,001 single nucleotide polymorphisms (SNP), with 2,913 of these being cows with phenotypic information such as single lactation milk yields, reproductive records, and linear type conformation. From the genotypic information, three separate genomic inbreeding measures were derived: the first was the percentage of homozygous SNP markers (F_{PH}), genomic inbreeding derived from a genomic relationship matrix (F_{GRM}), and genomic

inbreeding derived from runs of homozygosity (F_{ROH}). The effects of each of the three measures of genomic inbreeding on the economically important traits were then derived.

Chapter 4 presents research comparing genomic inbreeding measures between several domestic cattle breeds and genetic groups, as well as between varying densities of SNP panels. A total of 54,001 SNP markers were available for 6,600 commercial Holsteins, 2,402 Angus, and 2,302 Nelore cattle. Additionally, 43,485 SNP markers were available for 7,883 genetically elite Holsteins and 3,146 Jerseys. Two lower density SNP panels were derived for each of the genetic groups, the first including the 6,909 SNP contained within the Illumina BovineLD BeadChip (Illumina Inc, San Diego, CA) and the second containing roughly 6,400 equally spaced SNP. From each of the 3 SNP panels, F_{PH} values were calculated for each genetic group. Additionally, two measures of F_{ROH}, one with a minimum length of ROH of 4,000 kb and one with a minimum length of ROH of 8,000 kb, were calculated for each SNP panel within each genetic group. Comparisons between each of the genomic inbreeding measures calculated were then made.

The research presented in Chapter 5 aims to compare the genomic inbreeding expected between a sire-dam mating pair to the genomic inbreeding of their progeny. Genotype (43,485 SNP markers) and recent pedigree information was available for 11,484 Holstein cattle. A total of 374 sire-dam-progeny trios with genotype information were discovered for analysis. Several methods were developed to predict genomic inbreeding for the mating pair. The first method examines each SNP independently and determines the probability that the progeny will be homozygous at that specific locus. The calculation was compared to the F_{PH} values derived from the actual progeny of the mating pair. Another method utilized phased haplotypes from the sire-dam mating pair and simulated progeny which could be possible through this mating. Measures of F_{PHE} and F_{ROHE} were then calculated from the simulated progeny, averaged together, and

compared to the F_{PH} and F_{ROH} values of the actual progeny. The theory behind this comparison is that high inbreeding may be detrimental to the developing embryo, which may result in the actual progeny born having lower genomic inbreeding values than expected. Furthermore, a total of 3,906 full sibling pairs with genotype information were discovered. Comparisons were made between these sibling pairs to determine the variation present among genomic inbreeding due to Mendelian sampling.

CHAPTER 2

Review of Literature

Calculation of inbreeding from pedigrees

Inbreeding is defined as the mating of related individuals, with the key effect being the increase in homozygosity. The most commonly used measure of inbreeding is based on pedigree information, is defined as the proportion of genes in which an individual's parents share in common, and as F_{ped} (Wright, 1922). This measure is usually in reference to a base population in which all of the individuals are assumed to be unrelated. If the parents of an individual have alleles in common, it can be said that these alleles have been passed down from an ancestor who is shared between these two individuals. By tracing the pedigrees back to a common ancestor and computing probabilities at each segregation, the calculation of the amount of DNA which is identical by decent (**IBD**) can be made with the following formula:

$$F_{ped} = \sum_{i=1}^{n_a} [\left(\frac{1}{2}\right)^{n_a} (1 + F_a)]$$

where the sum is over all possible paths from parents to the common ancestor *a*, n_a is the number of individuals in that path, and F_a is the inbreeding coefficient of individual *a*. As F_{ped} produces a probability that a given gene is IBD, the values for F_{ped} range from 0 to 1.

As pedigrees have become large, with many inbred ancestors, calculation using the above method may be tedious. Henderson (1976) described an alternative tabular method to compute F_{ped} using a matrix of relationships. This matrix is computed recursively beginning with the base population, which has initial relationships among each other set to 0. The diagonal elements are assigned $1 + F_i$, where $F_i = a_{sd}/2$ and a_{sd} is the numerator relationship between the sire and dam of individual *i*. The off-diagonal elements are calculated as $½(a_{jsi} + a_{jdi})$ where s_i and d_i are the parents of *i* and *j* is younger than *i*. Any unknown or missing values are set to 0, meaning that an

individual in which only one of the parents is known is given the diagonal value of 1, ie. they are not inbred.

Effects of increases in pedigree inbreeding

The average inbreeding coefficient in US Holsteins has risen from 0.4% in 1970 to 5.8% in 2012 (USDA-AIPL, 2012). One reason for this increase may be the use of the animal model for genetic evaluation. Because the animal model uses all relationships among individuals, the related animals tend to rank together, and will result in the selection of closely related animals (Wiggans and VanRaden, 1995). The extensive use of artificial insemination and the intense selection pressure on bulls during this time period may also have contributed to a large increase of inbreeding values. Weigel (2001) noted that some popular Holstein bulls have sired as many as 250,000 milking daughters and 3,000 progeny tested sons worldwide. Also, of the nearly 5,000 progeny tested Holstein bulls at that time, nearly 50% were sired by the 10 most popular sires.

Increases in inbreeding, leading to decreases in the overall heterozygosity of individuals, have been known to cause decreases in fitness for some time. In fact, inbreeding studies in dairy cattle began in 1912, and Woodward and Graves (1946) analyzed successive inbreeding in US Holstein-Friesian cattle and its effects on reproduction, body size, conformation, and milk and fat production. Results of this study revealed that after each generation of inbreeding, average services per conception increased and average birth weight and body weight at various ages decreased. Average milk and butterfat production was not severely affected until cows exhibited large amounts of inbreeding. Furthermore, the most intensely inbred animals had a characteristic shape to the head, set to the ears, rougher coat, and sluggish gait which distinguished them from the more outbred animals. Even though these negative effects of inbreeding have been known for

some time, many studies analyzing the exact effects of increases in inbreeding among dairy cattle populations have been performed in the last 15 years.

Smith et al. (1998) analyzed the effects of inbreeding on lifetime performance of Holstein cows using data for milk production, reproduction, somatic cell score and linear type traits. A total of 2,610,123 cows were available for the analysis, having both F_{ped} coefficients and phenotypes. Per 1% increase in inbreeding, relative lifetime net income adjusted for opportunity cost was depressed by $14.79 in a fluid milk market and by $12.40 in a manufacturing milk market. Smith et al. (1998) also compared registered (full 5-generation pedigrees) and grade (incomplete pedigrees) cows. The incomplete pedigrees in grade cows resulted in poor estimates of inbreeding coefficients, and in turn, poor estimates of the inbreeding depression. When only registered cows were used in the analysis, relative lifetime net income over adjusted opportunity cost was depressed by $24.43 per 1% increase in inbreeding in a fluid market and by $21.78 in a manufacturing market. When analyzing fitness traits, much greater inbreeding depression was observed in the grade cows as well; age at first calving increased by 0.36 d, productive life decreased by 13.1 d, and total days in milk decreased by 10.3 d per 1% increase in inbreeding. Production traits also declined with an increase in inbreeding; total lifetime milk yield decreased by 358.4 kg, total lifetime fat yield decreased by 13.2 kg, and total lifetime protein yield decreased by 11.4 kg per 1% increase in inbreeding. Calving interval also increased by 0.26 d per 1% increase in inbreeding, while somatic cell score was unaffected. Linear type traits, such as stature, strength, and body depth, were also affected by increases in inbreeding.

Effects of inbreeding on production traits in a large sample of first lactation Canadian Holsteins were analyzed by Miglior et al. (1995). For each 1% increase in inbreeding, total

lactation milk yield decreased by 25 kg, total lactation fat yield decreased by 0.9 kg, and total lactation protein yield decreased by 0.8 kg.

Thompson et al. (2000) analyzed the effects of increases in inbreeding on both production traits and survival in US Holstein cows. This study examined varying severities of inbreeding and estimated single lactation milk yield losses of 35 kg per 1% increase in inbreeding for an F_{ped} ranging from 0 to 7%, and losses of 55 kg per 1% increase in inbreeding for F_{ped} values ranging from 7 to 10%. The chance of survival to lactations 2, 3, 4, and 5 also decreased as inbreeding increased. For a cow with an F_{ped} of 1%, the probability of initiating second lactation was 75%, third lactation was 54%, fourth lactation was 35%, and fifth lactation was 21%. By comparison, a cow with an F_{ped} of 10% had the probability of initiating second lactation of 66%, third lactation of 45%, fourth lactation of 25%, and fourth lactation of 10%.

Adamec et al. (2006) analyzed the effects of inbreeding on dystocia and stillbirth in US Holstein cows. In this study, only registered cows with complete 5 generation pedigrees were included. Results indicated that first parity heifers giving birth to bull calves had an increased probability of dystocia of 0.42% per 1% increase in F_{ped} while first parity heifers giving birth to heifer calves had an increased probability of dystocia of 0.30% per 1% increase in F_{ped}. The incidence of stillbirth for bull calves (+0.25% per 1% increase in F_{ped}) and heifer calves (+0.20% per 1% increase in F_{ped}) also increased in first parity heifers with an increase in inbreeding. Later parities were also negatively affected by increases in inbreeding for both probability of dystocia, ranging from an increase of 0.13 to 0.20% per 1% increase in F_{ped}, and probability of stillbirth, ranging from an increase of 0.005 to 0.05% per 1% increase in F_{ped}.

Further study on the effects of inbreeding on milk production, calving performance, fertility, and conformation traits in Irish Holstein-Friesian cows was performed by Mc Parland et

al. (2007). In this study, comparisons were made between cows that were the product of a half sib mating (F_{ped} = 12.5%) and cows which were not inbred (F_{ped} = 0%). The inbred cows had first lactation milk yield reduced by 61.8 kg, first lactation fat yield reduced by 5.3 kg, and first lactation protein yield reduced by 1.2 kg, compared to the non-inbred cows. The inbred cows also had a 2% greater probability of dystocia, 1% greater probability of stillbirth, 8.8 d increase in calving interval, and 2.5 d increase in age at first calving, compared to non-inbred cows. Survival to second lactation was also affected, where a non-inbred cow had a 4% greater chance to initiate a second lactation than an inbred cow.

Croquet et al. (2007) examined both linear and nonlinear effects of inbreeding on productions traits. The linear, quadratic, and cubic regression models all resulted in significant negative effects due to increases in inbreeding. The linear model predicted losses of 22.1 kg for first lactation milk yield, 1.1 kg for first lactation fat yield, and 0.7 kg for first lactation protein yield per 1% increase in F_{ped} for Holstein cows. The quadratic and cubic models indicated a nonlinear relationship between inbreeding levels and inbreeding depression, but between the F_{ped} values of 0 and 10%, where the vast majority of cows were present, the differences between linear and nonlinear models were negligible. Due to the small number of cows with an F_{ped} greater than 10%, the effects at this level were difficult to interpret accurately.

Negative effects of inbreeding have also been observed in other breeds of cattle. Carrillo and Siewerdt (2010) presented results from an Angus nucleus herd which had been closed to outside breeding for 70 years. Average F_{ped} for all animals in the herd was 6.8%, with the last generation of calves having an average F_{ped} of 12.0%. Increases in F_{ped} of 1% resulted in decreases in birth weight (2.19 kg), weaning weight (25.76 kg), adjusted 205 d body weight (25.28 kg), and average daily gain (0.11 kg/d).

Faria et al. (2009) analyzed the inbreeding accumulation in several Brazilian Zebu breeds. Results indicated that F_{ped} values in Nelore cattle increased from an average of 0.9% for animals born between 1979 and 1983 to an average of 2.1% for animals born between 1994 and 1998. Santana et al. (2010) analyzed the effects of increases in F_{ped} on growth and reproduction traits in Nelore cattle, where weaning weight, weight gain from weaning to 18 months of age, hip height, scrotal circumference, probability of heifer pregnancy at 18 months of age, and stayability were all negatively affected by increases in F_{ped}.

Limitations of pedigree inbreeding

Although F_{ped} has shown to be useful in the estimation of inbreeding depression, and it has been widely used to measure the negative effects of increases in inbreeding on economically important traits in domestic cattle, several limitations exist. First, the estimates of F_{ped} are only as accurate as the pedigrees that are used to calculate them. Ron et al. (1996) found a sire-daughter misidentification rate of 5.2% in 173 Israeli cows using microsatellite markers. Beechinor and Kelly (1987) found misidentification rates of 8 to 20% in dairy cattle in Ireland. Gelderman et al. (1986) determined sire-daughter misidentification using blood groups and several biochemical polymorphisms in a group of German bulls. Of the 15 test bulls with a total of 1,221 daughters (between 53 and 99 per bull), an overall misidentification rate of 13.2% was found, with individual bull misidentification rates between 4 and 23%. Christensen et al. (1982) discovered misidentification rates of between 5 and 15% in Danish dairy cattle. Dechow et al. (2008) analyzed paternity verification data from Alta Advantage progeny testing herds from Alta Genetics Inc. (Watertown, WI) using a 15-microsatellite marker analysis and herds from the Accelerated Genetics (Baraboo, WI) PACE young sire program using 6 to 9 microsatellite markers and 32 single nucleotide polymorphism (**SNP**) markers. Herds used in this test were

larger than average and were either enrolled in a progeny test program or considering using progeny test semen. The average misidentification rate among the 396 herds was 26%, whereas individual herds with greater than 50 cows had misidentification rates of 1.5 to 50%. Also, Visscher et al. (2002) discovered a sire misidentification rate of 10% in 568 UK dairy cows and 96 bulls using 11 unlinked microsatellite markers.

Although errors have been shown previously to be present in dairy cattle pedigrees, genomic tools such as the Illumina Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA) can be used to correct these errors. This practice has already been implemented by the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) and presented by Wiggans et al. (2011).

Even if pedigrees can be corrected, F_{ped} values are still just an estimate of the percentage of the genome which is IBD. This estimate can vary drastically, as presented by Carothers et al. (2006). First, consider an individual with a single loop of inbreeding; its parents share one common ancestor. Let T denote the length of the genome that is identical by descent and G denote the total length of the genome (both in Morgans, M). The genome is then covered by n evenly spaced markers, at a spacing of G/n. Then assign a random Bernouilli variable δ_i which takes the value 1 if the ith marker is IBD and 0 otherwise with $Pr(\delta_i = 1) = F_{ped}$. Then,

$$T = \frac{G}{n}\sum_{i} \delta_i$$

and

$$Exp(T) = \frac{G}{n}\sum_{i} Exp(\delta_i) = \frac{G}{n}F_{ped}n = GF_{ped}$$

We then can denote the realized proportion of the IBD genome to be f,

$$f = \frac{T}{G}$$

with

$$Exp(f) = F_{ped}$$

and a variance of

$$Var(f) \approx \frac{2F_{ped}(1 - F_{ped})}{\rho G}$$

where $\rho = n_p + n_m$, or the sum of the number of meioses from the common individual to the ancestral pair in the paternal and maternal lines, respectively. Using this formula for variance, we can determine that an individual which is the product of a first cousin mating, with an F_{ped} equal to 0.0625 (or 6.25%), $n_p = n_m = 3$, and assume G = 33 M,

$$Var(f) \approx \frac{2(0.0625)(1 - 0.0625)}{6(33)} = \frac{0.117}{198} = (0.0243)^2$$

Alternatively, we can look at an individual which is the product of a double second cousin mating. In this case, F_{ped} would also equal 0.0625, but n_p and n_m would each equal 4, denoting the difference in how the IBD segments of DNA came together in the current individual. The variance for this individual would be,

$$Var(f) \approx \frac{2(0.0625)(1 - 0.0625)}{8(33)} = \frac{0.117}{264} = (0.0211)^2$$

In either case, the standard deviation of the amount of DNA which is IBD can vary greatly from what is calculated from F_{ped}. Tools utilizing genomic information, such as SNP, should be able to more accurately detect this variation and provide a more precise measure of the amount of DNA which is IBD for a specific individual.

Use of genomic selection in dairy cattle and its effects on inbreeding

Traditional breeding value estimation in dairy cattle has involved utilizing phenotype information of individuals and their relatives and is commonly calculated using best linear unbiased prediction (**BLUP**; Henderson, 1984). Meuwissen et al. (2001) demonstrated how

prediction of breeding values could be performed using dense genotypic markers rather than pedigree information. This would allow breeding values to be predicted for animals which did not have any performance data or progeny of their own; an estimation which is not reliable using pedigree-based BLUP. This study proposed using Bayesian methods that assume a prior distribution of the variance associated with each genomic marker. Effects of specific markers would be estimated from a reference population of animals with performance data of progeny and summed together to predict the breeding values of young animals. This would allow a substantial increase in genetic gain, especially in dairy cattle, where bulls could be utilized based on their genomic prediction once they were sexually mature at just over a year of age, compared to around 5 years of age using the traditional BLUP method. The main reason for the large increase in genetic gain with genomic selection method is simply due to the dramatic decrease in generation interval of the sires used to produce young males or females.

VanRaden (2008) performed a simulation analysis with 2,967 bulls and 50,000 SNP markers randomly distributed across 30 chromosomes. Linear and nonlinear predictions were made with results indicating that genotyping provided information equivalent to about 20 daughters with phenotypic records, although the simulation did not account for linkage disequilibrium (**LD**) and results may be underestimated. The dairy genetics industry used this information to develop genomic breeding values once commercially available genotyping tools, such as the Illumina Bovine SNP50 BeadChip, became available. Other studies (Cole et al., 2009) confirmed that, apart from genes such as *DGAT1* (Grisart et al., 2004) and several large effect markers on BTA6 (Cohen-Zinder et al., 2005), the majority of traits are the product of a large number of genes with small additive effects, rather than few genes with large effects.

Much of the initial use of genomic information in dairy cattle focused on the identification and selection of elite bulls. Once the genotyping technology became both cheaper and more widely accepted, the question of how to use this technology on commercial dairies arose. Weigel et al. (2012) simulated different possible strategies to implement genomic selection on cows, heifers, and calves on commercial dairy farms. Results indicated that the genetic gains made by genotyping replacement heifers generally exceeded the costs of genotyping the animals. The gains were much greater when the animals had missing or incomplete pedigrees. Genotyping lactating cows with production records generally did not provide the same positive economic benefit as genotyping replacement heifers. If accurate pedigrees were available, presorting the replacement heifers and genotyping either the top genetic group to identify potential heifers to use as embryo donors or genotyping the bottom genetic group to identify potential cull heifers had the greatest economic potential. As more commercial dairy cows are being genotyped, genomic inbreeding measures, and the genomic relationships between cows and potential mates can be incorporated into commercial breeding programs.

As the use of genomic selection in dairy cattle and other livestock species has become more prevalent, questions have arisen as to how this will affect inbreeding. Daetwyler et al. (2007) analyzed the effects of inbreeding due to genomic selection and compared it to sib selection and BLUP selection. Results indicated that, since genomic selection will be able to differentiate between the Mendelian sampling terms of the estimated breeding values, better selection decisions will be made between siblings. This will reduce the between family variance and will result in less coselection of siblings, reducing the overall genetic impact of certain

families. Overall, genome-wide selection will be able to increase the rate of genetic gain while reducing or maintaining the change in inbreeding per generation.

A simulation study was performed by de Roos et al. (2011) analyzing the effects of genetic gain and inbreeding using either young genomically selected animals or progeny tested animals in breeding programs. Results indicated that, given a realistic genomic selection scenario, the rate of genetic gain per year would nearly double compared to conventional progeny testing scenarios. The increases in inbreeding per generation were nearly identical between the two systems, but because the generation interval was drastically shortened by genomic selection, the rate of inbreeding per year increased from 0.18 to 0.52%. As accuracies of the marker prediction increased, and less of the information for breeding values was derived from the parents' breeding values, the rate of inbreeding decreased dramatically with the reduction in the coselection of siblings. The negative consequences of inbreeding, such as the reduction in genetic variance, inbreeding depression in the phenotype, and accumulation of deleterious alleles, are most associated with the rate of inbreeding per generation rather than per year, since processes that compensate for the increase in inbreeding also occur per generation (Villanueva et al., 2000). These results indicate that, as long as marker predictions are relatively accurate, large increases in genetic gain can be made with genomic selection with a similar rate of inbreeding per generation. If marker predictions are less reliable, selection decisions based on SNP markers will reflect family similarities rather than correlations with desirable genes, and greater rates of inbreeding per generation may occur. Although rates of inbreeding in genomic selection may be similar to traditional selection, methods to control inbreeding are still needed.

Sonesson et al. (2012) performed a simulation study to determine how inbreeding should be monitored and controlled in genomic selection programs. The study measured genetic gain,

pedigree inbreeding, genomic inbreeding, and localized genomic inbreeding. When pedigree inbreeding was used as a constraint during genomic selection, genomic inbreeding measures were much higher than expected. But when using genomic inbreeding as a constraint in genomic selection, considerable progress was made in limiting the accumulation of IBD segments of DNA. Results indicate a need for a reliable and efficient measure of genomic inbreeding when performing selection based on genomic marker information.

Calculating inbreeding from genomic information

Several methods have been developed to estimate the level of inbreeding from genomic markers. The current genomic inbreeding index calculated by the USDA-ARS Animal Improvement Programs Laboratory is derived from a genomic relationship (**G**) matrix and is detailed by VanRaden et al. (2011b). The G matrix is utilized in the calculation of routine genetic evaluations for dairy cattle, with the diagonal element of the matrix correlating to the animal's relationship to itself, or its genomic inbreeding index derived from a G matrix ($\mathbf{F_{GRM}}$). The G matrix is calculated using the following formula:

$$G = \frac{ZZ'}{2\sum p(1-p)}$$

where Z, a matrix containing the deviations from base population allele frequency for the specific markers, contains (0 - 2p) for homozygotes, (1 - 2p) for heterozygotes, and (2 - 2p) for the opposite homozygotes, where p is the base population allele frequency. Originally, F_{GRM} values calculated by the USDA-ARS Animal Improvement Programs Laboratory used base population allele frequencies estimated using the algorithm of Gengler et al. (2007), which uses a pedigree relationship matrix and linear mixed model equations to account for selection and drift in allele frequencies across time. This estimation can pose difficulties if pedigree information is unavailable or unreliable. Furthermore, VanRaden et al. (2011b) discovered that using a base

population allele frequency of p = 0.5 for all SNP markers provides a higher correlation between F_{ped} and F_{GRM} (0.59 for Holsteins) when compared to estimating a base population allele frequencies (0.50 for Holsteins). Hence, when providing F_{GRM} with routine genetic evaluations, USDA-ARS Animal Improvement Programs Laboratory now uses a base population allele frequency p = 0.5. VanRaden et al. (2011b) also varied the calculation of F_{GRM} by either regressing or not regressing the G matrix on a pedigree relationship matrix, as described by VanRaden (2008). This regression is performed in order to provide F_{GRM} values with a distribution similar to the more commonly used F_{ped} values. With the regression of the G matrix on the pedigree relationship matrix, VanRaden et al. (2011b) reported F_{GRM} values of 11.0 ± 3.2% for Holsteins, 4.6 ± 4.6% for Jerseys, and 5.4 ± 3.9% for Brown Swiss. As both Jersey and Brown Swiss cattle have been previously shown to be, an average, more inbred than Holstein cattle, these results are somewhat confusing. The differences arise due to the calculation and correction using the genomic relationship matrix, with the values being dependent on the population used to make the calculations. This limits the comparisons that can be made between breeds, and between calculations made a different points in time with a different set of animals.

Another method to estimate genomic inbreeding based on the excess in homozygosity over a set of SNP markers uses the formula:

$$F_E = \frac{O(H_j) - E(H)}{m - E(H)}$$

where $O(H_j)$ is the observed homozygosity over all SNP for individual j, $E(H) = \Sigma\, 1 - 2p_i(1 - p_i)$ is the expected homoyzgosity for all individuals, and p_i is the base population allele frequency for all SNP i = 1, …, m. This estimate is calculated using the –het command in PLINK (Purcell et al., 2007).

Yang et al. (2010) developed a method which is predicted to have a lower error rate and is calculated with the formula:

$$F_{alt} = 1 - \frac{\sum_i \delta_i}{m}$$

where $\delta_i = 1/p_i$ and $1/q_i$ for a homozygote for the minor and major allele, respectively, and 0 for a heterozygote at SNP I, and m is the total number of non-missing markers. The frequency of the major allele is q_i at the ith SNP and $p_i = 1 - q_i$.

Another method to quantify genomic inbreeding is calculation of the proportion of a genome which is contained within a run of homozygosity (ROH). A run of homozygosity is essentially a long stretch of DNA in which all of the makers are homozygous. When DNA is passed from one generation to the next, it is contained within large sections or haplotypes instead of single markers inherited independently. So, when attempting to discover DNA which is IBD, methods that discover long sections of homozygous DNA should be more accurate than methods that ascertain whether each individual marker is IBD.

Howrigan et al. (2011) performed a simulation study to determine which software and program parameters best discovered ROH that were truly IBD. This study compared the three most popular detection programs, PLINK (Purcell et al., 2007), GERMLINE (Gusev et al., 2009), and BEAGLE (Browning and Browning, 2010), and varied detection thresholds within each program. After optimal thresholds were found for each program, PLINK was determined to outperform both GERMLINE and BEAGLE when identifying both recent and ancient inbreeding. When utilizing PLINK, LD pruning was required to remove redundant SNP within SNP-dense regions of the genome. This makes the SNP coverage more uniform in regards to recombination distance. Other parameters included: 1) not allowing any heterozygous SNP within the ROH, and 2) allowing up to 5% missing SNP.

When calculating F_{ped}, changing the year of the base population or changing the number of generations utilized in the calculations can change the estimate of F_{ped}. Varying the minimum length of ROH discovered in a ROH analysis is analogous to changing the base population in an F_{ped} analysis. Fisher (1954) noted that the expected length of a DNA segment which is IBD follows an exponential distribution with a mean equal to $\frac{1}{2g}$ morgans, where g equals the number of generations since the common ancestor. Using an average of 1.25 cM/Mb (Arias et al., 2009), estimates of how recently a common ancestor occurred in an individual's pedigree can be obtained from a ROH analysis.

Once ROH are discovered, an inbreeding index ($\mathbf{F_{ROH}}$) can be calculated using the following formula:

$$F_{ROH} = \frac{\sum_k length(ROH_k)}{L}$$

where k = number of ROH discovered for each individual and L = total length of the genome. The ROH are measured in kilobases, and when analyzing ROH in cattle, L = 2,612,820 kb (Zimin et al., 2009). As each ROH represents a segment of the genome which is IBD, F_{ROH} basically determines in the percentage of the genome which is IBD.

Keller et al. (2011) compared F_{ped} with measures of genomic inbreeding such as F_{PH}, F_{alt}, and F_{ROH}. Results indicated that at effective population sizes similar to those observed in domesticated animal species, especially modern dairy cattle breeds, F_{ROH} had the highest correlation to F_{ped} (0.25), followed by F_{alt} (0.19), and F_{PH} (0.17). When comparing each measure of inbreeding to homozygous mutation load, calculated from the simulated markers with allele frequency of less than 0.5, the causal mechanism underlying inbreeding depression, F_{ROH} had the highest correlation (0.60), followed by F_{alt} (0.53), and F_{PH} (0.45), with F_{Ped} having the lowest correlation (0.25). This study also analyzed each estimate of inbreeding after controlling for its

correlation with F_{ROH}, and this resulted in much lower correlations between homozygous mutation load and F_{alt} (0.15), F_{PH} (0.07), and F_{ped} (0.09). These same reductions in partial correlations were not observed for F_{ROH} when controlling for F_{PH} (0.57), F_{alt} (0.50), or F_{ped} (0.68). These results indicate that F_{ROH} provides additional information about homozygous mutation load that is not provided by any of the other measures, and therefore F_{ROH} should be the most accurate and useful measure of inbreeding.

Quantification of genomic inbreeding and genomic inbreeding depression

Very few studies about genomic inbreeding have been performed in domestic animal species, and those performed have focused on the correlation between genomic and pedigree measures of inbreeding, not the effects of inbreeding on quantitative traits. VanRaden et al. (2011b) reported correlations between F_{ped} and F_{GRM} of 0.59 for Holsteins, 0.68 for Jerseys, and 0.61 for Brown Swiss. This study also noticed stretches of homozygous DNA (ROH) of greater than 1,500 SNP in length. These segments represented very large percentages of the genome; for example the largest chromosome (BTA 1) contained a total of 2,748 markers. Pedigree analysis of these highly inbred individuals indicated that their parents had at least 1 parent or grandparent in common, and one individual had a famous bull represented three times as a great-grandsire of the parents.

Ferencakovic et al. (2011) studied the relationship between F_{ROH} and F_{ped} in Austrian Flechvieh cattle. F_{ROH} values were calculated with varying minimum length of ROH, which would be analogous to varying the year of the base population when calculating F_{ped}. Correlations between the various F_{ROH} values and F_{ped} ranged between 0.61 and 0.68, with the highest correlation between F_{ped} and F_{ROH} with a minimum length of 8,000 kb. Mean F_{ROH}

ranged from 9.0 ± 2.2% when a minimum of 1,000 kb was used in ROH determination, to 1.8 ± 1.3% when a minimum of 16,000 kb was used.

Huang et al. (2012) compared genome wide genetic architecture of closed sub-population of Hereford cattle to that of a general US population of Hereford cattle. This study indicated much higher average LD ($R^2 = 0.36$) between SNP in the closed sub-population when compared to the overall US population ($R^2 = 0.16$). Incidences of extended haplotype homozygosity (ROH) were also greater in the closed sub-population than in the overall US population, suggesting much higher levels of inbreeding.

Although the effects of genomic inbreeding depression in domestic animal species have yet to be analyzed, measures of genomic inbreeding have been utilized recently in human studies to determine possible negative effects due to increases in the percentage of the genome which is IBD. Keller et al. (2012) examined the effects of increases in F_{ROH} as a risk factor for schizophrenia. An increase in F_{ROH} of 1% resulted in an increase in the odds of schizophrenia by about 17%. This increased risk was not due to a few regions of the genome, but rather from an overall increase in the number of IBD segments.

McQuillan et al. (2012) analyzed the effects of increases in F_{ROH} and F_{PH} on human height. Increase in both F_{ROH} and F_{PH} resulted in reductions in height among the 21 different populations sampled, with an average reduction of 3 cm in height for the offspring of first cousins as compared with the offspring of unrelated individuals. Although the negative relationship between genomic inbreeding and height was consistent across populations, the severity of inbreeding depression varied greatly. These results indicate that many rare recessive variants influence human height, and specific variants may or may not be present in different human populations.

The relationship between increases in ROH and early onset Parkinson's disease was examined by Simon-Sanchez et al. (2012). This study indicated that, as the minimum length of ROH increased in the analysis, the ratio between the incidence rate in cases and incidence rate in controls increased as well. This suggests that as genomic inbreeding (length of ROH in an individual) increases, the risk for early onset Parkinson's disease also increases. Once again, these results suggest that a number of rare recessive variants have an effect on disease status.

Reproduction challenges in Holstein cattle

Dairy cattle have experienced large gains in milk production traits due to genetic selection and management improvements over the past several decades. But, along with the increase in milk production, reproductive traits have suffered. Washburn et al. (2002) analyzed reproductive trends in Holstein cows from the southeastern US between 1976 and 1999. Results indicated that average days open increased from 124 ± 0.7 d in 1976 to 168 ± 0.7 d in 1999. The number of services per conception also increased during this time from 1.91 ± 0.02 services in 1976 to 2.94 ± 0.04 in 1996. Days to first service and estrus detection rates also declined during this time period. Norman et al. (2009) analyzed trends in Holstein reproduction from 1996 to 2006. These results indicated an overall decrease in conception rate from 33% in 1996 to 30% in 2006, with a low of 26% occurring in 2001. Number of breedings per lactation increased from 2.1 in 1996 to 2.5 in 2006, while average calving interval increased from 410 to 422 d over the same time frame. These results indicate that, despite efforts to stem the decline, overall reproductive performance of Holstein cattle continues to wane.

Although failure to conceive is often the underlying problem, many studies have also suggested that embryonic death may play a large role. Moreira et al. (2001) reported a loss of 20.7% of pregnancies between 27 and 45 d of gestation in a sample of 139 pregnant Holstein

cows. Cartmill et al. (2001) compared pregnancy rates at days 28 and 38 through day 58 of gestation and discovered a 28% loss in pregnancies in a sample of 110 Holstein cows. In an extensive study, Chebel et al. (2004) compared pregnancy rates at days 31 and 45 of gestation in 1,465 Holstein cows and discovered a rate of pregnancy loss of 12.5%. In regards to inbreeding, embryonic death could be caused by maternal inbreeding, where the increase in inbreeding makes it more difficult for the dam to maintain the pregnancy, or inbreeding of the embryo, where the accumulation of negative recessive alleles leads to an embryo that is not viable and aborts at some point during gestation.

VanRaden et al. (2011a) discovered several haplotypes that may cause a decrease in fertility due to an increase in embryo inbreeding. When analyzing haplotypes from 58,453 Holstein cattle, 3 haplotypes in Holsteins were present in a heterozygous state and one homozygous state, but not in the opposite homozygous state. Phenotypic effects confirmed that these 3 haplotypes caused a reduction in fertility among two heterozygous parents. The effects of each of these mutations and other potential haplotypes on the conception rate of carrier sire mated to a cow with a carrier maternal grandsire ranged from a decrease of 0.9 to 3.2%. An additional deleterious haplotype was found for Brown Swiss and Jersey breeds as well. While a large degree of inbreeding is not necessarily required to obtain two copies of these deleterious haplotype, increases in inbreeding could increase the probability that these, or similar detrimental genes, would be present in the homozygous state.

Fritz et al. (2013) discovered an additional deleterious haplotype which has a carrier frequency of 7.2% of French Holsteins, but was too uncommon in US Holsteins to be confirmed by VanRaden et al. (2011a). A total of 49 homozygotes were expected in the French population, but none were found. The haplotype was then mapped to a 1.4 megabase regions within

chromosome 1, and further analysis discovered the possible point mutation within the GlycinAmide Ribonucleotide Transformylase gene. Fritz et al. (2013) also found the deleterious haplotypes first presented by VanRaden et al. (2011a) and additionally discovered several possible point mutations which may be the underlying cause for this disorder.

Summary

As has been shown, accumulation of pedigree inbreeding is prevalent among domestic cattle. High levels of inbreeding has also been found to be the cause of many negative effects on production, fertility, and longevity, even though some limitations in measuring and reporting pedigree inbreeding may exist. New genotyping technologies have allowed scientist to accurately determine genetic sequences in many species. While this has allowed dairy cattle geneticists to use this information to predict breeding values, it may also be used to predict measures of inbreeding more accurately than can be predicted using pedigrees. The following chapters outline the calculation of various measures of genomic inbreeding, the effects of genomic inbreeding on economically important traits in dairy cattle, a comparison of genomic inbreeding among breeds and genetic groups, prediction of genomic inbreeding from mating pairs and how it may affect fertility.

REFERENCES

Adamec, V., B. G. Cassell, E. P. Smith, and R. E. Pearson. 2006. Effects of inbreeding in the dam on dystocia and stillbirths in US Holsteins. J. Dairy Sci. 89:307–314.

Arias, J.A., M. Keehan, P. Fisher, W. Coppieters, and R. Spelman. 2009. A high density linkage map of the bovine genome. BMC Genetics. 10:18 doi:10.1186/1471-2156/10/18.

Beechinor, J. G., and E. P. Kelly. 1987. Errors of identification amongst cattle presented as progeny of some bulls used in the artificial insemination service in Ireland. Ir. Vet. J. 41:348–353.

Browning, S. R., and B. L. Browning. 2010. High-resolution detection of identity by descent in unrelated individuals. Am. J. Hum. Genet. 86(4):526-539.

Carothers, A. D., I. Rudan, I. Kolcic, O. Polasek, C. Hayward, A. F. Wright, H. Campbell, P. Teague, N. D. Hastie, and J. L. Weber. 2006. Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. Ann. Hum. Genet. 70:666–676.

Carrillo, J. A., and F Siewerdt. 2010. Consequences of long-term inbreeding on preweaning traits in a closed nucleus Angus herd. J. Anim. Sci. 88:87-95.

Cartmill, J. A., S. Z. El-Zarkouny, B. A. Hensley, G. C. Lamb, and J. S. Stevenson. 2001. Stage of cycle, incidence, and timing of ovulation, and pregnancy rates in dairy cattle after three timed breeding protocols. J. Dairy Sci. 84:1051–1059.

Chebel, R. C., J. E. P. Santos, J. P. Reynolds, R. L. A. Cerri, S. O. Juchem, and M. Overton. 2004. Factors affecting conception rate after artificial insemination and pregnancy loss in lactating dairy cows. Anim. Reprod. Sci. 84:239-255.

Christensen, L. G., P. Madsen, and J. Petersen. 1982. The influence of incorrect sire-identification on the estimates of genetic parameters and breeding values. Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain 7:200–208.

Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loor, A. Evertsvan der Wind, J. H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res. 15:936–944.

Cole, J. B., P. M. VanRaden, J. R. O'Connell, C. P. Van Tassell, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and G. R. Wiggans. 2009. Distribution and location of genetic effects for dairy traits. J. Dairy Sci. 92:2931–2946.

Croquet, C., P. Mayeres, A. Gillon, H. Hammami, H. Soyeurt, S. Vanderick, and N. Gengler. 2007. Linear and curvilinear effects of inbreeding on production traits for Walloon Holstein cows. J. Dairy Sci. 90:465–471.

Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124:369–376.

Dechow, C. D., H. D. Norman, N. R. Zwald, C. M. Cowan, and O. M. Meland. 2008. Relationship Between Individual Herd-Heritability Estimates and Sire Misidentification Rate. J. Dairy Sci. 91:1640-1647.

de Roos, A. P. W., C. Schrooten, R. F. Veerkamp, and J. A. M. van Arendonk. 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J. Dairy Sci. 94:1559–1567.

Faria, F. J. C., A. E. V. Filho, F. E. Madelena, and L. A. Josahkian. 2009. Pedigree analysis in the Brazilian Zebu breeds. J. Anim. Breed. Genet. 126:148-153.

Ferencakovic, M., E. Hamzic, B. Gredler, I. Curik, and J. Sölkner. 2011. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. Agriculturae Conspectus Scientifcus 76:325–328.

Fisher, R.A. 1954. A fuller theory of “junctions” in inbreeding. Heredity. 8:187-1973.

Fritz S, Capitan A, Djari A, Rodriguez SC, Barbat A, et al. 2013. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. PLoS ONE 8(6): e65550.

Gelderman, H., U. Pieper, and W. E. Weber. 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. J. Anim. Sci. 63:1759–1768.

Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate the gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal 1:21-28.

Grisart, B., F. Farnir, L. Karim, N. Cambisano, J.-J. Kim, A. Kvasz, M. Mni, P. Simon, J.-M. Frère, W. Coppieters, and M. Georges. 2004. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. Proc. Natl. Acad. Sci. USA 101:2398–2403.

Gusev, A., J. K. Lowe, M. Stoffel, M. J. Daly, D. Altshuler, J. L. Breslow, J. M. Friedman, I. Pe'er. 2009. Whole population, genome-wide mapping of hidden relatedness. Genome Res. 19(2):318-326.

Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in the prediction of breeding values. Biometrics. 32:69-83.

Howrigan, D.P., M.A. Simonson, and M.C. Keller. 2011. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics. 12:460.

Huang, Y., C. Maltecca, M. D. MacNeil, L. J. Alexander, W. M. Snelling, and J. P. Cassidy. 2012. Using 50 K single nucleotide polymorphisms to elucidate genomic architecture of line 1 Hereford cattle. Front. Genet. 3:285. doi: 10.3389/fgene.2012.00285.

Keller, M. C., M. A. Simonson, S. Ripke, B. M. Neale, P. V. Gejman, et al. 2012. Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. PLoS Genet. 8(4): e1002656. doi:10.1371/journal.pgen.1002656.

Keller, M.C., P.M. Visscher, and M.E. Goddard. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189:237-249.

McQuillan, R., N. Eklund, N. Pirastu, M. Kuningas, B. P. McEvoy, et al. 2012. Evidence of Inbreeding Depression on Human Height. PLoS Genet. 8(7): e1002655. doi:10.1371/journal.pgen.1002655.

Mc Parland, S., J. F. Kearney, M. Rath, and D. P. Berry. 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. J. Dairy Sci. 90:4411–4419.

Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Miglior, F., E. B. Burnside, and B. W. Kennedy. 1995. Production traits of Holstein cattle: estimation of nonadditive genetic variance components and inbreeding depression. J. Dairy Sci. 78:1174–1180.

Moreira, F., C. Orlandi, C. A. Risco, R. Mattos, F. Lopes, and W. W. Thatcher. 2001. Effects of presynchronization and bovine somatotropin on pregnancy rates to a timed artificial insemination protocol in lactating dairy cows. J. Dairy Sci. 84:1646–1659.

Norman, H. D., J. R. Wright, S. M. Hubbard, R. H. Miller, and J. L. Hutchinson. 2009. Reproductive status of Holstein and Jersey cows in the United States. J. Dairy Sci. 92:3517-3528.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, et al. 2007. PLINK: a toolset for whole genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.

Ron, M., Y. Blanc, M. Band, E. Ezra, and J. L. Weller. 1996. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. J. Dairy Sci. 79:676–681.

Santana Jr., M. L.,P. S. Oliveira, V. B. Pedrosa, J. P. Eler, E. Groeneveld, J. B. S. Ferraz. 2010. Effect of inbreeding on growth and reproductive traits of Nellore cattle in Brazil. Livestock Sci. 131:2-3:212-217.

Simón-Sánchez, J., L. L. Kilarski, M. A. Nalls, M. Martinez, C. Schulte, and P. Holmans., International Parkinson's Disease Genomics Consortium, Wellcome Trust Case Control Consortium, T. Gasser, J. Hardy, A. B. Singleton, N. W. Wood, A. Brice, P. Heutink, N. Williams, and H. R. Morris. 2012. Cooperative genome-wide analysis shows increased homozygosity in early onset Parkinson's disease. PLoS ONE 7:e28787 http://dx.doi.org/10.1371/journal.pone.0028787.

Smith, L. A., B. G. Cassell, and R. E. Pearson. 1998. The effects of inbreeding on the lifetime performance of dairy cattle. J. Dairy Sci. 81:2729–2737.

Sonesson, A. K., J. A. Woolliams, and T. H. E. Meuwissen. 2012. Genomic selection requires genomic control of inbreeding. Genet. Sel. Evol. 44:27.

Thompson, J. R., R. W. Everett, and N. L. Hammerschmidt. 2000. Effects of inbreeding on production and survival in Holsteins. J. Dairy Sci. 83:1856–1864.

US Department of Agriculture-Animal Improvement Laboratories. 2012 Bovine Inbreeding Trends. Accessed Aug. 30, 2012. http://aipl.arsusda.gov/eval/summary/inbrd.cfm.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P. M., K. M. Olson, D. J. Null, and J. L. Hutchison. 2011a. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94:6153–6161.

VanRaden, P. M., K. M. Olson, G. R. Wiggans, J. B. Cole, and M. E. Tooker. 2011b. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 94:5673–5682.

Villanueva, B., P. Bijma, and J. A. Wooliams, 2000. Optimal selection policies for schemes with overlapping generations and restricted inbreeding. Genet. Sel. Evol. 32:339-355.

Visscher, P. M., J. A. Woolliams, D. Smith, and J. L. Williams. 2002. Estimation of Pedigree Errors in the UK Dairy Population using Microsatellite Markers and the Impact on Selection. J. Dairy Sci. 85:2368-2375.

Washburn, S. P., W. J. Silvia, C. H. Brown, B. T. McDaniel, and A. J. McAllister. 2002. Trends in reproductive performance in Southeastern Holstein and Jersey DHI Herds. J. Dairy Sci. 85:244-251.

Weigel, K.A. 2001. Controlling inbreeding in modern breeding programs. J. Dairy Sci. 84:E177-E184.

Weigel, K. A., P. C. Hoffman, W. Herring, T. J. Lawlor Jr. 2012. Potential gains in lifetime net merit from genomic testing of cows, heifers, and calves on commercial dairy farms. J. Dairy Sci. 95:2215-2225.

Wiggans, G. R., and P. M. VanRaden. 1995. Calculation and use of inbreeding coefficients for genetic evaluation of United States dairy cattle. J. Dairy Sci. 78:1584-1590.

Wiggans, G. R., P. M. VanRaden, and T. A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, and future. J. Dairy Sci. 94:3202–3211.

Woodward, T. E., R. R. Graves. 1946. Results of Inbreeding Grade Holstein-Friesian Cattle. US Dept. of Ag. Tech. Bulletin #27.

Wright, S. 1922. Coefficients of inbreeding and relationship. Amer. Natur. 56:330-338.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders et al. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565-569.

Zimin A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz., et al. 2009. A whole-genome assembly of the domestic cow, Bos Taurus. Genome Biol. 10:R42.

CHAPTER 3

Evaluation of inbreeding depression in Holstein cattle using whole genome SNP markers and alternative measures of genomic inbreeding

ABSTRACT

The effects of increased pedigree inbreeding in dairy cattle populations have been well documented and result in a negative impact on profitability. Recent advances in genotyping technology have allowed researchers to move beyond pedigree analysis and study inbreeding at a molecular level. In this study, 5,853 animals were genotyped for 54,001 single nucleotide polymorphisms (SNP); 2,913 cows had phenotypic records including a single lactation for milk yield (from either lactation 1, 2, 3, or 4), reproductive performance, and linear type conformation. After removing SNPs with poor call rates, low minor allele frequencies, and departure from Hardy-Weinberg equilibrium, 33,025 SNPs remained for analyses. Three measures of genomic inbreeding were evaluated, percent homozygosity (**F_{PH}**), inbreeding calculated from runs of homozygosity (**F_{ROH}**), and inbreeding derived from a genomic relationship matrix (**F_{GRM}**). Average F_{PH} was 60.5 ± 1.1%, average F_{ROH} was 3.8 ± 2.1%, and average F_{GRM} was 20.8 ± 2.3%, where animals with larger values for each of the genomic inbreeding indices were considered more inbred. Decreases in total milk yield to 205 d postpartum of 53, 20, and 47 kg per 1% increase in F_{PH}, F_{ROH}, and F_{GRM}, respectively, were observed. Increases in days open per 1% increase in F_{PH} (1.76 d), F_{ROH} (1.72 d), and F_{GRM} (1.06 d) were also noted, as well as increases in maternal calving difficulty (0.09, 0.03, and 0.04 on a 5-point scale for F_{PH}, F_{ROH}, and F_{GRM}, respectively). Several linear type traits, such as strength (-0.40, -0.11, -0.19), rear legs rear view (-0.35, -0.16, -0.14), front teat placement (0.35, 0.25, 0.18), and teat length (-0.24, -0.14, -0.13) were also affected by increases in F_{PH}, F_{ROH}, and F_{GRM}, respectively. Overall, increases in each measure of genomic inbreeding in this study were associated with negative effects on production and reproductive ability in dairy cows.

INTRODUCTION

Inbreeding in US dairy cattle has increased steadily over the past several decades. The average inbreeding coefficient of Holsteins rose from 0.4% in 1970 to 5.8% in 2012 (USDA-AIPL, 2012), with the majority of this increase attributed to intense selection pressure on bulls. Some Holstein bulls used in artificial insemination have had as many as 250,000 milking daughters or 5,000 progeny tested sons (Weigel, 2001). In addition, of the roughly 5,000 young bulls that were progeny-tested each year globally at that time, nearly 50% were offspring of the 10 most popular sires. Even with extensive pedigree records, avoidance of inbreeding is increasingly difficult.

The negative effects of inbreeding have been well documented and tend to fall into two categories. The first is an increased prevalence of rare lethal or harmful recessive disorders, such as BLAD (Kehrli et al., 1990) or DUMPS (Shanks et al., 1984), when closely related individuals are mated. The second is an overall decrease in functionality, performance, and profitability of inbred animals. Many studies have concluded that increased pedigree inbreeding in dairy cattle is associated with decreases in production (Miglior et al., 1995; Smith et al., 1998; Thompson et al., 2000), reproductive ability (Smith et al., 1998; Mc Parland et al., 2007), and survivability (Thompson et al., 2000; Mc Parland et al., 2007). Smith et al. (1998) indicated that for each 1% increase in inbreeding, lifetime total milk yield decreased by 94.5 kg, lifetime total fat yield decreased by 3.3 kg, and lifetime total protein yield decreased by 2.9 kg. Effects of inbreeding on reproductive traits and survival were measured by McParland et al. (2007), and results indicated a 0.7 d increase in calving interval and a 0.3% decrease in survival to second lactation per 1% increase in inbreeding. Furthermore, Adamec et al. (2006) noted an increase in probability of maternal dystocia (0.42 and 0.30% for male and female calves, respectively) and stillbirths (0.25 and 0.20% for male and female calves, respectively) per 1% increase in inbreeding.

With the recent development of high-throughput genomic tools, such as the Illumina Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA), many new questions have arisen regarding inbreeding. Results from a simulation study by de Roos et al. (2011) noted that, with the reduction in generation interval from the use of genomic selection, the rate of genetic gain per generation could double at a given rate of inbreeding per generation. At the same time, however, the rate of inbreeding per year will increase, due to the reduction in generation interval. Daetwyler et al. (2007) noted that genomic selection will be able to predict the Mendelian sampling between full sibs without progeny testing. This will reduce the incidence of co-selection of siblings, as well as the overall genetic impact of single influential animals on the population, which may lead to slower accumulation of inbreeding.

Previous studies have considered only pedigree based estimates of inbreeding, but with the availability of whole genome marker panels the next logical step is to quantify inbreeding genomically. The inbreeding coefficient is defined as the probability that a pair of alleles are identical-by-descent (**IBD**). Historically, geneticists have estimated this probability using pedigree information. Utilizing genomic information should lead to a more accurate depiction of inbreeding. For example, consider an organism whose parents are first cousins (with no other previous common ancestors). The pedigree inbreeding coefficient ($\mathbf{F_{ped}}$) would be 6.25%; on average 6.25% of this organism's genome would be identical, having originated from either of the common great-grandparents. Carothers et al. (2006) noted that this value varies greatly due to Mendelian sampling, with a standard deviation of 2.43%. This deviation depends on the recombination events that occurred during gamete formation in the parents, as well as the chance meeting of the successful gametes. While this deviation is present when estimating inbreeding

from pedigrees, genomic estimates of inbreeding should determine the actual product of the recombination events and provide a more accurate estimation.

In the dairy industry, genomic inbreeding coefficients of genotyped animals are currently calculated from a SNP-derived genomic relationship matrix (**F_{GRM}**). VanRaden et al. (2011a) compared F_{GRM} values to F_{ped} values, and reported correlations of 0.59, 0.68, and 0.61 for Holstein, Jersey, and Brown Swiss, respectively. Slightly higher correlations of 0.69 were obtained by Hayes and Goddard (2008) in a population of Australian Angus bulls. To date, no studies have determined the effects of F_{GRM} on lactation yield, fertility, or survivability in dairy cattle.

Increased levels of inbreeding would appear genomically as an increase in the frequency of homozygous alleles. One simple method to determine inbreeding genomically would be to look at the percentage of alleles that are homozygous. A problem with this method is that alleles that are IBD and identical-by-state (**IBS**) cannot be distinguished and are both included in this measure of inbreeding. An alternative method involving genomic runs of homozygosity (**ROH**) attempts to distinguish these differences and has been utilized in human genomic studies for nearly a decade, examining population history (Kirin et al., 2010; Li et al., 2008) and the effects of inbreeding on disease risk (Simon-Sanchez et al., 2012). Keller et al. (2011) indicated that inbreeding estimates using ROH (**F_{ROH}**) are preferable to F_{ped} and other measures of genomic inbreeding, because it correlates strongly with homozygous mutation load. More specifically, at an effective population size similar to that of the Holstein cattle population, F_{ROH} has a correlation of about 0.6 to the homozygous mutation load, while F_{Ped} only had a correlation of about 0.25. Very few studies involving ROH have been performed in cattle, but Ferencakovic et

al. (2011) noted positive correlations (0.61 to 0.68) between varying measures of F_{ROH} and F_{Ped} in a population of 500 Simmental bulls.

The current study aims to quantify various measures of genomic inbreeding and determine their associations with economically important traits in dairy cattle.

MATERIALS AND METHODS

Data

A total of 5,853 animals were genotyped for 54,001 single nucleotide polymorphism (SNP) markers. After editing SNPs for minor allele frequency (MAF < 0.05), call rate (percent missing > 0.1), and Hardy-Weinberg equilibrium ($p < 0.0001$), 33,025 SNPs remained for analysis. Animals with more than 10% missing SNPs were also removed from the analysis. Phenotypes were available for 2,913 cows in 9 herds from various geographical regions of the US. Data compiled for each cow consisted of one lactation record through 205 d from either first, second, third, or fourth parity, with data from 854 cows in first lactation, 1,088 cows in second lactation, 592 cows in third lactation, and 379 cows in fourth lactation. All cows were born between July 1999 and December 2005. Lactations included in this study were initiated between June 2006 and March 2007.

Daily milk yields were collected from all animals from parturition through 205 days in milk (**DIM**). From this, total milk yield to 205 DIM, average daily milk yield, and peak milk yield values were derived. Fat percentage, protein percentage, somatic cell score (**SCS**), and milk urea nitrogen (**MUN**) were recorded at 60 d intervals throughout the lactation and averaged over the lactation prior to analysis.

Days open, conception rate, DIM at first breeding, calf birth weight, and calving ease were derived from on-farm reproductive data. Cows were required to be at least 250 d

postpartum if they had not been confirmed pregnant, and all cows with greater than 250 d open were set to 250 d, following the process of VanRaden et al. (2004) for routine genetic evaluation of daughter pregnancy rate. Conception rate for each cow was defined as 1 divided by the number of times bred (if confirmed pregnant), and zero otherwise. For periods of estrus in which more than one breeding occurred (e.g., 2 breedings within 3 d), only one breeding was counted towards the number of times bred. Calving ease was recorded on an ordinal scale from 1 (no assistance) to 5 (extremely difficult birth).

At an average of 97.3 ± 44.7 DIM, all cows were scored for linear type traits, including stature, strength, body depth, dairy form, rump angle, rump width, rear legs side view, rear legs rear view, foot angle, fore udder attachment, udder height, udder width, udder cleft, udder depth, front teat placement, rear teat placement, teat length, and udder tilt. All linear type traits were scored on a 50 point scale by trained evaluators.

Genomic Inbreeding Coefficients

The first measure of genomic inbreeding considered was the percent homozygosity (**F_{PH}**) of all SNPs. F_{PH} was derived with the formula:

$$F_{PH} = \frac{N_{AA} + N_{BB}}{N_{AA} + N_{AB} + N_{BB}}$$

where N_{AA}, N_{AB}, and N_{BB} refer to the number of SNPs that are classified as AA, AB, and BB respectively.

While the F_{PH} of an animal can provide some indication of its level of inbreeding, it does not distinguish between markers that are IBS and those that are IBD. One possible method to alleviate this problem is to consider genomic ROH. A ROH is defined as a specific number of consecutive SNPs which are all homozygous. Inbreeding increases overall homozygosity in an individual, but this increase does not simply present itself as single randomly dispersed

homozygous SNPs, but rather as long runs of homozygous SNPs that were inherited together. Furthermore, the length of the ROH correlates to the distance within the pedigree until the common ancestor is observed. Longer ROH indicate more recent common ancestors, because recombination has had fewer generations to break up the segments, whereas shorter ROH are indicative of common ancestors further back in the pedigree. If the minimum length of ROH is increased, the results would be focused on more recent inbreeding. Fisher (1954) noted that the expected length of the DNA segment which is IBD follows an exponential distribution with mean equal to $\frac{1}{2g}$ Morgans, where g equals the number of generations since the common ancestor. Common ancestors occurring 10 generations back would have an average ROH length of 0.05 Morgans, or 5 cM. At an average of 1.25 cM/Mb (Arias et al., 2009), converting the minimum length of ROH discovered in this study (about 4 Mb) to cM results in an average ROH length of 5 cM. In practice, discovery of ROH is slightly more complex and was determined using PLINK (Purcell et al., 2007). Figure 1 describes the manner by which ROH were discovered in PLINK using a sliding window of SNP along the chromosome. First, paternal and maternal chromosomal segments are presented. A sliding window of 10 SNP then moves along the chromosome one SNP at a time. This determines whether every SNP inside this window is homozygous. The number of completely homozygous windows, as well as the total number of windows, is summed for each SNP. If, at minimum, 10 consecutive SNP are determined to have greater than 5% of these windows homozygous, a ROH is called. Recommendations for many of the input parameters for ROH discovery were derived from Howrigan et al. (2011). No heterozygous SNPs and 1 missing SNP were allowed within the sliding window. The minimum ROH length of 30 SNP (compared to 10 SNP in Figure 1) was used in order to capture inbreeding occurring in about the previous 10 generations in this study (Fisher, 1954). This

essentially means, in comparison to pedigree inbreeding, that F_{ROH} is calculated with 10 generations of complete pedigrees, or the base population used for determining F_{ROH} is 10 generations back. Since high linkage disequilibrium (**LD**) within given sections of DNA can lead to detection of ROH that are not truly IBD, LD pruning was also performed on the SNP set prior to the ROH determination to increase power, as suggested by Purcell et al. (2007). LD pruning was performed using PLINK, and SNPs which had an $r^2 > 0.5$ with all other SNPs in a 50 SNP window were removed. This resulted in a total of 7,997 SNPs being used for the ROH analysis.

The results of the ROH discovery were utilized to create an inbreeding coefficient for each animal, denoted as F_{ROH}. F_{ROH} was calculated by the formula:

$$F_{ROH} = \frac{\sum_k length(ROH_k)}{L}$$

where k = number of ROH discovered for each animal, and L = total length of the genome. Length of ROH was measured in kilobases (kb) with L = 2,612,820 kb (Zimin et al., 2009).

Measures of inbreeding from a genomic relationship matrix, G, were denoted as F_{GRM}, and were calculated using the method and programs described by VanRaden et al. (2011a). This is the method utilized by the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) for routine genomic evaluation of US dairy cattle, as well as calculation of published genomic inbreeding values. The G matrix was calculated using the following formula:

$$\boldsymbol{G} = \frac{\boldsymbol{ZZ'}}{2\sum p(1-p)}$$

where Z contains the values: 0 - 2p for homozygotes, 1 - 2p for heterozygotes, and 2 - 2p for opposite homozygotes. The F_{GRM} used in the depression analyses was calculated utilizing a p = 0.5, which is the current method utilized by USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) for their presentation of genomic inbreeding values. The matrix Z

then contains values of 1 or -1 for homozygotes and 0 for heterozygotes. This essentially makes F_{GRM} a measure of homozygosity that has been transformed to follow a distribution similar to traditional F_{Ped}. The values on the diagonal of **G** denote the relationship of the animal to itself, or its genomic inbreeding coefficient. A second genomic inbreeding index (**$F_{GRM\text{-}BP}$**) was calculated by estimating the allele frequencies (p) in the base population with the algorithm of Gengler et al. (2007). This method utilizes the very limited pedigree information available as well as linear mixed model equations to provide an estimate of the selection and drift of allele frequencies. This method had been previously used by the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) to report genomic inbreeding (VanRaden et al., 2011a), and will only be used to compare against F_{PH}, F_{ROH}, and F_{GRM} in the present study.

Statistical Analysis

All statistical analyses were performed using the MIXED procedure in SAS software version 9.1 (SAS Institute, Cary, NC), and all phenotypic traits were analyzed using linear models that regressed the trait by the measures of inbreeding. All traits were adjusted for effects of herd-year-season and parity. Because some cows were missing up to 15% of daily milk records, the percentage of days with missing records was included in the model for 205 d total milk yield and average daily milk yield. The DIM at peak milk was included as a covariate when analyzing peak milk, and DIM at time of evaluation was included as a covariate when analyzing linear type traits. The type of birth (single/twins) and sex of calf were included as covariates when analyzing body weight of the calf and calving ease. Linear regression coefficients corresponding to the change in each trait per 1% increase in genomic inbreeding, as well as significance tests, were derived from these models.

RESULTS

Figure 2 displays the distributions of F_{PH}, F_{ROH}, and F_{GRM} respectively, with means of 60.5 ± 1.1%, 3.8 ± 2.1%, and 20.8 ± 2.3%. For each distribution, animals with smaller genomic inbreeding values are considered as the least inbred animals in the population, whereas animals with larger genomic inbreeding values are considered as the most inbred. Correlations between the three measures of genomic inbreeding were large, with correlations between F_{PH} and F_{ROH} of 0.81, F_{PH} and F_{GRM} of 0.99, and F_{ROH} and F_{GRM} of 0.81. Furthermore, $F_{GRM\text{-}BP}$ had more modest correlations to F_{PH} (0.77), F_{ROH} (0.55), and F_{GRM} (0.78).

Estimates of inbreeding depression for production traits are presented in Table 1. Increases in 1% of F_{PH}, F_{ROH}, or F_{GRM} resulted in decreases in 205 d milk yield of 53, 20, or 47 kg, respectively. Average daily milk yield also exhibited a decrease due to a 1% increase in F_{PH} (0.28 kg/d), F_{ROH} (0.11 kg/d), or F_{GRM} (0.25 kg/d). Furthermore, a slight decrease in MUN was observed when F_{PH} (0.06 mg/dL) or F_{GRM} (0.03 mg/dL) increased by 1%. Peak milk, average fat percentage, average protein percentage, and average SCS were not affected by changes in F_{PH}, F_{ROH}, or F_{GRM}.

Table 2 displays the estimates of inbreeding depression for reproductive traits. A 1% increase in F_{PH}, F_{ROH}, or F_{GRM} resulted in an increase in days open of 1.76, 1.72, or 1.06 d, respectively. Conception rate also decreased with a 1% increase in F_{ROH} (-0.82%) or F_{GRM} (-0.53%), whereas increases in F_{PH} had no effect. DIM at first breeding was not affected by changes in any of the genomic inbreeding measures. Increases in F_{PH} or F_{GRM} resulted in decreases in the BW of the calves born from these cows of 0.4 or 0.2 kg/1% increase, respectively. Furthermore, calving ease scores (measured on a 5-point scale) increased per 1% increase in F_{PH} (0.09), F_{ROH} (0.03), or F_{GRM} (0.04).

Estimates of inbreeding depression for linear type traits are presented in Table 3. Stature, dairy form, rump width, rear legs side view, foot angle, fore udder attachment, udder height, udder width, and udder cleft did not change with an increase in any measure of genomic inbreeding. With each 1% increase in F_{PH}, F_{ROH}, or F_{GRM}, strength decreased (-0.40, -0.11, -0.19), rear legs rear view tended towards closer hocks (-0.35, -0.16, -0.14), front teats were closer together (0.35, 0.25, 0.18), and teat length was shorter (-0.24, -0.14, -0.13). Furthermore, an increase F_{PH} resulted in more shallow body depth (-0.25) and a greater forward tilt to the udder (0.24). An increase in F_{ROH} resulted in higher udders (0.14), more forward placement of rear teats (0.25), and greater forward tilt to the udder (0.15). Increases in F_{GRM} resulted in a more shallow body depth (-0.14) and higher pins (-0.14). All values denoted are the estimated change in the linear type trait (measured on a 50-point scale) per 1% increase in the corresponding measure of genomic inbreeding.

In order to more accurately depict the expected differences in performance associated with changes in each of the three measures of genomic inbreeding, differences between cows with small and large genomic inbreeding coefficients were compared in Table 4. Predicted phenotypes are shown for cows with genomic inbreeding coefficients 2 standard deviations above or below the mean of the corresponding genomic inbreeding measurement (F_{PH}, F_{ROH}, or F_{GRM}). Because the F_{ROH} distribution is slightly skewed to the right, and the mean minus twice the standard deviation results in a negative number, 0 was used as the lower bound for F_{ROH}. Phenotypic values shown in Table 4 were calculated from the phenotypic mean and the estimated regression coefficient for the corresponding trait and genomic inbreeding measure. Cows with high (plus 2 standard deviations) inbreeding coefficients produced less total milk to 205 d (-242, -161, -438 kg) and had a lower average daily milk yield (-1.28, -0.89, -2.33 kg) for F_{PH}, F_{ROH},

and F_{GRM}, respectively than cows with low (minus 2 standard deviations) inbreeding coefficients. Cows with high values for F_{PH} and F_{GRM} also had lower average MUN levels (-0.3 and -0.3 mg/dL, respectively). An increase in 8, 14, and 10 days open was noted between cows with high and low F_{PH}, F_{ROH}, and F_{GRM}, respectively, whereas a decrease in conception rate of 6.6 and 4.9% was noted between cows with high and low F_{ROH} and F_{GRM} values, respectively. Cows with high F_{PH} and F_{GRM} tended to have calves that were lighter (-1.8 and -1.9 kg, respectively) than cows with low values. With a linear regression analysis, an increase of 0.41, 0.24, and 0.37 in average maternal calving ease scores was observed between cows with low and high F_{PH}, F_{ROH}, and F_{GRM}, respectively. Most linear type traits exhibited a difference of between 1 and 2 points when comparing cows with low and high genomic inbreeding coefficients. The largest difference (2.0) was observed for front and rear teat placement when the comparison was made with F_{ROH}, while the smallest difference of -0.9 was noted for the association between F_{ROH} and strength.

DISCUSSION

Because pedigree-based inbreeding coefficients were unavailable for animals in this study, comparisons must be based on previous studies. Comparison of genomic and pedigree inbreeding depression from the same animals would be preferred, but lack of reliable pedigree data is a common occurrence on most commercial dairy operations, as was the case with those used in the present study.

As expected, the correlation between F_{PH} and F_{GRM} was extremely high (0.99), as utilizing a p=0.5 for an allele frequency in F_{GRM} is essentially a measure of homozygosity which has been adjusted to conform to a distribution similar to pedigree inbreeding. Utilizing the results from Keller et al. (2011) in which F_{ROH} was determined as the optimal method of genomic inbreeding, the correlations between F_{ROH} and F_{GRM} (0.81) and the correlations between F_{ROH}

and F_{GRM-BP} (0.55) would suggest that utilizing a uniform base population allele frequency (p=0.5) may be more beneficial than attempting to estimate a base population. These results are similar to VanRaden et al. (2011a), in which utilizing allele frequencies of 0.5 resulted in higher correlations with F_{Ped}. The results of the present study further demonstrate that the current method of using allele frequencies of 0.5 the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) is preferred to a estimating a base population.

Results from the present study are consistent with previous studies involving F_{Ped}. Decreases in total milk yield (Smith et al., 1998; Thompson et al., 2000; McParland et al., 2007), as well as decreases in overall reproductive ability (Smith et al., 1998; McParland et al., 2007), as inbreeding increases, have been noted previously. Furthermore, Adamec et al. (2006) had also noted an increase in dystocia as inbreeding increased. Smith et al. (1998) provided estimates of expected differences in milk yield and calving interval between cows with 0% inbreeding and those with 12.5% inbreeding, which would correspond to a cow produced from a half-sib mating. At this inbreeding level, first lactation mature equivalent milk yield would decrease by 464 kg. The genomic inbreeding measures of F_{PH} (-242 kg) and F_{ROH} (-161 kg) in Table 4 indicate less overall milk yield depression than Smith et al. (1998), but the estimate for F_{GRM} in Table 4 is similar (-438 kg). Note that values presented in Table 4 span 4 standard deviations, whereas comparing F_{Ped} of 0 and 12.5% spans a range of 5 standard deviations. Also, the current study considered only milk yield to 205 d postpartum, while Smith et al. (1998) considered milk yield to 305 d. Much larger realized depression was observed in reproductive ability in this study than Smith et al. (1998). When increasing F_{Ped} from 0 to 12.5%, an increase in calving interval of 3.3 d was predicted, but for the range of 4 standard deviations in F_{PH}, F_{ROH}, and F_{GRM} shown in Table 4, increases of 8, 14, and 10 days open were predicted.

Several linear type traits in the current study and Smith et al. (1998) shared significance (strength, body depth, udder depth, and front teat placement). For both studies, greater levels of inbreeding resulted in a narrower chest with less body depth, as well as a higher udder with closer front teats. Croquet et al. (2006) presented similar results, noting that more inbred cows tended to have narrower, smaller frame size with less body depth. McParland et al. (2007) examined the same traits in Irish Holstein cows and reported opposite results with respect to strength and body depth, although they noted that this unexpected result may have been due to linebreeding or directional selection for these traits. The association with shallower udder depth in this study and others (Smith et al., 1998; Mc Parland et al., 2007) may reflect the fact that highly inbred animals tend to produce less milk, which may result in less volume and depth of udder.

As shown in Figure 2, the three measures of genomic inbreeding considered in this study have very different probability density functions, means, and standard deviations. Furthermore, the number of SNPs, size of the SNP chip, and selection criteria for the SNPs used to determine the genomic inbreeding values can have a huge impact on these values. For example, the selection of SNPs for the Illumina Bovine3K BeadChip focused on SNPs that were more polymorphic than SNPs on the Bovine SNP50 BeadChip (Illumina Inc., 2011). An index for F_{PH} from the 3K chip would provide lower inbreeding values, simply due to pre-selection that has occurred among the SNPs. Although each of the genomic inbreeding indexes in this study were associated with inbreeding depression, values presented to dairy producers should be consistent if genomic inbreeding is to be used effectively in selection decisions.

Utilizing ROH may provide the most effective, consistent, and easily understood genomic inbreeding values. As shown in Figure 2B, the distribution of F_{ROH} values is very

similar to what is normally seen with pedigree inbreeding. Changes in the number of SNPs in a SNP panel can be accommodated easily by simply changing the minimum number of SNPs in ROH determination. For example, when utilizing 3K genomic data, a minimum length of 15 SNPs may correlate closely to inbreeding that occurred in the previous 10 generations, as was the case for the minimum length of 30 SNPs used in this study. Furthermore, the basic definition of F_{ROH}, the percentage of the genome that is IBD, is the definition of pedigree inbreeding as well. Determination of whether a homozygous SNP is IBD or IBS is important when examining genomic inbreeding, and utilizing ROH is the most effective method presented herein to distinguish between IBD and IBS. In pedigree inbreeding, the determination of whether an allele is IBD is in reference to a base population. As has been mentioned, the estimation of a base population for use in a genomic relationship matrix is a difficult problem, and many of the methods may not proved a better estimation than simply using an allele frequency of 0.5 (VanRaden et al., 2011a). This same difficulty would occur when attempting to correct F_{PH} to a base population. This is alleviated with the use of runs of homozygosity, though. Varying the minimum length of ROH discovered is analogous to changing the base population in pedigree inbreeding. A shorter minimum ROH would provide more ancient inbreeding (a base population occurring many generations previously) while a longer minimum ROH would only include more recent inbreeding (a base population of just several generation previously). Also, Keller et al. (2011) determined that F_{ROH} values correlate much higher to homozygous mutation load (0.6) than another measure of genomic inbreeding which would be analogous to F_{PH} and F_{GRM} (0.45). All measures of genomic inbreeding presented had a higher correlation with homozygous mutation load than did F_{Ped} (0.25). Combining all of these aspects suggests that F_{ROH} would be the most effective and easily understood method of genomic inbreeding presented in this study.

Several challenges may occur when utilizing genomic inbreeding to predict inbreeding depression. One is that the traits analyzed to determine inbreeding depression are also traits that are under directional selection in the population, meaning that increased homozygosity at some loci, or presence of some ROH, may actually be beneficial. Results of this study indicate, however, that overall increases in homozygosity (genomic inbreeding) are associated with decreased functionality and productivity. Much as a tradeoff between pedigree inbreeding and selection intensity existed in traditional selection, a tradeoff between genomic inbreeding and selection for homozygosity of favorable alleles may exist in genomic selection. Sonesson et al. (2012) noted that, when calculating genomic breeding values, a correction based on genomic inbreeding (instead of traditional pedigree inbreeding) is required. Cole and VanRaden (2010) previously demonstrated the selection of a "supercow," which would have the 30 best possible chromosomes. This cow would have a PTA for lifetime net merit of +$3,148, which is about 3.5 times greater than the highest living animal at that time ($911). Most likely, this "supercow" would be homozygous at large proportions of its genome, and although its breeding value would be superior, results of this study suggest that the actual production and reproductive ability would most likely be reduced due to inbreeding depression.

Another possible limitation is due to errors in genotyping, which could be exacerbated by the lack of pedigree information for cows in this study. Wiggans et al. (2011) noted that with pedigrees and genotypes from parents, misclassified SNPs could often be corrected in the progeny. This type of correction, finding SNPs labeled as homozygous which were actually heterozygous and vice-versa, could provide more accurate estimates of genomic inbreeding, but correction of genotypes based on pedigrees may not be possible on many commercial dairies. Many genotyped animals on commercial farms would have genotyped sires, but few of the dams

of these animals would have been genotyped. Errors in genotyping would most likely result in underestimation of F_{ROH}. Because no heterozygous SNPs were allowed within a ROH, an error in which a homozygous SNP is mistakenly identified as heterozygous may result in a section of the genome that is not identified as IBD. This problem is less severe with F_{PH} and F_{GRM}, as those measures apply to single SNPs.

VanRaden et al. (2011b) discovered the presence of 5 recessive defects (haplotypes affecting fertility) in the Holstein, Brown Swiss, and Jersey populations. These defects were discovered by noting that no homozygous recessive animals were present in the population, despite a large number of heterozygous animals, suggesting that individuals that were homozygous for the haplotype in question did not survive full-term. This demonstrates both an effect of increased inbreeding and another possible way that genomic inbreeding measures presented by this study may have been underestimated with regard to the matings. If a sire and dam that were heterozygous for one of these haplotypes were mated, the resulting live progeny may be less genomically inbred than expected. As the present study focused on global genomic inbreeding, the results of VanRaden et al. (2011b) suggest further examination into local inbreeding may also be beneficial when examining production, reproduction, and health traits.

CONCLUSIONS

The three methods to quantify genomic inbreeding discussed in this study all demonstrate inbreeding depression for economically important traits in dairy cattle. Lactation performance and reproductive ability were negatively affected when any measure of genomic inbreeding increased. Among the methods considered, F_{PH}, F_{ROH}, and F_{GRM}, only F_{ROH} can distinguish between markers that are IBS and markers that are IBD. This information is important when evaluating the impact of inbreeding and when attempting to control inbreeding using

computerized mate selection algorithms. Furthermore, F_{ROH} has been previously shown to be the most correlated with homozygous mutation load and can also exploit the concept of a base population more effectively than the other measures of genomic inbreeding. Further work is needed to estimate the lifetime economic effects of increases in genomic inbreeding, as well as optimal strategies to balance genomic inbreeding and response to selection.

ACKNOWLEDGEMENTS

D. W. Bjelland acknowledges financial support from the USDA National Institute of Food and Agriculture National Needs Graduate Fellowship Grant Number 2010-38420-30477. I would also like to acknowledge the contribution of Donald Nkrumah, Natascha Vukasinovic, Sue DeNise, Prashanth Boddhireddy, Jason Osterstock, and Doug Ricke of Pfizer Animal Genetics (now Zoetis Genetics, Kalamazoo, MI).

REFERENCES

Adamec V., B.G. Cassell, E.P. Smith, and R.E. Pearson. 2006. Effects of inbreeding in the dam on dystocia and stillbirths in US Holsteins. J. Dairy Sci. 89:307-314.

Arias, J.A., M. Keehan, P. Fisher, W. Coppieters, and R. Spelman. 2009. A high density linkage map of the bovine genome. BMC Genetics. 10:18 doi:10.1186/1471-2156/10/18.

Carothers, A.D., I. Rudan, I. Kolcic, O. Polasek, C. Hayward, A.F. Wright, H. Campbell, P. Teague, N.D. Hastie, and J.L. Weber. 2006. Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. Ann. Hum. Genet. 70,666-676.

Cole, J.B and P.M. VanRaden. 2010. Visualization of results from genomic evaluations. J. Dairy Sci. 93:2727-2740.

Croquet, C., P. Mayeres, A. Gillon, H. Hammami, H. Soyeurt, S. Vanderick, and N. Gengler. 2006. Linear and curvilinear effects of inbreeding on production traits for Walloon Holstein Cows. J. Dairy Sci. 90:465-471.

Daetwyler, H.D., B. Villanueva, P. Bijma, and J.A. Woolliams. 2007. Inbreeding in genome-wide selection. J. Anim. Breed. Genet. 124:369-376.

de Roos, A.P.W., C. Schrooten, R.F. Veerkamp, and J.A.M. van Arendonk. 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. J. Dairy Sci. 94:1559-1567.

Ferencakovic, M., E. Hamzic, B. Gredler, I. Curik, and J Solkner. 2011. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh Cattle. ACS. 4:325-328.

Fisher, R.A. 1954. A fuller theory of "junctions" in inbreeding. Heredity. 8:187-1973.

Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate the gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal 1:21-28.

Hayes, B.J and M.E. Goddard. 2008. Technical note: Prediction of breeding values using marker-derived relationship matrices. J. Anim. Sci. 86:2089-2092.

Howrigan, D.P., M.A. Simonson, and M.C. Keller. 2011. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics. 12:460.

Illumina Inc. 2011. GoldenGate Bovine3K Genotyping BeadChip. Accessed Aug. 3, 2011. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine3K.pdf.

Kehrli M.E., F.C. Schmalstieg, D.C. Anderson, M.J. Van Der Maaten, B.J. Hughes, M.R. Ackermann, C.L. Wilhelmsen, G.B. Brown, M.G. Stevens, and C.A. Whetstone. 1990. Molecular definition of the bovine granulocytopathy syndrome: identification of deficiency of the Mac-1 (CD11b/CD18) glycoprotein. *Am. J. Vet. Res.* 11:1826–1836.

Keller, M.C., P.M. Visscher, and M.E. Goddard. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189:237-249.

Kirin, M., R. McQullian, C.S. Franklin, H. Campbell, P.M. McKeigue, et al. 2010. Genomic runs of homozygosity record population history and consanguinity. PLos One. 5(11): e13996. doi:10.1371/journal.phone.0013996.

Li, J.Z., D.M. Absher, H. Tand, A.M. Southwick, A.M. Casto, et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 319:1100-1104.

Mc Parland, S., J.F. Kearney, M. Rath, and D.P. Berry. 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. J. Dairy Sci. 90:441-4419.

Miglior F., E.B. Burnside, and B.W. Kennedy. 1995. Production traits of Holstein cattle: estimation of nonadditive genetic variance components and inbreeding depression. J. Dairy Sci. 78:1174-1180.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, et al. 2007. PLINK: a toolset for whole genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.

SAS Institute Inc. 2011. Version 9.1. SAS Institute, Inc., Cary, NC.

Sewalem, A., G.J. Kistemaker, F. Miglior, and B.J. Van Doomaal. 2006. Analysis of inbreeding and its relationship with functional longevity in Canadian dairy cattle. J. Dairy Sci. 89:2210-2216.

Shanks, R. D., D. B. Dombrowski, G. W. Harpestad, and J. L. Robinson. 1984. Inheritance of UMP synthase in dairy cattle. J. Hered. 75:337-340.

Simon-Sanchez, J., L.L. Kilarski, M.A. Nalls, M. Martinez, C. Schulte, et al. 2012. Cooperative genome-wide analysis shows increased homozygosity in early onset Parkinson's Disease. PLoS One. 7(3): e28787. doi:10.1371/journal.pone.0028787

Smith, L.A., B.G. Cassell, and R.E. Pearson. 1998. The effects of inbreeding on the lifetime performance of dairy cattle. J. Dairy Sci. 81:2729-2737.

Sonesson, A.K., J.A. Wolliams, and T.H.E. Meuwissen. 2012. Genomic selection requires genomic control of inbreeding. GSE. 44:27.

Thompson, J.R., R.W. Everett, and N.L. Hammerschmidt. 2000. Effects of inbreeding on production and survival in Holsteins. J. Dairy Sci. 83:1856-1864.

US Department of Agriculture-Animal Improvement Laboratories, 2012. Bovine Inbreeding Trends. Accessed Aug. 3, 2012. http://aipl.arsusda.gov/eval/summary/inbrd.cfm.

VanRaden, P. M., A. H. Sanders, M. E. Tooker, R. H. Miller, H. D. Norman, M. T. Kuhn, and G. R. Wiggans. 2004. Development of a national genetic evaluation for cow fertility. J. Dairy Sci. 87:2285-2292.

VanRaden, P.M., K.M. Olson, G.R. Wiggans, J.B. Cole, and M.E. Tooker. 2011a. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 94:5673-5682.

VanRaden, P.M., K.M. Olson, D.J. Null, and J.L. Hutchison. 2011b. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy. Sci. 94:6153-6161.

Weigel, K.A. 2001. Controlling inbreeding in modern breeding programs. J. Dairy Sci. 84:E177-E184.

Wiggans, G.R., P.M. VanRaden, and T.A. Cooper. 2011. The genomic evaluation system in the United States: Past, present, and future. J. Dairy Sci. 94:3202-3211.

Zimin A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz., et al. 2009. A whole-genome assembly of the domestic cow, Bos Taurus. Genome Biol. 10:R42.

Table 1. Estimates of inbreeding depression for production traits, expressed as change in phenotype per 1% increase in percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}).

	Phenotypic	Phenotypic	F_{PH}		F_{ROH}		F_{GRM}	
Item	mean	SD	Estimate	SE	Estimate	SE	Estimate	SE
205-d milk yield (kg)	8473	1586	-53***	19	-20**	10	-47**	22
Peak milk (kg)	64	13	-0.22	0.16	-0.14	0.09	-0.17	0.19
Average daily milk (kg)	44	8	-0.28***	0.10	-0.11**	0.05	-0.25**	0.11
Average fat (%)	3.63	0.59	-0.003	0.008	-0.001	0.005	-0.003	0.004
Average protein (%)	3.01	0.24	-0.002	0.004	0.001	0.002	-0.001	0.002
Somatic cell score (log_2 cells/ml)	2.99	0.42	0.001	0.006	0.004	0.003	0.001	0.003
Milk urea nitrogen (mg/dL)	13.5	2.4	-0.06**	0.03	-0.02	0.02	-0.03**	0.01

*P < 0.1, **P < 0.05, ***P < 0.01

Table 2. Estimates of inbreeding depression for reproductive traits, expressed as change in phenotype per 1% increase in percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}).

	Phenotypic	Phenotypic	F_{PH}		F_{ROH}		F_{GRM}	
Item	mean	SD	Estimate	SE	Estimate	SE	Estimate	SE
Days open (d)	123	60	1.76*	1.00	1.72***	0.54	1.06**	0.52
Conception rate (%)	59.5	35.5	-0.82	0.60	-0.82**	0.33	-0.53*	0.31
Days in milk at 1st breeding (d)	72.6	19.6	0.27	0.24	0.20	0.13	0.14	0.12
Calf birth weight (kg)	40.1	4.8	-0.4**	0.2	-0.1	0.1	-0.2***	0.1
Calving ease (5-point scale)	1.7	0.9	0.09***	0.03	0.03**	0.02	0.04**	0.02

*P < 0.1, **P < 0.05, ***P < 0.01

Table 3. Estimates of inbreeding depression for linear type traits on a 50-point scale, expressed as change in phenotype per 1% increase in percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}).

	Phenotypic	Phenotypic	F_{PH}		F_{ROH}		F_{GRM}	
Item	mean	SD	Estimate	SE	Estimate	SE	Estimate	SE
Stature	31.7	8.7	-0.06	0.12	0.07	0.06	-0.06	0.06
Strength	29.0	8.8	-0.40***	0.11	-0.11*	0.06	-0.19***	0.06
Body depth	27.9	8.9	-0.25**	0.11	-0.01	0.06	-0.14**	0.06
Dairy form	29.3	8.1	-0.04	0.11	0.07	0.06	-0.03	0.05
Rump angle	22.6	9.5	-0.20	0.15	-0.05	0.08	-0.14*	0.08
Rump width	28.8	10.2	-0.06	0.13	0.07	0.07	-0.03	0.07
Rear legs side view	27.5	8.2	-0.06	0.13	-0.04	0.07	-0.06	0.07
Rear legs rear view	23.5	10.3	-0.35**	0.16	-0.16*	0.09	-0.14*	0.08
Foot angle	25.5	9.4	-0.16	0.14	-0.05	0.08	-0.04	0.08
Fore udder attachment	23.7	11.1	-0.13	0.17	0.06	0.09	-0.02	0.09
Udder height	26.0	10.9	-0.15	0.16	0.01	0.09	-0.05	0.08
Udder width	29.9	10.6	-0.21	0.15	-0.05	0.08	-0.05	0.08
Udder cleft	32.4	10.7	0.13	0.16	0.12	0.09	0.04	0.08
Udder depth	22.4	11.1	0.11	0.13	0.14**	0.07	0.06	0.07
Front teat placement	29.3	9.9	0.35**	0.15	0.25***	0.08	0.18**	0.08
Rear teat placement	29.6	11.9	0.27	0.18	0.25***	0.10	0.12	0.09
Teat length	27.0	9.1	-0.24*	0.13	-0.14*	0.07	-0.13*	0.07
Udder tilt	25.2	9.5	0.24*	0.14	0.15**	0.07	0.10	0.07

*P < 0.1, **P < 0.05, ***P < 0.01

Table 4. Estimates of inbreeding depression for all significant traits, expressed as the difference in predicted phenotype between plus or minus 2 SD from the mean for percent homozygosity (F_{PH}), inbreeding coefficient derived from runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and inbreeding coefficient calculated from a genomic relationship matrix (F_{GRM}).

		F_{PH}			F_{ROH}			F_{GRM}	
Item	58.20%	62.76%	Diff	0%	8.06%	Diff	16.10%	25.42%	Diff
205-d milk yield (kg)	8,594	8,352	-242	8,554	8,392	-161	8,692	8,254	-438
Average daily milk (kg)	45	43	-1.28	44	43	-0.89	45	43	-2.33
Milk urea nitrogen (mg/dL)	13.7	13.4	-0.3	-	-	-	13.7	13.4	-0.3
Days open (d)	119	127	8	116	130	14	118	128	10
Conception rate (%)	-	-	-	62.8	56.2	-6.6	62.0	57.0	-4.9
Calf birth weight (kg)	41.0	39.2	-1.8	-	-	-	41.0	39.2	-1.9
Calving ease (5-point scale)	1.5	1.9	0.41	1.6	1.8	0.24	1.5	1.9	0.37
Strength	29.9	28.1	-1.8	29.5	28.6	-0.9	29.9	28.1	-1.8
Body depth	28.4	27.3	-1.1	-	-	-	28.5	27.2	-1.3
Rump angle	-	-	-	-	-	-	23.2	21.9	-1.3
Rear legs rear view	24.3	22.7	-1.6	24.2	22.9	-1.3	24.2	22.9	-1.3
Udder depth	-	-	-	21.8	23.0	1.1	-	-	-
Front teat placement	28.5	30.1	1.6	28.3	30.3	2.0	28.4	30.1	1.7
Rear teat placement	-	-	-	28.6	30.6	2.0	-	-	-
Teat length	27.5	26.4	-1.1	27.5	26.4	-1.1	27.6	26.3	-1.2
Udder tilt	24.6	25.7	1.1	24.6	25.8	1.2	-	-	-

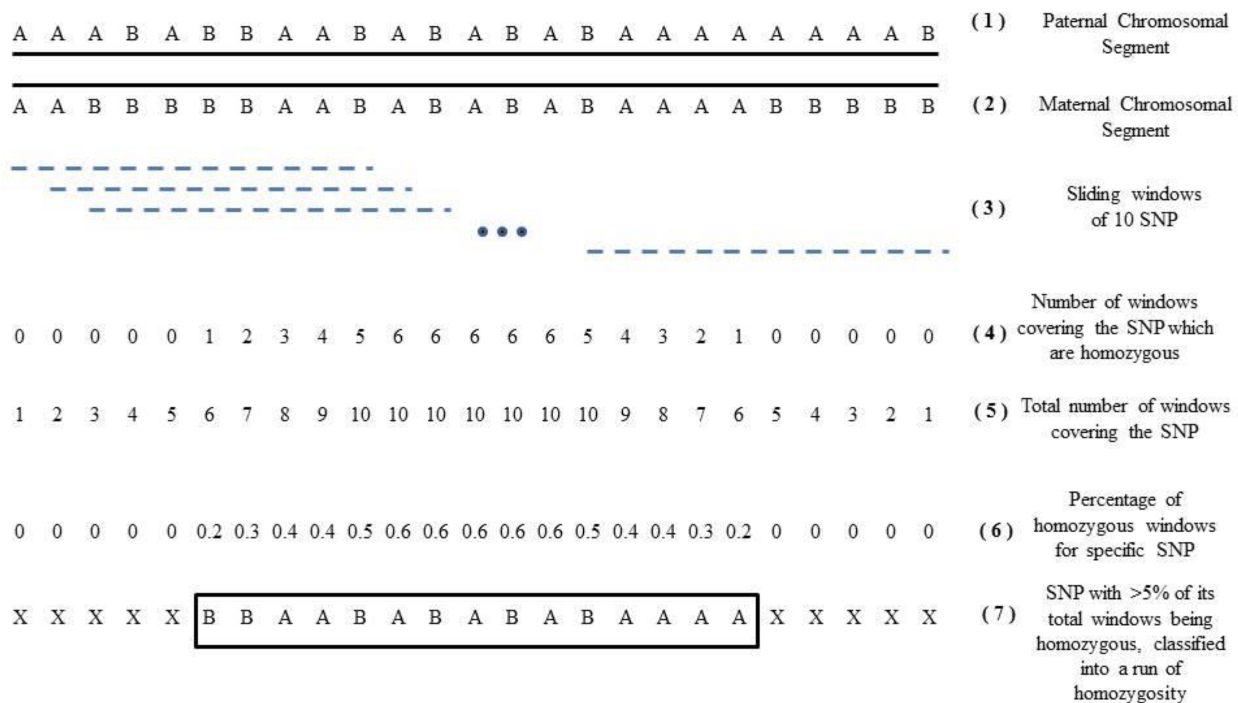


Figure 1. Description of the process for discovery of runs of homozygosity (ROH) using a sliding window of SNP markers along the chromosome, as implemented with PLINK software (Purcell et al., 2007)

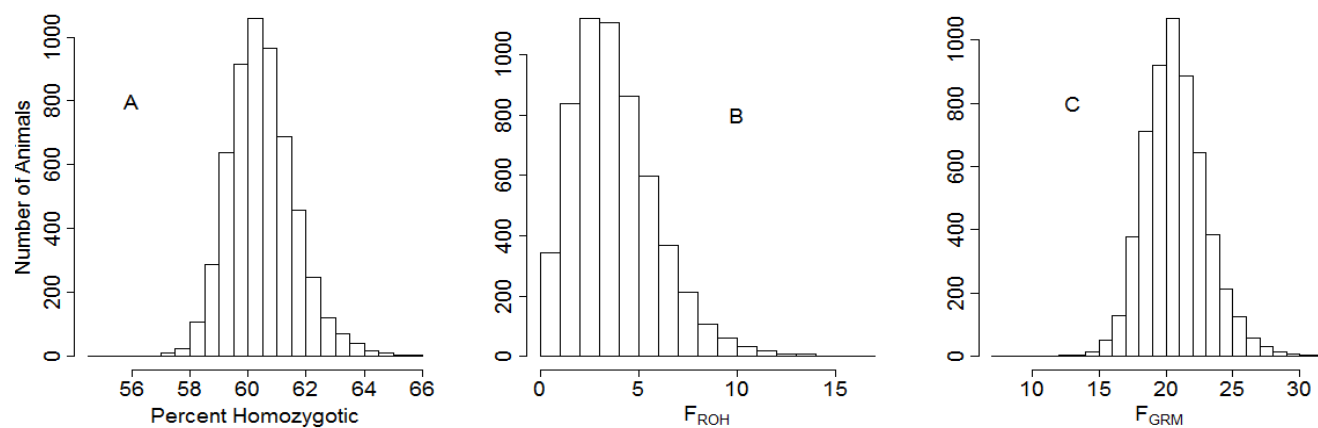


Figure 2. Frequency distribution of animals in the present study, according to (**A**) percent homozygosity (F_{PH}), (**B**) runs of homozygosity (F_{ROH}) with a minimum length of 30 SNP, and (**C**) inbreeding coefficient derived from a genomic relationship matrix (F_{GRM}) using the method of VanRaden et al. (2011a).

CHAPTER 4

Comparison of genomic inbreeding in Holstein, Jersey, Angus, and Nelore cattle using dense and reduced SNP marker panels

ABSTRACT

Increases in pedigree inbreeding levels, as well as losses in performance due to inbreeding depression, have been well documented among commercial livestock populations. With recent advances in genomic tools available for animal breeders, calculation of more precise measures of inbreeding based on genomic markers is now possible. Comparison of genomic inbreeding between breeds will give researchers a broader understanding of the current structure of commercial livestock breeds and implications of intense selection. A large number (54,001) of single nucleotide polymorphism (SNP) markers were available from 6,600 commercial Holsteins, 2,402 Angus, and 2,302 Nelore cattle while 43,485 SNP markers were available from 7,883 genetically elite Holsteins and 3,146 Jerseys. A subset of 6,909 SNP contained within the Illumina BovineLD BeadChip (Illumina Inc, San Diego, CA) and a subset of roughly 6,400 equally spaced SNP were extracted to create two lower density SNP sets. From each of the three SNP panels, measures of genomic inbreeding based on percent homozygosity, runs of homozygosity (ROH) with a minimum length of 4,000 kb, and ROH with a minimum of 8,000 kb were derived. Jersey cattle had the highest inbreeding values when derived from ROH, followed by elite Holsteins, Angus, and commercial Holsteins, with Nelore having the lowest levels of genomic inbreeding. Between the two Holstein groups, elite Holsteins had higher percent homozygosity and higher inbreeding coefficients derived from ROH than commercial Holsteins, indicating that the animals under greater selection intensity also exhibited higher inbreeding values. Ascertainment bias in the selection of the SNP for lower density SNP panels was evident with much lower average percent homozygosity for each breed, except Nelore, when calculated with SNP from the low density panel rather than the 50K or equally spaced 6K SNP panels. In lower density SNP panels, minimal differences between inbreeding derived from ROH

with a minimum of 4,000 kilobases and 8,000 kilobases indicate that low density SNP panels may be capable of detecting only recent inbreeding. Measures of genomic inbreeding vary by breed, SNP panel density, and intensity of the selection within the population. These measures may be incorporated into a genomic selection program to help control future inbreeding.

INTRODUCTION

Inbreeding in livestock populations has been traditionally calculated through pedigrees and has steadily increased over the past several decades. The average pedigree inbreeding (F_{ped}) coefficient for US Holsteins has risen from 0.4% in 1970 to 5.8% in 2012, while the F_{ped} for US Jerseys has risen from 0.8 to 7.1% over the same time period (USDA-AIPL, 2012). Increases in F_{ped} are present in the Brazilian Nelore population as well, with the F_{ped} of animals born from 1979 to 1983 of 0.9% and F_{ped} of animals born from 1994 to 1998 of 2.1% (Faria et al., 2009).

Development of high-throughput genomic tools, such as the Illumina Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA), have provided new tools for animal breeders, such as the development of breeding values based on genomic information (VanRaden et al. 2009). Initially, the majority of animals genotyped were genetically elite young bulls and females by artificial insemination (AI) companies. As costs of genotyping have decreased and acceptable reliabilities have been achieved from lower density (3,000 and 6,000 SNPs) SNP panels (Wiggans et al., 2012), more commercial cows have been genotyped to help producers make selection decisions in their herds. Economic studies have also determined that genotyping replacement dairy heifers on commercial operations, even with low density SNP panels, could be beneficial to producers (Weigel et al, 2012).

Information from genomic markers will also be able to provide a more accurate quantification of inbreeding. The inbreeding coefficient derived from pedigrees (F_{ped}) is defined

as the probability that a pair of alleles is identical by decent (IBD). Although significant inbreeding depression has been observed due to increases in F_{ped} in Holsteins (Smith et al., 1998) and Angus (Carrillo and Siewerdt, 2010), this measure also contains a large amount of variability. For example, consider an organism whose parents are first cousins (with no other previous common ancestors): F_{ped} would be 6.25%; an average of 6.25% of this organism's genome would be identical, having originated from either of the common great-grandparents. Carothers et al. (2006) noted that the actual value varies greatly due to Mendelian sampling, with a standard deviation of 2.43%. This deviation depends on the recombination events that occurred during gamete formation in the parents, as well as the chance meeting of the successful gametes. A measure of inbreeding utilizing genomic marker information will be able to more accurately depict the exact proportion of alleles which are IBD. Furthermore, Sonesson et al. (2012) noted that when calculating genomic breeding values a correction based on genomic inbreeding, instead of F_{ped}, is required for accurate prediction of breeding values.

Increased levels of inbreeding depicted in genomic markers would appear as increases in the frequency of homozygous alleles. A simple method to determine inbreeding genomically would be to calculate the percentage of alleles which are homozygous, with the higher number indicating a more inbred individual. One problem with this method is that this does not distinguish between alleles which are IBD and those that are identical by state (IBS). Utilizing runs of homozygosity (ROH) can differentiate the two. The ROH (with one chromosomal strand coming from the father and one from the mother) originate from an ancestor common to both the animal's sire and dam. Segments like these are transmitted through the pedigree to an animal in the current generation, where the segments come together to form a ROH which is IBD. The length of the ROH correlates to the distance within the pedigree until the common ancestor is

observed. Longer ROH indicate more recent common ancestors, because recombination has had fewer generations to break up the segments, whereas shorter ROH indicate a common ancestor further back in the pedigree. When discovering ROH in an individual, increasing the minimum length of the ROH would result in focusing on more recent inbreeding and would be analogous to changing the base population to a more recent year or generation with respect to calculating pedigree inbreeding. Fisher (1954) noted that the expected length of the DNA segment which is IBD follows an exponential distribution with mean equal to $\frac{1}{2g}$ Morgans, where g equals the number of generations since the common ancestor. Common ancestors occurring 10 generations back would have a mean ROH length of 0.05 Morgans, or 5 cM. Studies involving ROH as an inbreeding measure for population history (Kirin et al., 2010; Li et al., 2008) and disease risk (Simon-Sanchez et al. 2012) in human populations have been performed recently. Keller et al. (2011) noted that inbreeding involving ROH ($\mathbf{F_{ROH}}$) is more accurate at depicting DNA segments which are truly IBD than other measures of genomic and pedigree inbreeding due to a higher correlation with homozygous mutation load. Furthermore, in a comparison of varying measures of F_{ROH} and F_{ped} in a population of 500 Simmental bulls, Ferencakovic et al. (2011) noted positive correlations that ranged from 0.61 to 0.68.

The goals of this study were to compare genomic inbreeding values between populations of Jersey, commercial Holstein, elite Holstein, Angus, and Nelore cattle and to determine the differences in genomic inbreeding derived from medium density SNP marker panels and lower density SNP marker panels.

MATERIALS AND METHODS

Data

Genotypic information from 3 *Bos taurus* breeds, Holstein, Jersey, and Angus, and 1 *Bos inducus* breed, Nelore, were available for this study. Furthermore, as genetically elite animals are a product of intense genetic selection from both the maternal and paternal pathways, inbreeding in these elite animals may differ from commercial animals. To study this, two sets of Holsteins, one of genetically elite animals and the other of commercial cows, were included. The numbers of total animals and SNP for each genetic group in this study are presented in Table 1. Genotypes for the 7,883 elite Holstein and 3,146 Jersey cattle (2,656 males and 490 females) were provided by the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) and consisted of 43,485 SNP makers throughout the 29 *Bos taurus* autosomes and the X chromosome. These SNP represent the subset of markers on the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) that are used for routine evaluation of US dairy cattle, after removal of SNP with a call rate of <90%, greater than 1% parent-progeny conflicts, complete linkage disequilibrium with an adjacent SNP, or minor allele frequency (MAF) of <1% in each of the Holstein, Jersey, and Brown Swiss breeds (Wiggans et al., 2009). Genotypes for 6,660 commercial Holstein, 2,402 Angus, and 2,302 Nelore cattle were provided by Zoetis Genetics (Kalamazoo, MI) and consisted of 54,001 SNP markers throughout the 29 autosomes and the X chromosome. All genetic groups went through additional quality control with the removal of SNP for MAF < 0.05, call rate (percent missing > 0.1), and violation of Hardy-Weinberg equilibrium ($p < 0.0001$). Individual animals with greater than 10% missing SNP were also removed from the analysis. In order to provide a more uniform genetic population, all animals born before 1995 were also removed. After edits, 2,253 Jersey cattle with 31,873 SNP, 6,510 elite Holstein cattle with 37,374 SNP, 4,386 commercial Holstein cattle with 32,045 SNP, 2,337 Angus cattle with 38,559 SNP, and 2,216 Nelore cattle with 23,319 SNP remained for analysis in the 50K SNP set.

The genetically elite Holsteins were the first group of Holstein cattle genotyped in the US and were comprised of sires from AI organizations which had very reliable breeding value estimations and also genetically elite dams of AI sires. The commercial Holsteins consisted of cattle from 9 herds throughout the US and contained every animal for a single generation. The Jersey and Angus groups were similar to the elite Holsteins, in that they were AI sires with highly reliable breeding values and dams of sires. The Nelore group contained genotypes on all cattle within 2 separate Brazilian herds.

A lower density SNP subset was extracted from the original data and was comprised of the 6,909 SNP contained within the Illumina BovineLD BeadChip (Illumina Inc, 2012). After the extraction of the SNP, the same quality control edits of removing SNP for $MAF < 0.05$, call rate (percent missing > 0.1), and violation of Hardy-Weinberg equilibrium ($p < 0.0001$) were made, as well as the removal of individuals with greater than 10% missing SNP. A total of 2,253 Jersey cattle with 5,318 SNP, 6,503 elite Holstein cattle with 5,556 SNP, 4,385 commercial Holstein with 5,902 SNP, 2,335 Angus with 6,417 SNP, and 2,215 Nelore with 4,201 SNP remained for analysis for the low density SNP set.

The Illumina BovineLD BeadChip was designed by selecting SNP which were the most reliable, polymorphic, uniformly distributed, and useful for imputation in the most common breeds of cattle, with the majority of the influence coming from the Holstein, Jersey, Angus, and Brown Swiss breeds (Illumina, 2012). In addition, sets of nearly 6,400 equally spaced SNP were selected for each genetic group in order to detect any ascertainment bias that may have been present due to the pre-selection of the SNP on the Illumina BovineLD BeadChip. After SNP editing on the original genotypes for each breed, SNP were ordered on each chromosome, with the equally spaced 6K SNP set including every fourth SNP for the Nelore cattle, every fifth SNP

for the Jersey and commercial Holstein cattle, and every sixth SNP for the elite Holstein and Angus cattle. As the selected SNP had already gone through quality control checks, only the removal of animals with greater than 10% missing SNP was required. A total of 2,253 Jersey cattle with 6,375 SNP, 6,510 elite Holstein cattle with 6,096 SNP, 4,386 commercial Holstein cattle with 6,329 SNP, 2,337 Angus cattle with 6,356 SNP, and 2,215 Nelore cattle with 5,597 SNP remained for analysis in the equally spaced 6K SNP set.

Pedigree information and F_{ped} values were available only for elite Holstein and Jersey cattle and were obtained from the USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD).

Genomic Inbreeding Coefficients

The first measure of genomic inbreeding considered was percent homozygosity (**F_{PH}**) of all SNP in each of the 50K, low density, and equally spaced 6K SNP sets. F_{PH} was derived with the formula:

$$F_{PH} = \frac{N_{AA} + N_{BB}}{N_{AA} + N_{AB} + N_{BB}}$$

where N_{AA}, N_{AB}, and N_{BB} refer to the number of SNP that are classified as AA, AB, and BB, respectively.

While the F_{PH} of an animal can provide some indication of its level of inbreeding, it does not distinguish between markers that are IBS and those that are IBD. One possibility to alleviate this problem is to consider genomic ROH, which is defined as a specific number of consecutive SNP which are all homozygous. Discovery of ROH was performed using PLINK (Purcell et al., 2007). Figure 1, previously presented by Bjelland et al. (2013), describes the manner by which ROH were discovered using a sliding window of SNP along the chromosome. An individual's maternal and paternal chromosomes are presented, and a sliding window of 10 SNP (dashed line

below the chromosomal sequence) moves along the chromosome one SNP at a time. Each of these windows is determined to be either completely homozygous or not, with the total number of completely homozygous windows and total number of windows summed for each SNP (lines 4 and 5, respectively). Next, the percentage of sliding windows which are homozygous for each SNP is calculated (line 6). If this percentage is greater than 5% for at minimum 10 consecutive SNP, a ROH is discovered. Recommendations for many of the input parameters for ROH discovery were derived from Howrigan et al. (2011). Within the sliding window, no heterozygous SNP and 1 missing SNP were allowed. Since high linkage disequilibrium within given sections of DNA can lead to detection of ROH that are not truly IBD, linkage disequilibrium pruning was performed on the SNP set prior to ROH determination to increase power, as suggested by Purcell et al. (2007). Linkage disequilibrium pruning was also performed using PLINK, and SNP which had an $r^2 > 0.5$ with all other SNP in a 50 SNP sliding window were removed. Table 1 displays the total number of SNP remaining for the Jersey (4,580, 1,908, and 2,098), elite Holstein (7,775, 2,378, and 2,609), commercial Holstein (7,780, 3,144, and 3,365), Angus (7,061, 3,049, and 3,107), and Nelore (9,915, 3,512, and 4,378) populations, for the 50K SNP set, low density SNP set, and equally spaced 6K SNP set, respectively. For ROH discovery, a minimum length of 15 SNP was used to discover ROH for the 50K SNP, and a minimum length of 10 SNP was used to discover ROH for both the low density and 6K SNP sets. All ROH with a SNP density (kb per SNP) of greater than the mean plus 3 times the standard deviation were removed from each genetic group and SNP set. With the greater distance between SNP, these ROH would have the greatest chance of being false positives. In order to provide a more uniform comparison between genetic groups and SNP sets, two datasets were formed for each of these combinations; the first included all ROH discovered that were greater

than 4,000 kb and the second included all ROH discovered that were greater than 8,000 kb. These two cutoff values give genomic inbreeding measures that encompass more distant and more recent inbreeding, respectively.

Results of the ROH discovery were then utilized to create an inbreeding coefficient for each animal, denoted as F_{ROH4} when the minimum length of ROH was 4,000 kb and F_{ROH8} when the minimum length of ROH was 8,000 kb, which was calculated by the formula:

$$F_{ROH4} or\ F_{ROH8} = \frac{\sum_k length(ROH_k)}{L}$$

where k = number of ROH for each animal, and L = total length of the genome. Length of ROH were measured in kilobases (kb) with L = 2,612,820 kb (Zimin et al., 2009)

RESULTS AND DISCUSSION

The number of SNP remaining after SNP editing and linkage disequilibrium pruning varied greatly between genetic groups. Many more SNP were removed due to low minor allele frequency for the Nelore cattle than any other genetic group. This would be expected since the majority of the SNP selected for the 50K SNP set were the most polymorphic for Holstein, Jersey, and Angus cattle (Illumina, 2012). The Bovine Hapmap Consortium (2009) noted that when *taurine* breeds were used for SNP discovery based on high MAF, about 30% of the SNP had a MAF > 0.3 within the *taurine* breeds, while only 19% of the SNP had a MAF > 0.3 within the *indicine* breeds. Jersey cattle had many fewer SNP remaining after linkage disequilibrium pruning than the other genetic groups, while Nelore cattle had more SNP remaining. This is consistent with previous results from the Bovine Hapmap Consortium (2009), which indicated that Jersey had much higher r^2 values among SNP while *Bos indicus* cattle had lower 4^2 values at short distances when compared to other breeds.

Frequency distributions of F_{PH}, F_{ROH4}, and F_{ROH8} derived from the 50K SNP set for each genetic group are provided in Figure 2. Nelore (68.4 ± 1.0%) had the largest mean F_{PH}, followed by Jersey (63.5 ± 1.7%), Angus (62.8 ± 1.3%), and elite Holstein cattle (62.6 ± 1.3), with the commercial Holstein cattle (60.3 ± 1.1%) having the lowest mean F_{PH}. The Jersey cattle (11.1 ± 4.0%) had the largest mean F_{ROH4} value, followed by the elite Holstein (8.1 ± 2.9%), Angus (7.9 ± 2.8%), and commercial Holstein cattle (6.8 ± 2.4%), with Nelore (5.2 ± 2.4%) having the lowest mean F_{ROH4} value. When the minimum length of ROH discovered was increased from 4,000 kb to 8,000 kb, Jersey cattle (10.1 ± 3.9%) still had the largest mean F_{ROH8} value, followed by elite Holstein cattle (6.3 ± 2.8%), Angus (5.2 ± 2.7%), and commercial Holstein cattle (4.9 ± 2.2%), with Nelore (2.9 ± 2.1%) having the lowest mean F_{ROH8} value.

The higher inbreeding values for the elite Holsteins compared to the commercial Holsteins is expected, since greater selection intensity is placed on both the paternal and maternal pathways of the elite animals, whereas commercial animals usually only have intense selection pressure from the paternal side. Smith et al. (1998) had previously reported that Holstein cattle from registered (elite) herds had higher pedigree inbreeding than cows from grade (commercial) herds, although part of the difference in this study may be attributed to the incomplete pedigrees of the grade cattle. The Jersey cattle, which were also a group of elite genetic animals, had by far the largest average F_{ROH4} and F_{ROH8} values. This is not surprising when comparing pedigree inbreeding of Jersey (7.9%) and Holstein (5.8%) cattle calculated by the USDA-AIPL (2012). The lower F_{ROH4} and F_{ROH8} values from Nelore cattle are also consistent with previous estimates of low levels of pedigree inbreeding (2.1%), although differences in base population and year of calculation may also reduce the estimates of Nelore when compared to either Holstein or Jersey (Faria et al., 2009). When calculating genomic

inbreeding, a constant minimum length of ROH between breeds would essentially be analogous to using the same number of generations as a base population in pedigree inbreeding.

Although the Nelore cattle had the highest mean F_{PH} and the lowest mean F_{ROH4} and F_{ROH8}, the correlations between F_{PH} and F_{ROH4} (0.82) and between F_{PH} and F_{ROH8} (0.79) were fairly high. Commercial Holstein cattle had slightly higher correlations between F_{PH} and F_{ROH4} (0.85) and between F_{PH} and F_{ROH8} (0.83). Elite Holstein (0.89 and 0.87), Angus (0.90, and 0.86), and Jersey (0.91 and 0.90) cattle all had greater correlations between F_{PH} and F_{ROH4}, and F_{PH} and F_{ROH8}, respectively. As F_{PH} is essentially a ROH with a minimum length of 1 SNP, high correlations are expected. Correlations also seem to depend on the number of SNP remaining for analysis after linkage disequilibrium pruning, with Jersey cattle having the highest correlations and the fewest remaining SNP and Nelore cattle have the lowest correlations and the most remaining SNP. The similar correlations between F_{PH} and F_{ROH4} and between F_{PH} and F_{ROH8} in Jersey cattle may also be due to the fact that fewer SNP remained after linkage disequilibrium pruning. The lower density of SNP resulted in fewer ROH being discovered in the range between 4,000 and 8,000 kb.

The mean number of ROH found was similar between genetic groups with a minimum ROH length of 4,000 kb. Angus (23 ± 5) averaged the greatest number of ROH, followed by Jerseys (20 ± 6), elite Holsteins (19 ± 5), commercial Holsteins (17 ± 5), and Nelore (17 ± 6). When the minimum length of ROH was increased to 8,000 kb, Jersey cattle (16 ± 5) averaged the greatest number of ROH, followed by elite Holstein (11 ± 4), Angus (10 ± 4), commercial Holstein (9 ± 3), and Nelore cattle (6 ± 3). For the Nelore cattle, 51 out of the 2,215 animals had no ROH of at least 8,000 kb in length when discovered using the 50K SNP set, although all of these animals had a ROH with a minimum length of 4,000 kb. These animals would have an

F_{ROH8} value of 0, indicating no genomic inbreeding. By contrast, 2 elite Holstein cattle had an F_{ROH8} of 0; 2 commercial Holstein cattle had an F_{ROH4} of 0 while 5 had an F_{ROH8} of 0, and 1 Angus had an F_{ROH8} value of 0. All Jersey cattle had at least 2 ROH for both the minimum lengths of 4,000 and 8,000 kb.

The large drop in average number of ROH for all genetic groups except Jersey cattle may suggest higher levels of more ancient inbreeding, while in recent generations matings of closely related individuals have been largely avoided. Angus and Nelore seemed to show the largest drop, with a difference of 13 and 11 ROH discovered, respectively. Although Angus cattle averaged the greatest number of ROH with a minimum of 4,000 kb, it is evident that the majority of these are very short, considering the large drop in average number of ROH with a minimum of 8,000 kb, as well as the fact that both the Jersey (11.1%) and elite Holstein (8.1%) cattle had greater overall average F_{ROH4} values than Angus (7.9%). This may suggest that although a population bottleneck had occurred during breed formation, as suggested by the Bovine Hapmap Consortium (2009), measures to avoid inbreeding in recent generations have been effective. Furthermore, natural service is more widespread in Angus than in either the Holstein or Jerseys. This would mean there are more sires contributing DNA to the next generation, which would lead to lower inbreeding for the Angus population. In Jersey cattle, where very little change was seen, there is evidence of both a population bottleneck and difficulty in avoiding recent inbreeding, possibly due to a small effective population size and widespread use of AI.

The most inbred animals varied greatly between breeds, while the animal with the largest F_{ROH4} and F_{ROH8} values was an Angus (36.3 and 32.6% respectively). This indicates that about one third of this animal's genome is IBD and contained within a ROH. This animal also had a very high F_{PH} of 74.9%. There were only a total of 9 Angus cattle with an F_{ROH4} greater than

20%, and only 5 with an F_{ROH8} greater than 20%. The most inbred Jersey had an F_{ROH4} equal to 32.4%, F_{ROH8} equal to 30.7%, and F_{PH} equal to 72.0%. The F_{ped} for this animal was 18.6%. Other than this bull, no other Jersey had an F_{ROH4} greater than 25%, but there were a total of 47 Jersey cattle with an F_{ROH4} greater than 20% and 22 with an F_{ROH8} greater than 20%. The most inbred elite Holstein had an F_{ROH4} equal to 26.3%, an F_{ROH8} equal to 24.4%, F_{PH} equal to 69.7%, with an F_{ped} of 15.2%. Thirteen (out of 6,509) elite Holsteins had an F_{ROH4} greater than 20% and 6 had an F_{ROH8} greater than 20%. No commercial Holstein or Nelore cattle were greater than 20% inbred for either F_{ROH4} or F_{ROH8}, with the most inbred commercial Holstein having an F_{ROH4} of 16.8% and an F_{ROH8} of 14.4% and the most inbred Nelore having an F_{ROH4} of 18.4% and an F_{ROH8} of 15.8%.

Comparisons of F_{ped} to F_{PH}, and F_{ped} to F_{ROH4} from the 50K SNP panel, are displayed in Figure 6 for elite Holsteins (A and B) and Jersey cattle (C and D). Comparisons made utilizing F_{ROH8}, the low density SNP panel, and the equally spaced 6K SNP panels were very similar to the illustrations in Figure 6, and thus are not presented herein. For elite Holstein cattle, F_{ROH4} had the highest correlation with F_{ped}, followed by F_{ROH8}, and then F_{PH} (0.65, 0.63, and 0.58) with genomic inbreeding derived from the 50K SNP panel. Similar results were observed with the low density SNP panel (0.64, 0.63, and 0.57) and the equally spaced 6K SNP panels (0.61, 0.61, and 0.53) when comparing F_{ped} to F_{ROH4}, F_{ROH8}, and F_{PH}, respectively. Slightly higher correlations were observed with Jersey cattle for the 50K SNP panel (0.69, 0.68, and 0.62), low density SNP panel (0.67, 0.67, and 0.61), and equally spaced 6K SNP panels (0.67, 0.67, and 0.59), when comparing F_{ped} to F_{ROH4}, F_{ROH8}, and F_{PH}, respectively. The correlations regarding F_{ped} compared to both F_{ROH4} and F_{ROH8} are similar to the results of Ferencakovic et al. (2011) who presented correlations between 0.61 and 0.68 when comparing F_{ped} to FROH with varying minimum length

of ROH. For both the elite Holstein and Jersey cattle, the higher density 50K SNP panel had higher correlations with F_{ped} than the low density and equally spaced 6K SNP panels.

As expected, some of the highly selected, genetically elite animals exhibited the greatest amount of genomic inbreeding. The fact that the Jerseys and elite Holsteins with the highest genomic inbreeding also had a very high pedigree inbreeding coefficient supports the usefulness of these measures. VanRaden et al. (2011) had also noticed long segments of DNA with completely homozygous SNP in animals who had several common grandparents and great-grandparents. An exact match between F_{ped} and genomic inbreeding would not be expected for several reasons. One is that F_{ped} is an estimate of the percentage of alleles that are IBD while genomic inbreeding actually attempts to calculate the true value. Another is that the base populations between the two are very different. The F_{ped} can vary depending on how many generations of pedigree data are used; genomic inbreeding does essentially the same thing by varying the minimum length of the ROH discovered. One of the main challenges in looking at more ancient inbreeding is the current density of the conventional SNP marker panels. With a denser SNP panel, shorter ROH may be detected, which could possibly provide a better understanding of the population history of breeds of cattle, as in human populations involved in the Human Genome Diversity Project (Kirin et al., 2010). For example, a large number of shorter ROH could suggest population bottlenecks due to breed formation in our current domestic cattle. A denser SNP panel would also improve the accuracy of ROH discovery. However, in utilizing linkage disequilibrium pruning, as is required with the current method to determine ROH, determining very short ROH may still be difficult due to the population structure and long haplotypes present in domestic cattle.

Figure 3 displays the F_{PH} derived from 50K, low density, and equally spaced 6K SNP marker panels. Evidence of ascertainment bias in selection of the low density SNP marker panel is evident, with means for F_{PH} derived from the 50K and equally spaced 6K SNP sets being greater than F_{PH} derived from the low density SNP for commercial Holsteins (60.3, 60.4, and 54.6%), elite Holsteins (62.6, 62.7, and 55.8%), Angus (62.8, 62.9, and 58.5%), and Jersey (63.5, 63.4, and 60.1%) cattle. The largest drop in F_{PH} is seen in the two Holstein groups, as expected since 82.5% (91,081 out of 110,409) of the animals used in construction of the low density SNP panel, which were selected in part based on a high minor allele frequency, were Holsteins (Illumina, 2012). Angus and Jersey cattle made up 6.0 and 5.9% of the samples used in construction of the low density SNP panel, respectively. Both of these breeds also exhibited a large decrease when comparing F_{PH} from 50K to F_{PH} from the low density panel. As no Nelore, and less than 0.1% *Bos indicus* cattle, were included in the construction of the low density SNP panel, the similarities between all three measures of F_{PH} (68.4, 68.4, and 67.1%) are not surprising.

Figures 4 and 5 display the F_{ROH4} and F_{ROH8} measures from the 50K, low density, and equally spaced 6K SNP panels for each of the genetic groups, respectively. One result that is clearly evident in these two figures is that there is a much lower average inbreeding in Nelore cattle using the 50K SNP panel (5.2 and 2.9%) than either the low density (8.7 and 6.0%) or equally spaced 6K (10.0 and 5.4%) SNP panels for F_{ROH4} and F_{ROH8}, respectively. One possible reason may be that the Nelore cattle have more overall homozygosity than any of the other breeds. Then, when reducing the number of SNP from the original 50K to the lower density 6K and attempting to find consecutive SNP that are all homozyogous, it would be much easier to find ROH which are false positives. These false positive ROH may actually contain

heterozygous markers, but since the SNP selection process only included *Bos taurus* breeds, polymorphic SNP within the Nelore breed are masked by SNP which only have a high minor allele frequency in the *Bos taurus* breeds. This is also evident, although not as visually conspicuous, when comparing F_{ROH4} values derived from the equally spaced 6K and low density SNP panels for elite Holsteins (9.2 vs. 6.2%), commercial Holsteins (7.7 vs. 6.0%), Jersey (11.3 vs. 7.9%), and Angus (8.5 vs. 6.9%). This may suggest that the greater F_{PH} in the equally spaced 6K SNP sets results in higher values of F_{ROH4} and F_{ROH8}.

Ignoring the results from the Nelore, since their frequency distribution varies drastically compared to the other genetic groups, the differences between F_{ROH4} and F_{ROH8} were calculated separately for each 50K, low density, and equally spaced 6K SNP panels. Averaging across the genetic groups, the smallest difference (0.6%) was observed when comparing F_{ROH4} and F_{ROH8} using the low density panel, followed by the equally spaced 6K SNP sets (1.1%), with the largest difference found with the 50K SNP panel (1.9%). While F_{ROH4} includes shorter ROH, which are evident of inbreeding further back in the pedigree, these ROH may be more difficult to detect with a lower density SNP panel. The small difference between F_{ROH4} and F_{ROH8} values when discovered using the low density SNP panel suggest that many of the shorter ROH found with the 50K SNP panel are undiscovered by the equally spaced 6K panels, and particularly the low density SNP panel.

CONCLUSIONS

Genomic inbreeding measures were calculated for 5 genetic groups of cattle: elite Holsteins, commercial Holsteins, Jerseys, Angus, and Nelore, using 50K, low density, and equally spaced 6K SNP panels. Nelore cattle had the highest average F_{PH}, but also the lowest average F_{ROH4} and F_{ROH8} values. Jersey cattle had much higher average F_{ROH4} and F_{ROH8} than any

of the other genetic groups. Genetically elite Holsteins had higher F_{PH}, F_{ROH4}, and F_{ROH8} values than commercial Holsteins, suggesting that the intense selection pressure on the genetically elite animals has resulted in greater levels of DNA which is IBD. Ascertainment bias in the selection of the low density SNP panel was shown, as the F_{PH} from the 50K and equally spaced 6K SNP panels tended to be similar for all breeds, while F_{PH} from the low density SNP panel was lower for elite Holsteins, commercial Holsteins, Jersey, and Angus cattle. As genomic information from Nelore cattle did not go into selection of SNP for the low density panel, measures of F_{PH} for all three SNP sets were similar for Nelore. Furthermore, due to the decreased density of the low density SNP panels, the shorter ROH that account for inbreeding further back in a pedigree are more difficult to detect. This may suggest that only recent genomic inbreeding can be detected accurately using lower density SNP panels.

ACKNOWLEDGEMENTS

DWB acknowledges financial support from the USDA National Institute of Food and Agriculture National Needs Graduate Fellowship Grant Number 2010-38420-30477. I would also like to acknowledge the contribution of current and former Zoetis Genetics (Kalamazoo, MI) employees, Natascha Vukasinovic, Prashanth Bohdreddy, Donald Nkrumah, Sue DeNise, Jason Osterstock, and Doug Ricke.

REFERENCES

Bjelland, D.W., K.A. Weigel, N. Vukasinovic, and J. D. Nkrumah. Evaluation of inbreeding depression in Holstein Cattle using SNP markers and alternative measures of genomic inbreeding. J. Dairy Sci. 2013. in press.

Bovine HapMap Consortum, The. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 2009. 324:528-532.

Carothers, A.D., I. Rudan, I. Kolcic, O. Polasek, C. Hayward, A.F. Wright, H. Campbell, P. Teague, N.D. Hastie, and J.L. Weber. Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. Ann. Hum. Genet. 2009, 70,666-676.

Carrillo, J.A., and F. Siewerdt. Consequences of long-term inbreeding accumulation on preweaning traits in a closed nucleas Angus herd. J. Anim. Sci. 2010, 88:87-95.

Faria, F.J.C., A.E.V. Filho, F.E. Madalena, and L.A. Josahkian. Pedigree analysis in the Brazilian Zebu breeds. J. Anim. Breed. Genet. 2009, 126:148-153.

Ferencakovic, M., E. Hamzic, B. Gredler, I. Curik, and J Solkner. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh Cattle. ACS. 2011, 4:325-328.

Fisher, R.A. A fuller theory of “junctions” in inbreeding. Heredity. 1954, 8:187-1973

Howrigan, D.P., M.A. Simonson, and M.C. Keller. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics. 2011, 12:460.

Illumina Inc. 2012. BovineSNP50 Genotyping BeadChip. Accessed Nov. 1, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp5O.pdf

Keller, M.C., P.M. Visscher, and M.E. Goddard. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 2011, 189:237-249.

Kirin, M., R. McQullian, C.S. Franklin, H. Campbell, P.M. McKeigue, et al. Genomic runs of homozygosity record population history and consanguinity. PLos One. 2010, 5(11): e13996. doi:10.1371/journal.phone.0013996.

Li, J.Z., D.M. Absher, H. Tand, A.M. Southwick, A.M. Casto, et al. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008, 319:1100-1104.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, et al. PLINK: a toolset for whole genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007, 81:559-575.

Simon-Sanchez, J., L.L. Kilarski, M.A. Nalls, M. Martinez, C. Schulte, et al. Cooperative genome-wide analysis shows increased homozygosity in early onset Parkinson's Disease. PLoS One. 2012, 7(3): e28787. doi:10.1371/journal.pone.0028787

Smith, L.A., B.G. Cassell, and R.E. Pearson. The effects of inbreeding on the lifetime performance of dairy cattle. J. Dairy Sci. 1998, 81:2729-2737.

Sonesson, A.K., J.A. Wolliams, and T.H.E. Meuwissen. Genomic selection requires genomic control of inbreeding. GSE. 2012, 44:27.

US Department of Agriculture-Animal Improvement Laboratories. 2012 Bovine Inbreeding Trends. Accessed Aug. 30, 2012. http://aipl.arsusda.gov/eval/summary/inbrd.cfm

VanRaden, P.M. Efficient methods to compute genomic predictions. J. Dairy Sci. 2009, 91:4414-4423.

VanRaden, P.M., K.M. Olson, G.R. Wiggans, J.B. Cole, and M.E. Tooker. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J. Dairy Sci. 2011, 94:5673-5682.

Weigel, K.A., P.C. Hoffman, W. Herring, and T.J. Lawlor Jr. Potential gains in lifetime net merit from genomic testing of cows, heifers, and calves on commercial dairy farms. J. Dairy Sci. 2012, 95:2215-2225.

Wiggans, G.R., T.S. Sonstegard, P.M. VanRaden, L.K. Matukumalli, R.D. Schnabel, J.F. Taylor, F.S. Schenkel, and C.P. Van Tassell. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J. Dairy Sci. 2009, 92:3431-3436.

Wiggans, G.R., T.A. Cooper, P.M. VanRaden, K.M. Olson, and M.E. Tooker. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. J. Dairy Sci. 2012, 95:1552-1558.

Zimin A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz., et al. A whole-genome assembly of the domestic cow, Bos Taurus. Genome Biol. 2009, 10:R42.

Table 1. Number of animals and SNP before and after data editing.

	Number of Animals[1]			Number of SNP after edits[2]			Number of SNP after LD pruning[3]		
	Initial	After Edits	Initial # of SNP	50K	Low Density	Equally Spaced 6K	50K	Low Density	Equally Spaced 6K
Commercial									
Holstein	6,600	4,386	54,001	32,045	5,902	6,329	7,780	3,144	3,365
Elite Holstein	7,883	6,510	43,485	37,374	5,556	6,069	7,775	2,378	2,609
Jersey	3,146	2,253	43,485	31,873	5,318	6,375	4,580	1,908	2,098
Angus	2,402	2,337	54,001	38,559	6,417	6,356	7,061	3,049	3,107
Nelore	2,302	2,216	54,001	23,319	4,201	5,597	9,915	3,512	4,378

[1]Animals removed with greater than 10% missing SNP

[2]SNP removed for percent missing (>10%), minor allele frequency (<5%), and violation of Hardy-Weinberg equilibrium (p< 0.0001)

[3]SNP removed for linkage disequilibrium (LD) pruning with an $R^2 > 0.5$ with all other SNP in a 50 SNP window

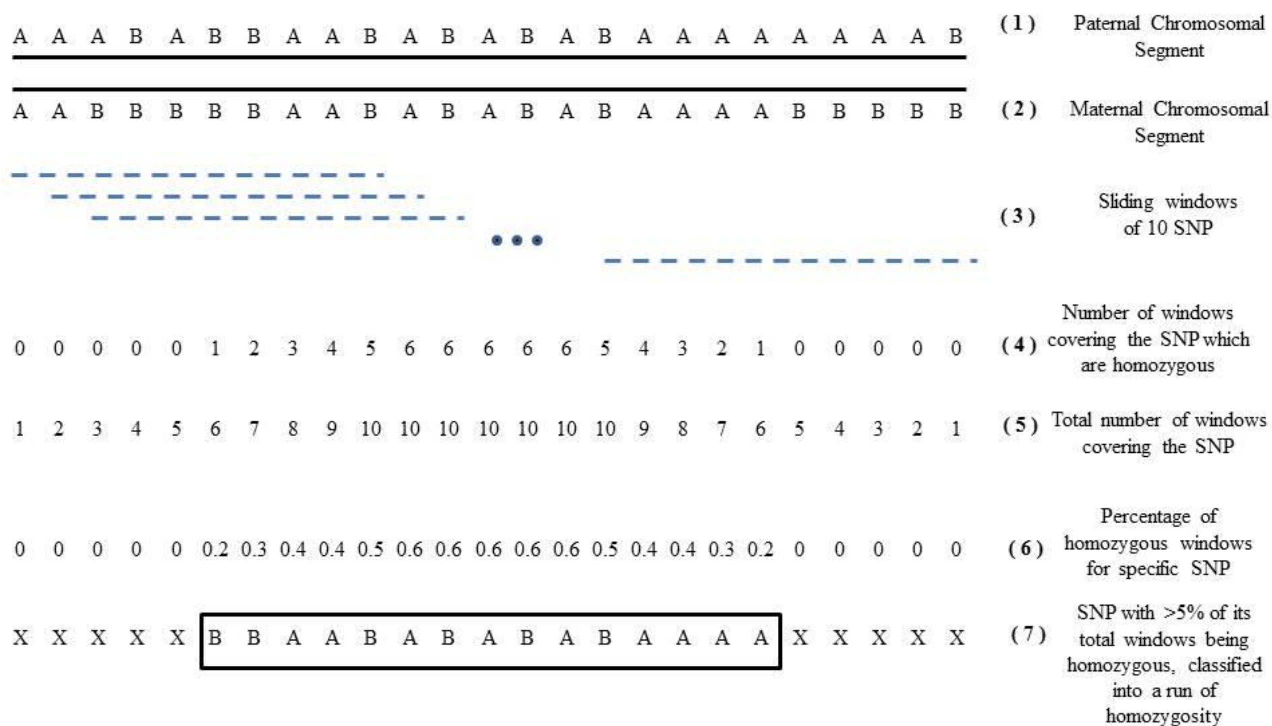


Figure 1. Description of the process for discovery of runs of homozygosity (ROH) using a sliding window of SNP markers along the chromosome, as implemented with PLINK software (Purcell et al., 2007)

Figure 2. Frequency distributions for percent homozygosity, F_{ROH4}, and F_{ROH8} derived from 50K SNP panel.

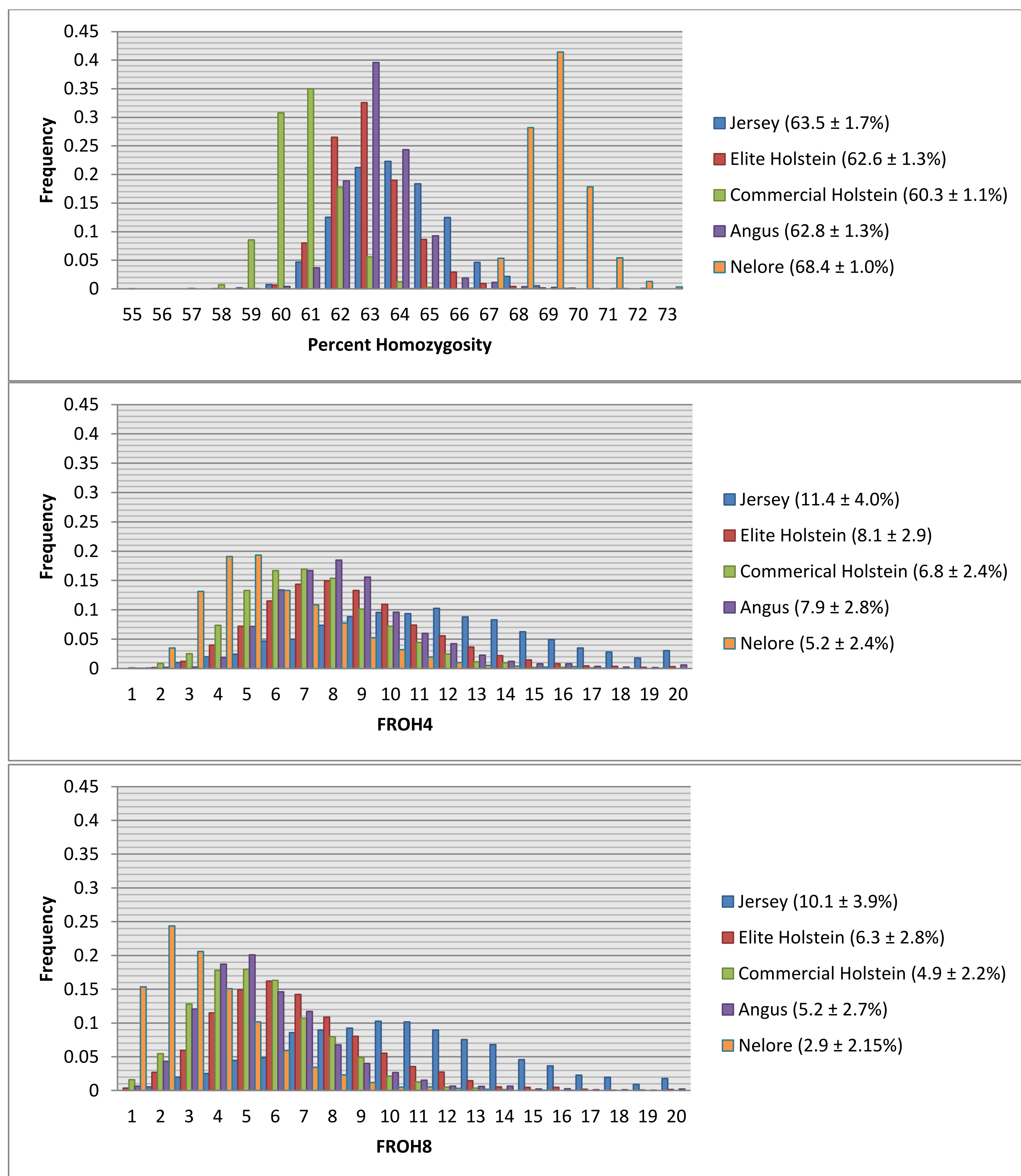


Figure 3. Frequency distributions for percent homozygosity from 50K, equally spaced 6K, and low density SNP panels.

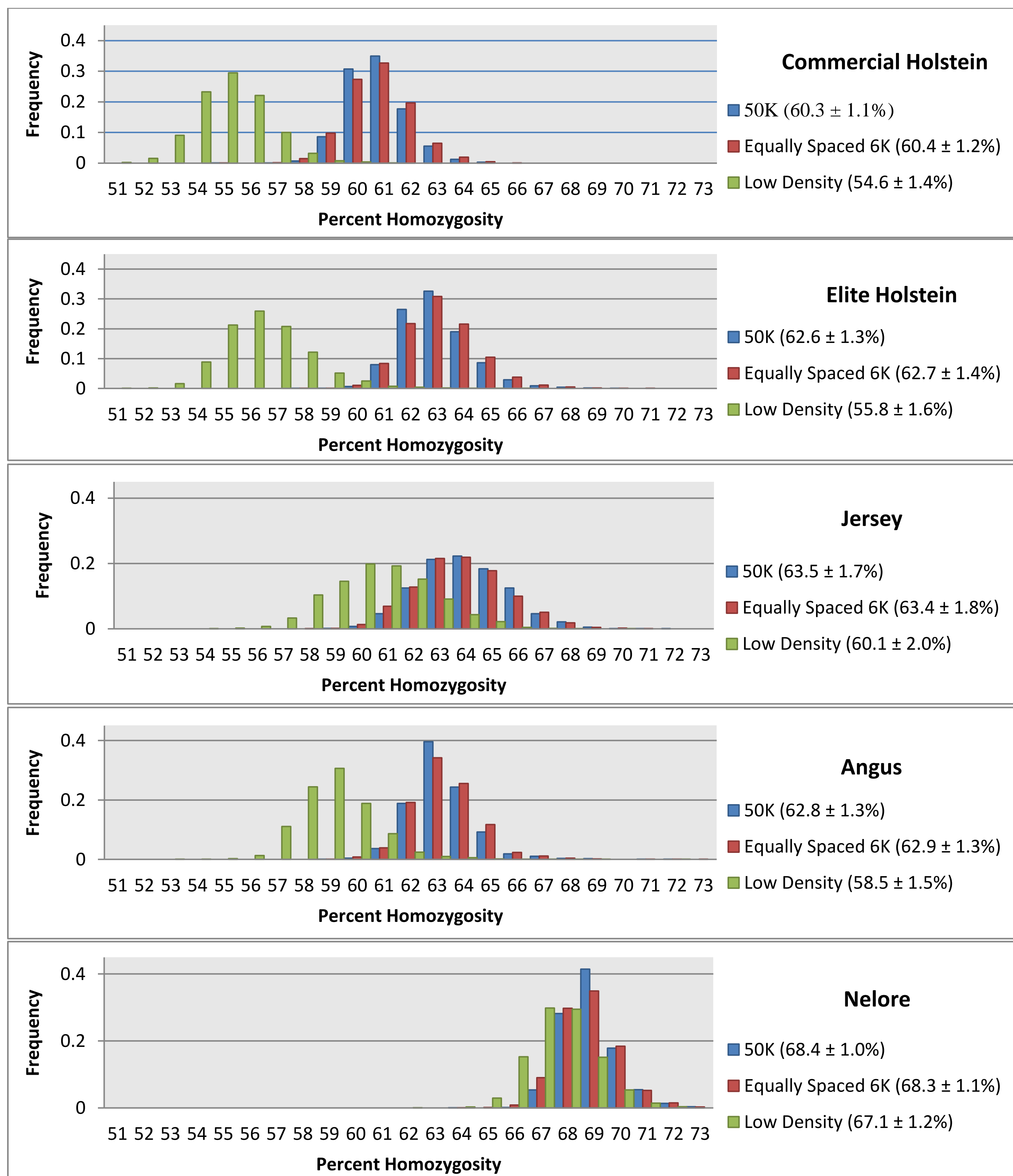


Figure 4. Frequency distributions for F_{ROH4} calculated from 50K, equally spaced 6K, and low density SNP panels.

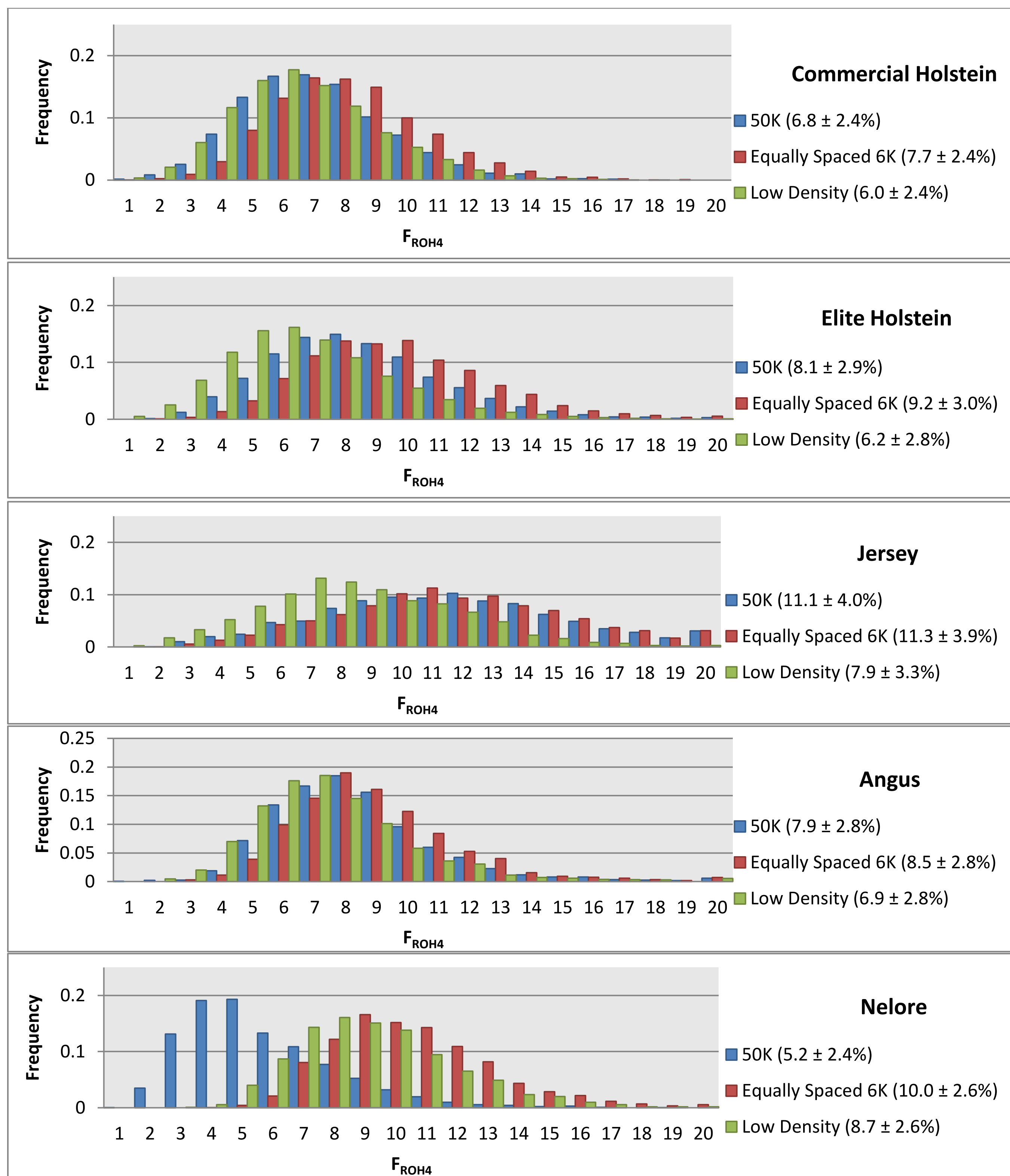


Figure 5. Frequency distributions for F_{ROH8} calculated from 50K, equally spaced 6K, and low density SNP panels.

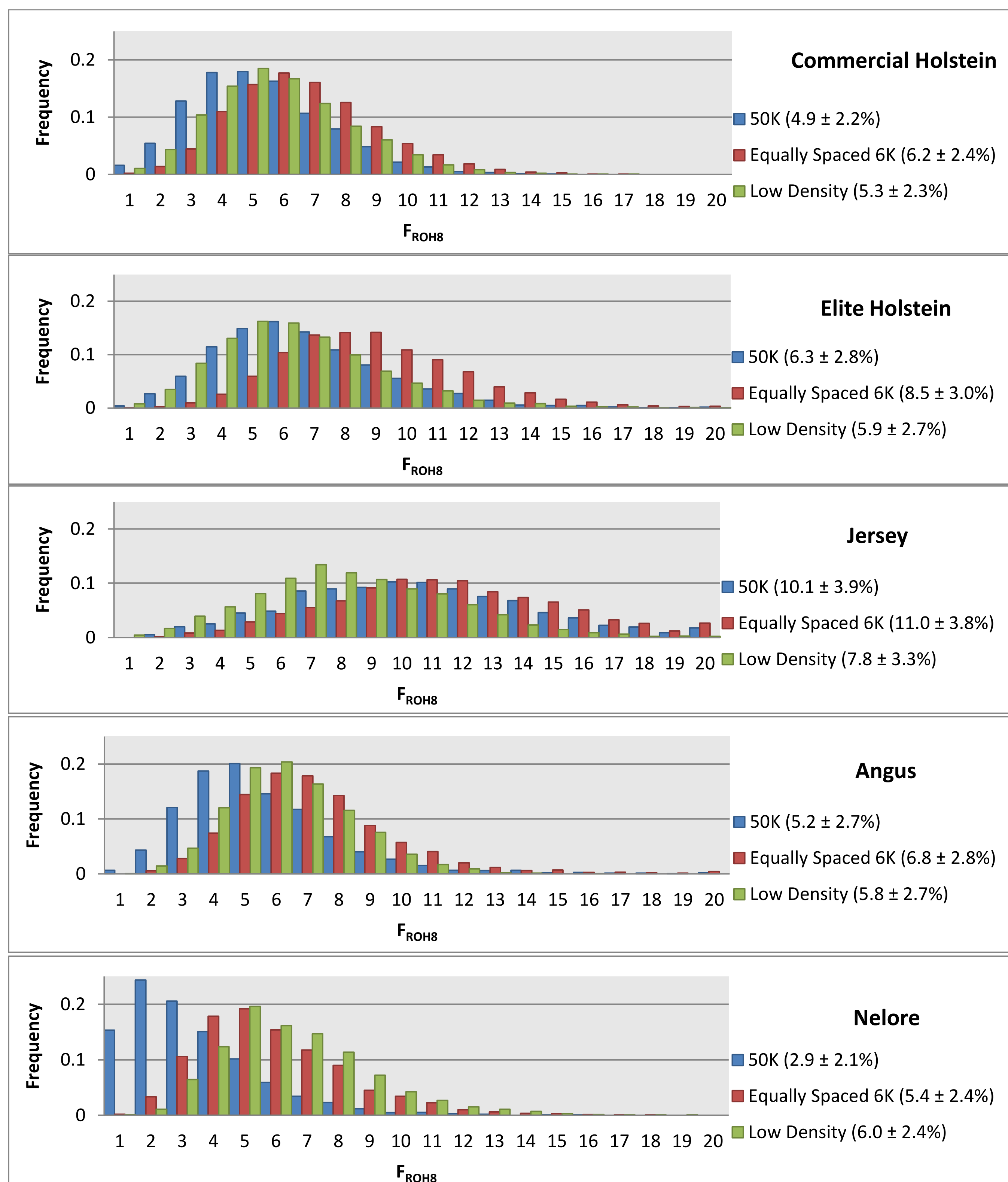
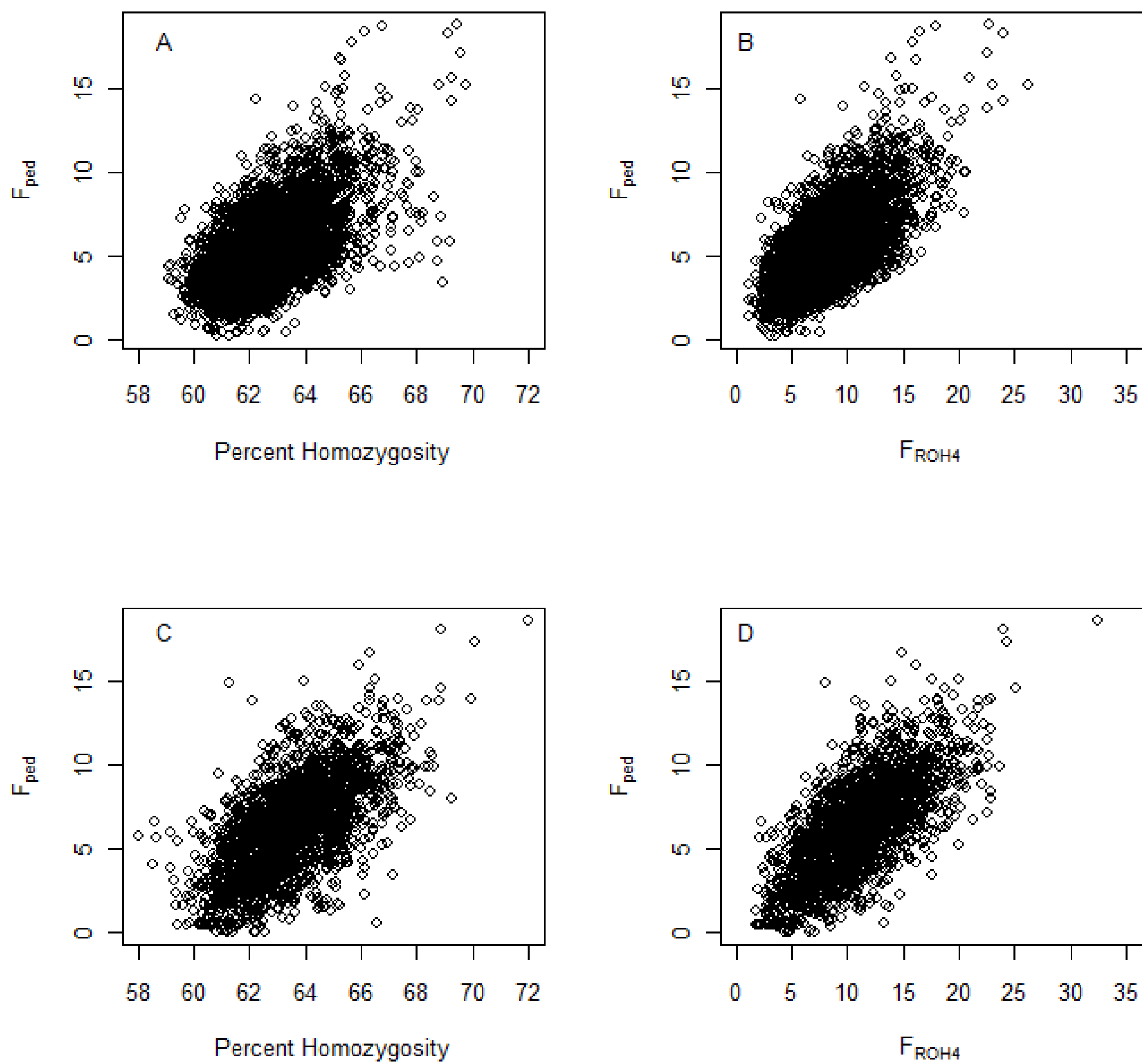


Figure 6. Relationship between F_{ped} and percent homozygosity, and F_{ped} and F_{ROH4} for elite Holstein (A and B, respectively) and Jersey (C and D, respectively) cattle.



CHAPTER 5

Comparison of genomic inbreeding within a family-based structure in Holstein cattle

ABSTRACT

As reproductive ability in dairy cattle has declined, some evidence has suggested embryonic loss may be a major cause. The current study compares expected genomic inbreeding from sire-dam mating pairs to genomic inbreeding from live progeny in an attempt to determine how embryonic inbreeding may impact fertility. A total of 412 sire-dam-progeny trios were available in which all animals had genotypes containing 43,485 SNP markers. After removal of trios due to parentage errors and SNP call rate, 374 remained for analysis. Additionally, 3,906 genotyped full sibling pairs were available for comparison. Expected genomic inbreeding measures were calculated by predicting homozygosity independently per SNP (F_{PHE}) in sire-dam mating pairs and by simulating progeny using phased haplotype information (F_{ROHE} and F_{PHE}). Actual genomic inbreeding measures were calculated using the percent homozygosity of all SNP (F_{PH}) and utilizing runs of homozygosity (F_{ROH}). Average F_{PHE} values (62.8 ± 0.78%) were similar to F_{PH} (63.1 ± 1.12%), when considering each SNP independently. After phasing haplotypes and simulated progeny with an average of 30 crossover events, F_{PHE} (62.5 ± 0.87%) was again similar to F_{PH} (62.7 ± 1.16%), and F_{ROHE} (3.01 ± 1.41) was slightly lower than F_{ROH} (3.53 ± 2.17%). Genomic inbreeding between full siblings was also compared, and while pedigree inbreeding measures between these pairs of animals would be the same, only moderate correlations between genomic inbreeding of one sibling with the other (0.47-0.52) were present. Results suggest increases in genomic inbreeding do not explain a large effect of the viability of the embryo at average levels of expected inbreeding. Higher variation in F_{ROH} and F_{PH} values were present with sire-dam mating pairs exhibiting high F_{ROHE} and F_{PHE}, respectively, which may suggest high levels of genomic inbreeding are required for a noticeable effect on overall

embryo viability. Overall, results did not suggest a large impact of expected inbreeding on embryo viability.

INTRODUCTION

Dairy cattle populations have seen a decline in reproductive ability over the past several decades. Washburn et al. (2002) analyzed reproductive traits in Holstein and Jersey cows of the southeastern US and reported an increase in days open from an average of 122 ± 2.8 d in 1978 to 152 ± 2.8 d in 1999 for Jerseys and from an average of 124 ± 0.7 d in 1978 to 168 ± 0.7 d in 1999 for Holstein cows. Similar negative trends were also seen in traits such as services per conception, days to first service, and estrus detection rate. Norman et al. (2009) analyzed reproductive trends in US Holstein and Jersey cows from 1996 to 2006 and found similar results. Average conception rate decreased from 33% in 1996 to 30% in 2006 for Holstein cows, with a low in 2001 of 26%, while a decrease from 39% in 1996 to 35% in 2006 for Jersey cows was observed, with a low of 30% in 2001. Between 1996 and 2006, number of breedings per lactation also increased for Holsteins (2.1 to 2.5 services) and Jerseys (2.0 to 2.3 services), as did the average calving interval for Holsteins (410 to 422 d) and Jerseys (398 to 410 d).

Failures in reproduction can be caused by many factors, such as increases in inbreeding, poor estrus detection, anestrus or abnormal luteal phases in high producing dairy cows, or low concentrations of key reproductive hormones such as progesterone and IGF-I (Lucy, 2001). Another issue that affects fertility is embryonic death. Moreira et al. (2001) compared pregnancies at days 27 and 45 of gestation in 139 Holstein cows and discovered a loss of 20.7% of the pregnancies. A similar analysis was performed by Cartmill et al. (2001) comparing pregnancies at day 28 with pregnancies at 38 through 58 d of gestation and a loss of 28% of the pregnancies was discovered in 128 Holstein cows. Chebel et al. (2004) performed a more

extensive study of 1,465 Holstein cows, comparing pregnancies at days 31 and 45 of gestation, with a total loss of 12.5% of the pregnancies. The causes of early embryonic loss are sometimes unknown, but may be due to increases in inbreeding. If inbreeding of the dam is high, this could lead to problems with the maternal recognition and maintenance of pregnancy, and lead to some of the negative results seen in inbreeding studies with regards to reproductive traits (Smith et al., 1998; Mc Parland et al., 2007). If inbreeding of the embryo is high, there is a greater chance that deleterious lethal disorders are present, such as BLAD (Kehrli et al., 1990) or DUMPS (Shanks et al., 1984), one of the recessive deleterious haplotypes discovered by VanRaden et al. (2011), or even the accumulation or interaction of genes with small negative effects on fertility (Khatib et al., 2009).

With the use of tools derived from whole genome sequencing, such as the Illumina Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA), methods to quantify inbreeding on a genomic scale have been developed (Keller et al., 2011). Bjelland et al. (2013) analyzed the effects of three measures of genomic inbreeding, inbreeding derived from runs of homozygosity (F_{ROH}), inbreeding derived from a genomic relationship matrix (F_{GRM}), and the overall percent homozygosity of the genome (F_{PH}), and discovered negative effects on both milk production and reproductive traits with increases in measures of genomic inbreeding. Increases in days open ranged from 1.06 to 1.76 d per 1% increase in inbreeding for the three measures of genomic inbreeding. Other studies in domestic animal species are lacking, but increases in both F_{ROH} and F_{PH} in human populations have been correlated with higher risk of disease (Simon-Sanchez et al., 2012; Keller et al., 2012) and decreases in fitness of quantitative traits (McQuillan et al., 2012).

The study herein attempts to determine whether the genomic inbreeding measures of progeny are similar when compared with expectations derived from the parents. Deviations from

expectations may suggest that embryonic inbreeding has a large effect on the survival of that embryo.

MATERIALS AND METHODS

Data

Data were provided by Genex Cooperative/CRI (Shawano, WI) and consisted of 54,001 SNP markers from a total of 3,601 Holstein cattle. Genotypes for 7,883 genetically elite Holstein cattle were also provided by USDA-ARS Animal Improvement Programs Laboratory (Beltsville, MD) and consisted of 43,485 SNP markers throughout the 29 *Bos taurus* autosomes and the X chromosome. These SNP represent the subset of markers on the Illumina BovineSNP50 BeadChip that are used for routine genetic evaluation of US dairy cattle, after removal of SNP with a call rate of $<90\%$, greater than 1% parent-progeny conflicts, complete linkage disequilibrium with an adjacent SNP, or minor allele frequency of $<1\%$ in each of the Holstein, Jersey, and Brown Swiss breeds (Wiggans et al., 2009). As the animals provided in the Genex dataset contained mored familial information (relatively equal number of cows and bulls) and the USDA dataset contained mainly genetically elite sires, the USDA data were utilized to fill in missing gaps in the sire-dam-progeny trios rather than provide unique information. SNP from the Genex dataset were reduced to the 43,485 SNP available in the USDA-ARS data. Further editing of SNP was performed on the complete dataset with the removal of SNP based on minor allele frequency ($MAF < 0.05$), call rate (percent missing > 0.10), and violation of Hardy-Weinberg equilibrium ($p < 0.0001$). Individual animals with missing SNP greater than 10% were also removed from the analysis.

Pedigree information was provided by both Genex Cooperative and the USDA-ARS Animal Improvement Programs Laboratory, and consisted of sire, dam, and progeny

information. A total of 412 sire-dam-progeny trios were available for analysis, with 374 sire-dam-progeny trios remaining after removal of trios due to greater than 5% parent-progeny conflicts. Additionally, 3,031 genotyped Holsteins within full sibling families, comprised of a total of 3,906 full sibling pairs, were available for analysis, with some of the full sibling families having a large number of members. When forming the full sibling pairs from a family of 4 full siblings (A, B, C, and D), for example, 6 different combinations were available for comparison (AB, AC, AD, BC, BD, and CD). This explains the discrepancy between the total number of animals in this part of the study and the total number of full sibling pairs.

Simple Method

The first method utilized to determine expected inbreeding from the mating pair treats each SNP independently and determines the probability that the SNP in the progeny will be homozygous. The probabilities are determined using the information given in Table 1. For example, if a sire has the genotype at a given SNP of AA, and the dam has the genotype at that SNP of AA, the progeny of that mating pair will have a homozygous genotype (AA) at that SNP with a probability equal to 1. Conversely, if at a given SNP, the sire had the genotype AB and the dam had the genotype AB, four possible genotypes are possible in the progeny: AA, AB, BA, and BB. Two of the 4 possible genotypes are homozygous, so the probability of the progeny being homozygous at that SNP is 0.5. The total expected homozygosity (F_{PHE}) from the sire-dam mating pair is then calculated with the following formula,

$$F_{PHE} = \frac{\sum_i^m P(Hom_i)}{m}$$

where m = number of non-missing markers and $P(Hom_i)$ = probability of producing a homozygous marker at locus *i*, based on the probabilities in Table 1. The actual homozygosity of the progeny is then calculated as,

$$F_{PH} = \frac{N_{AA} + N_{BB}}{N_{AA} + N_{AB} + N_{BB}}$$

where N_{AA}, N_{AB}, and N_{BB} are the number of SNP which are classified as AA, AB, and BB, respectively. For each sire-dam-progeny trio, F_{PHE} and F_{PH} were then compared using a Chi squared test.

Phased Haplotype Method

The second method utilized to determine expected inbreeding simulated mating pairs from phased haplotype data. Haplotypes were phased using the hidden Markov model methods developed in BEAGLE 3.0 (Browning and Browning, 2009). This method employed the sire-dam-progeny trio information, as well as the population data, to infer the haplotypes. Initially, missing genotypes are imputed based on allele frequency and random phasing in heterozygotes. Then, this algorithm alternates between model building and sampling. In the model building step, current estimates for each haplotype are utilized in building a new hidden Markov model. Then in the sampling step, new haplotypes are sampled for each sire-dam-progeny trio based upon the genotypic data and the current hidden Markov model. Once the haplotypes were phased, 250 possible progeny were simulated for each mating and average recombination rate. With no recombination events, for each chromosome, one of the two chromosomal haplotypes were selected at random for each parent. When selections were made for each chromosome, they were combined to form the genomes of the expected progeny. With 4 possible combinations per chromosome, a total of $4^{29} = 2.8 \times 10^{17}$ possible combinations were possible with each mating pair using this method.

The method which included recombination events in the simulations was performed in a similar manner. Simulations including 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 150, 200, 250, and 500 average recombination events per genome were performed. For each chromosome,

independently, 6 simulated gametes were created for each parent. These simulated gametes were produced by randomly starting at one of the possible haplotypes, then after 10 SNP, a probability corresponding to the average recombination rate used in this simulation was given as to whether a crossover would occur at this specific location. If a crossover occurred, the next SNP would read from the opposite haplotype. This process would continue every 10 SNP with the possibility of a crossover. Once each of the 6 gametes for each chromosome were formed for each parent, one was randomly selected from each parent, independently for each chromosome, and assembled into the 250 simulated progeny for that specific recombination rate. For this simulation, after the potential gametes were formed, $(6*6)^{29} = 1.4 \times 10^{45}$ possible progeny could be simulated from the 6 potential gametes for each parent.

Once the simulated progeny were produced, two separate methods were utilized to produce expected inbreeding coefficients. The first calculated both the F_{PH} and F_{PHE} utilizing the same formula used to calculate F_{PH} in the previous analysis. The second method utilized PLINK (Purcell et al., 2007) to determine runs of homozygosity (ROH) for each possible and real progeny. A ROH is essentially an extended haplotype in which all markers contained within a given section are homozygous. As DNA is passed from generation to generation, large sections of DNA are inherited together, rather than single SNP or markers. If inbreeding occurs, the large sections of DNA which have been passed from the common ancestor to the related mating pair come together in their progeny to form a ROH. The input parameters used in determining ROH were derived from a simulation study by Howrigan et al. (2011) in which 30 SNP were used as minimum length for the ROH and no heterozygote SNP were allowed within the ROH. LD pruning was also performed prior to ROH discovery, with all SNP having an $R^2 > 0.5$ with all other SNP in a 50 SNP window being removed. This was performed on the SNP and population

set prior to phasing in order to eliminate any bias the phasing process may produce. A total of 6,452 SNP remained for ROH analysis after LD pruning. As these data were phased and missing SNP were derived from family and population parameters, no missing SNP were present in the data. After ROH were discovered for each real and simulated progeny, the inbreeding measure was calculated using the formula,

$$F_{ROH} \text{ or } F_{ROHE} = \frac{\sum_k length(ROH_k)}{L}$$

where k = number of ROH discovered for each animal, and L = total length of the genome. Length of ROH was measured in kilobases (kb) with L = 2,612,820 kb (Zimin et al., 2009). The 250 F_{PHE} and F_{ROHE} from the simulated progeny for each mating pair were then averaged to provide a coefficient to compare to F_{PH} and F_{ROH} from the actual progeny, respectively. Comparisons were made between F_{PH} and F_{PHE} and between F_{ROH} and F_{ROHE} using Chi-squared tests.

Full Sibling Analysis

In addition to the sire-dam-progeny trios, a total of 3,906 full siblings pairs in which both siblings had genotype information were available for analysis. Two genomic measures of inbreeding previously outlined, F_{ROH} and F_{PH}, were calculated for each of the animals. Comparisons between each of the sibling pairs, and among full sibling families, were made to determine the variability present among the genomic measures of inbreeding. As the full siblings would have the exact same measures of pedigree inbreeding (F_{ped}) any differences between the siblings would be due to Mendelian sampling and would only be able to be determined utilizing genomic marker information.

RESULTS AND DISCUSSION

Expected versus Actual Genomic Inbreeding

Means of actual and expected genomic inbreeding measures are presented in Table 2. An average of 30 crossover events during simulations was utilized in creating the progeny for the results presented in Table 2. This was similar to the average recombination rates found in Angus and Limousin cattle presented by Weng et al. (2013).When using the simple method with SNP treated independently, F_{PHE} (62.8 ± 0.78%) was similar to F_{PH} (63.1 ± 1.12%). Similar results were present after simulation of progeny from phased haplotypes, where F_{PHE} (62.5 ± 0.87%) was similar to F_{PH} (62.7 ± 1.16%) and F_{ROHE} (3.01 ± 1.41%) was slightly lower than F_{ROH} (3.53 ± 2.17%). Differences in the values of F_{PH} are present because for the simple method, calculations were made with unphased genotypes while phased haplotypes were utilized in determining the coefficients for the simulated data. Missing SNP were classified and some inconsistencies were corrected during the phasing process which led to the difference in inbreeding measures. Correlations between F_{PH} (0.91) and F_{ROH} (0.97) calculated before and after haplotype phasing were high, indicating that the phasing method did not drastically affect these measures of inbreeding.

The results of study do not suggest that expected inbreeding of progeny has a large effect on the viability of the embryos at all severities of expected inbreeding. Results of studies such as the deleterious haplotypes discovered by VanRaden et al. (2011), do not seem to present themselves in a large effect with respect to embryonic inbreeding. Although with the sample size present in this study and the carrier rate of the deleterious haplotypes presented by VanRaden et al. (2011) of less than 5%, a large effect may not be possible to detect. Studies which have focused on inbreeding, using both pedigree (Smith et al., 1998) and genomic measures (Bjelland et al., 2013) have found large negative effects of increases in maternal inbreeding on reproductive ability. Results here may suggest that other than avoiding the large deleterious

effects, overall reproductive ability and viability of the embryo may be more affected by increases in maternal inbreeding rather than embryonic inbreeding.

Further visualizations of F_{PHE} compared with F_{PH} using the simple method are presented in Figure 1. A correlation between the two measures of 0.70 is shown in Figure 1A, suggesting that predicting the inbreeding of progeny using this method is acceptable and could be utilized in mate selection programs. Histograms including both F_{PHE} and F_{PH} present in Figure 1B demonstrate the overall mean and distribution of the two measures. Figure 2 and Figure 3 present similar comparisons between F_{PHE} and F_{PH} and between F_{ROHE} and F_{ROH}, respectively, when discovered using simulated progeny. Correlations between F_{PHE} and F_{PH} (0.68) and between F_{ROHE} and F_{ROH} (0.68) when calculated from progeny simulated with an average of 30 crossover events were also moderately high, suggesting that accurate prediction of genomic inbreeding in progeny can be made by utilizing phased haplotypes of potential mates. Figures 2A and 3A suggest possible differences in variation when comparing F_{PHE} and F_{ROHE} values at different severities. To determine this, data were ordered from least to greatest for both measures of genomic inbreeding separately and split into a low, median, and high predicted genomic inbreeding values. Standard deviations of F_{PH} varied greatly when the data were split into the three subsets, whereas the lowest F_{PHE} group had a standard deviation for F_{PH} of 0.89%, the median group had a standard deviation of 0.79%, and the group with the highest expected inbreeding had a standard deviation of 1.09% for F_{PH}. The large variation in the high F_{PHE} group may correlate to a negative effect of the increases in predicted inbreeding on embryo viability. Similar results are present when creating the three datasets for F_{ROHE}, with standard deviations of 1.27, 1.57, and 2.45%, for F_{ROH} in the low, median, and high F_{ROHE} subsets, respectively. The mean for F_{ROHE} (5.12%) in the highest group is also the only instance in which the expected

inbreeding was higher than the actual (F_{ROH} = 5.05%), although this difference was not significant. Overall, these results may suggest that at very high levels of expected inbreeding may cause a slight decrease viability of the embryo due to an accumulation of small negative effects on fertility, such as those discovered by Khatib et al. (2008), although more work may be needed to accurately assess this hypothesis. These result may also be simply the fact that if there is a possibility of large amount of genomic inbreeding (the mating pair are closely related), there is also a chance that the progeny born simply did not receive many of the common alleles. It is difficult to determine whether the variation is due to the highly inbred animals not surviving to term or due to the variation in Mendelian sampling.

Effects of variation in recombination rate

Effects of changes in recombination rate during progeny simulation were also examined and are presented in Figure 4. The horizontal lines present for F_{PH} and F_{ROH} represent the inbreeding coefficients from the actual progeny and were used to compare against all possible average recombination events. As the average number of recombination events increased, no effect was observed on the average F_{PHE} of the simulated progeny in Figure 4A. This would be expected since increasing the number of crossovers should be independent of which SNP were selected to be passed to progeny. This also suggests that the F_{PHE} values calculated using the simple method, and discussed previously, should provide adequate estimates of F_{PH} in possible progeny.

In contrast, the F_{ROHE} values simulated from progeny were highly impacted by increases in the average recombination rate. This is also expected, as increasing the average recombination rate would be analogous to increasing the distance to a common ancestor in this individual's pedigree. For example, if the actual average recombination rate were 30 recombination events

per meiosis, using an average of 60 recombination events would essentially create 2 meioses events and make the parents of the current individual into its grandparents, with respect to the number of recombination events. In this simulation, the recombination events are the only effects present which can break up the potential ROH, but there are no possible events, such as other inbred matings, which can create new ROH. So, as shown in Figure 4B, as the average number of recombination events increases, the average F_{ROHE} values decrease.

Correlations between F_{PH} and F_{PHE}, as well as F_{ROH} and F_{ROHE} when simulated with varying rates of recombination are presented in Figure 4C. Fluctuations when the average number of recombination events was less than 100 are most likely due to the relatively small (n = 250) number progeny which were simulated. The correlations between F_{PH} and F_{PHE} are largely unaffected by increases in recombination rate, with a slightly higher correlation present with an average of 500 recombination events. This slight increase may also be present due to the number of simulated progeny, as other simulations (not shown here) also had correlations ranging from 0.67 to 0.71. The correlation between F_{ROH} and F_{ROHE} decreases drastically once the average number of recombination events is greater than 100. As the increase in the recombination events breaks up many of the ROH, the majority of the predicted progeny have F_{ROHE} values much lower and much less accurate than when lower recombination rates are utilized.

Full Sibling Pairs

Genomic inbreeding measures for the 3,031 animals with full siblings ranged from 0 to 14.1% for F_{ROH}, with a mean of 3.1 ± 1.9%, and from 57.7 to 67.2% for F_{PH}, with a mean of 62.6 ± 1.1%. The two genomic inbreeding measures were moderately correlated, with an R^2 of 0.76. Scatterplots with one full sibling plotted against the other full sibling for the 3,906 full sibling pairs are presented in Figure 5A for F_{ROH} and 5C for F_{PH}, with histograms of the absolute value

of differences between siblings presented in 5B for F_{ROH} and 5D for F_{PH}. Moderate correlations between siblings were present for both F_{ROH} ($R^2 = 0.51$) and F_{PH} ($R^2 = 0.47$). As each of the full siblings would have the same F_{ped} value ($R^2 = 1.0$ when comparing full siblings), the difference shown in F_{PH} and F_{ROH} are differences which are undetectable using traditional measures of pedigree inbreeding. Measuring inbreeding using genomic markers provides more accurate information than simply using pedigrees. The majority of differences between full siblings for both measures of genomic inbreeding were low, but differences of up to 11.4% were present for F_{ROH} and up to 4.8% in F_{PH}. The mean difference for F_{ROH} was 0.9 ± 0.7%, while the mean difference for F_{PH} was 1.6 ± 1.3%.

A separate analysis was performed only looking the least and most inbred animals. The first subset included all sibling pairs (n = 898) in which one of the siblings had an F_{ROH} value less than 1.0%. The second subset included sibling pairs (n = 363) in which one of the siblings had an F_{ROH} value between 6.0 and 7.0%. The average difference for the least inbred animals was 1.7 ± 1.3%, while the average difference between the most inbred animals was 2.4 ± 1.6%. The higher mean and variation present in the most inbred group may suggest that if a sire and dam are closely related, there is a high probability of their progeny being highly inbred, but also a chance of their progeny receiving only a small proportion of DNA which is inbred. In fact, one sibling in the highly inbred subset had and F_{ROH} value of 0, while 52 out of the 362 (14.3%) total siblings had F_{ROH} values less than 2.5%. Conversely for the least inbred subset, only 57 out of the 897 (6.3%) total siblings had an F_{ROH} greater than 4.5%. Carothers et al. (2006) previously presented similar results on the variability with respect to F_{ped}. The product of a first cousin mating would have an F_{ped} of 6.25% with a standard deviation of 2.43%. For a product of double cousins, which would still have an F_{ped} of 6.25% but would have more chances for the

recombination events to break up the inbred segments, the standard deviation would only be 2.11%. The product of a mating with first cousins once removed has an F_{ped} of 3.125% with a standard deviation of 1.6%, while the product of a second cousin mating, with an F_{ped} of 1.56%, has a standard deviation of 1.1%. As the results presented in the current study and those presented by Carothers et al. (2006) suggest, as the common ancestor in a pedigree is further removed from the current individuals, both the inbreeding coefficients and the variation in the inbreeding coefficients decrease. Overall these results suggest that if a mating is closely related, there is a chance that its progeny may receive a large or small amount of DNA which is inbred, but if the mating pair is not closely related there is very little chance to receive a large proportion of DNA which is inbred.

CONCLUSIONS

Expected genomic inbreeding measures from sire-dam mating pairs were compared against genomic inbreeding values from actual live progeny to determine if embryonic inbreeding may have an influence on fertility in dairy cattle. Two methods were used to calculate expected genomic inbreeding: the first determined possible homozygosity based on single SNP independently and the other based on simulating progeny from phased haplotypes. Slight increases in the actual genomic inbreeding when compared to the expected were observed, which is the opposite of what one would expect if genomic inbreeding had a large impact on embryo survival, although high expected levels of genomic inbreeding showed more variability and some evidence of lower actual genomic inbreeding measures. Genomic inbreeding measures were also compared between full siblings, with only moderate correlations present, suggesting genomic measures of inbreeding are required to provide a more accurate measure of relatedness than simply using pedigrees.

ACKNOWLEDGEMENTS

D. W. Bjelland acknowledges financial support from the USDA National Institute of Food and Agriculture National Needs Graduate Fellowship Grant Number 2010-38420-30477. I would also like to thank the employees of Genex Cooperative/CRI (Shawano, WI), Angie Coburn, Glenn Gilbert, Roy Wilson, Keith Heikes, Gamal Azim, and Ami Lasecki, for their cooperation.

REFERENCES

Bjelland, D. W., K. A. Weigel, N. Vukasinovic, J. D. Nkrumah. 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J. Dairy. Sci. 96:4697-4706.

Browning, B. L. and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210-223

Carothers, A. D., I. Rudan, I. Kolcic, O. Polasek, C. Hayward, A. F. Wright, H. Campbell, P. Teague, N. D. Hastie, and J. L. Weber. 2006. Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. Ann. Hum. Genet. 70:666–676.

Howrigan, D.P., M.A. Simonson, and M.C. Keller. 2011. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. BMC Genomics. 12:460.

Kehrli M.E., F.C. Schmalstieg, D.C. Anderson, M.J. Van Der Maaten, B.J. Hughes, M.R. Ackermann, C.L. Wilhelmsen, G.B. Brown, M.G. Stevens, and C.A. Whetstone. 1990. Molecular definition of the bovine granulocytopathy syndrome: identification of deficiency of the Mac-1 (CD11b/CD18) glycoprotein. Am. J. Vet. Res. 11:1826–1836.

Khatib, H., W. Huang, X. Wang, A. H. Tran, A. B. Bindrim, V. Schutzkus, and R. L. Monson. 2009. Single gene and gene interaction effects on fertilization and embryonic survival rates in cattle. J. Dairy Sci. 92:2238-2247.

Keller, M. C., M. A. Simonson, S. Ripke, B. M. Neale, P. V. Gejman, et al. 2012. Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. PLoS Genet. 8(4): e1002656. doi:10.1371/journal.pgen.1002656.

Keller, M.C., P.M. Visscher, and M.E. Goddard. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189:237-249.

McQuillan, R., N. Eklund, N. Pirastu, M. Kuningas, B. P. McEvoy, et al. 2012. Evidence of Inbreeding Depression on Human Height. PLoS Genet. 8(7): e1002655. doi:10.1371/journal.pgen.1002655

Mc Parland, S., J.F. Kearney, M. Rath, and D.P. Berry. 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. J. Dairy Sci. 90:441-4419.

Norman, H. D., J. R. Wright, S. M. Hubbard, R. H. Miller, and J. L. Hutchinson. 2009. Reproductive status of Holstein and Jersey cows in the United States. J. Dairy Sci. 92:3517-3528.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, et al. 2007. PLINK: a toolset for whole genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.

Shanks, R. D., D. B. Dombrowski, G. W. Harpestad, and J. L. Robinson. 1984. Inheritance of UMP synthase in dairy cattle. J. Hered. 75:337-340.

Simon-Sanchez, J., L.L. Kilarski, M.A. Nalls, M. Martinez, C. Schulte, et al. 2012. Cooperative genome-wide analysis shows increased homozygosity in early onset Parkinson's Disease. PLoS One. 7(3): e28787. doi:10.1371/journal.pone.0028787

Smith, L.A., B.G. Cassell, and R.E. Pearson. 1998. The effects of inbreeding on the lifetime performance of dairy cattle. J. Dairy Sci. 81:2729-2737.

VanRaden, P.M., K.M. Olson, D.J. Null, and J.L. Hutchison. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy. Sci. 94:6153-6161.

Washburn, S. P., W. J. Silvia, C. H. Brown, B. T. McDaniel, and A. J. McAllister. 2002. Trends in reproductive performance in Southeastern Holstein and Jersey DHI Herds. J. Dairy Sci. 85:244-251.

Weng, Z. Q., M. Saatchi, R. Schnabel, J. Taylor, and D. Garrick. 2013, Abstract. Factors associated with recombination in beef cattle. ADSA Nat. Conv. Indianapolis, IN.

Zimin A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz., et al. 2009. A whole-genome assembly of the domestic cow, Bos Taurus. Genome Biol. 10:R42.

Table 1. All possible genotypes in progeny for given sire and dam genotypes, with probability the genotype is homozygous.

Dam Genotypes	Sire Genotypes AA	Sire Genotypes AB	Sire Genotypes BB
AA	AA (1)	AA, AB (0.5)	AB (0)
AB	AA, AB (0.5)	AA, AB, BA, BB (0.5)	AB, BB (0.5)
BB	AB (0)	AB, BB (0.5)	BB (1)

Table 2. F_{PHE} and F_{PH} values calculated using the simple method and F_{PHE}, F_{PH}, F_{ROHE}, and F_{ROH} values calculated using simulated progeny with 30 crossover events

Item	Mean	SD
Independent SNP Method		
F_{PHE}	62.83	0.78
F_{PH}	63.14	1.12
Difference	-0.31	
Simulated Progeny		
F_{PHE}	62.49	0.87
F_{PH}	62.67	1.16
Difference	-0.18	
F_{ROHE}	3.01	1.41
F_{ROH}	3.53	2.17
Difference	-0.52*	

* $P < 0.1$

Figure 1. F_{PHE} predicted using the simple method compared to actual F_{PH} of live progeny (A) and a histogram (B) with F_{PHE} predicted using the simple method (light bars) and F_{PH} from actual live progeny (dark bars).

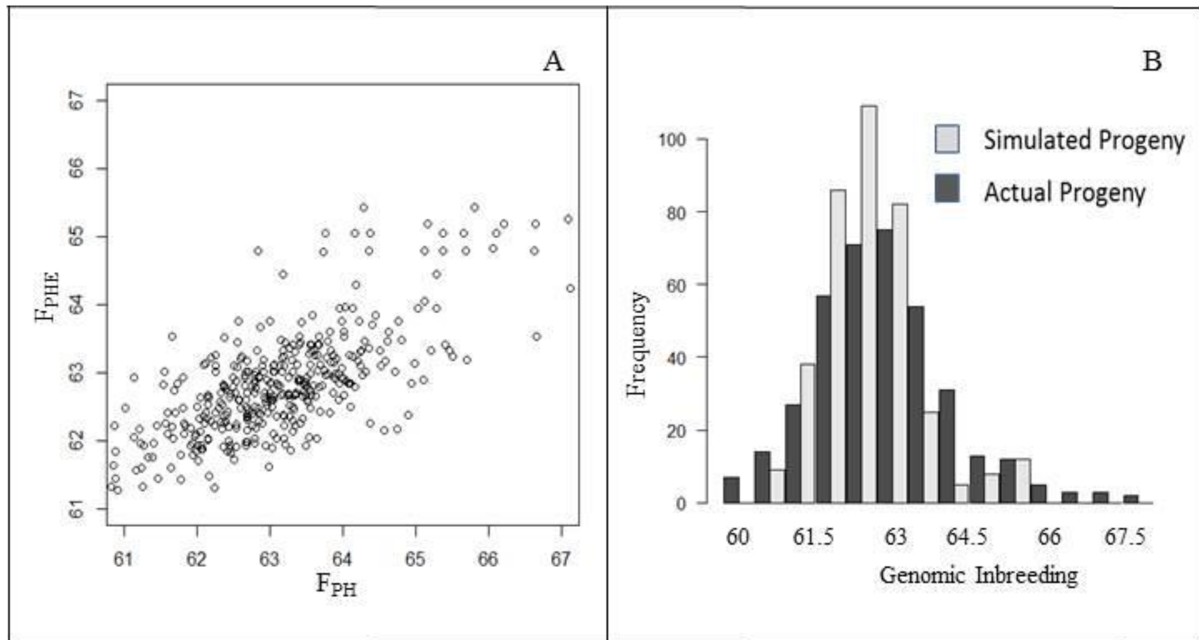


Figure 2. F_{PHE} predicted using simulated progeny with an average of 30 crossover events to actual F_{PH} of live progeny (A) and a histogram (B) with F_{PHE} predicted using the simulated progeny (light bars) and F_{PH} from actual live progeny (dark bars).

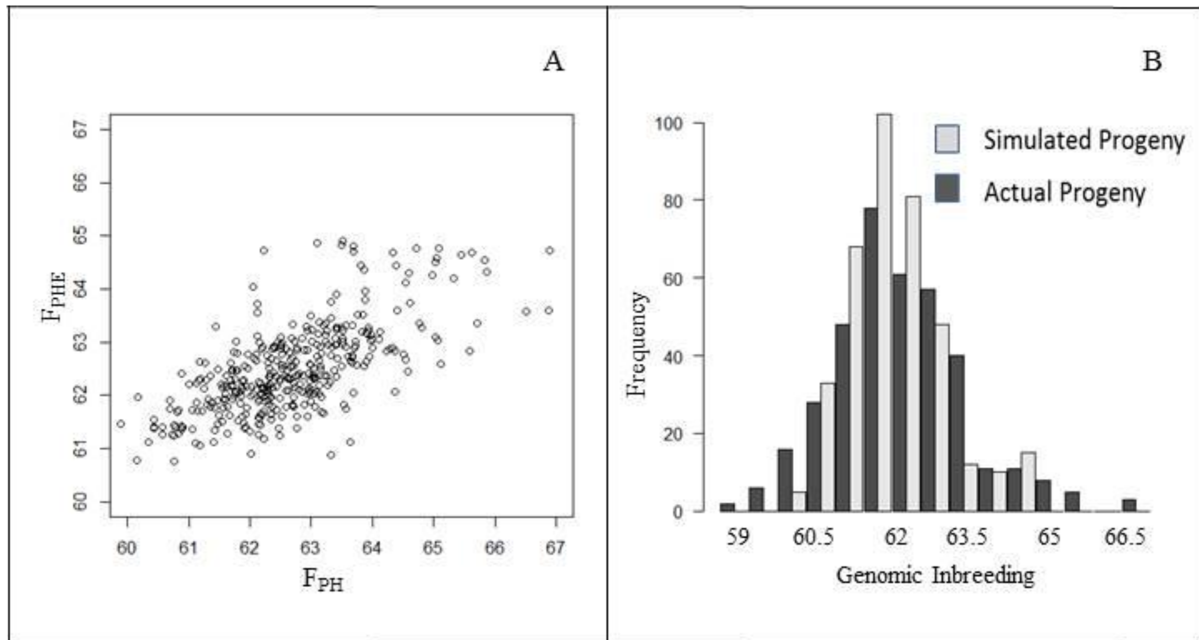


Figure 3. F_{ROHE} predicted using the simulated progeny with an average of 30 crossover events compared to actual F_{ROH} of live progeny (A) and a histogram (B) with F_{ROHE} predicted using the simulated progeny (light bars) and F_{ROH} from actual live progeny (dark bars).

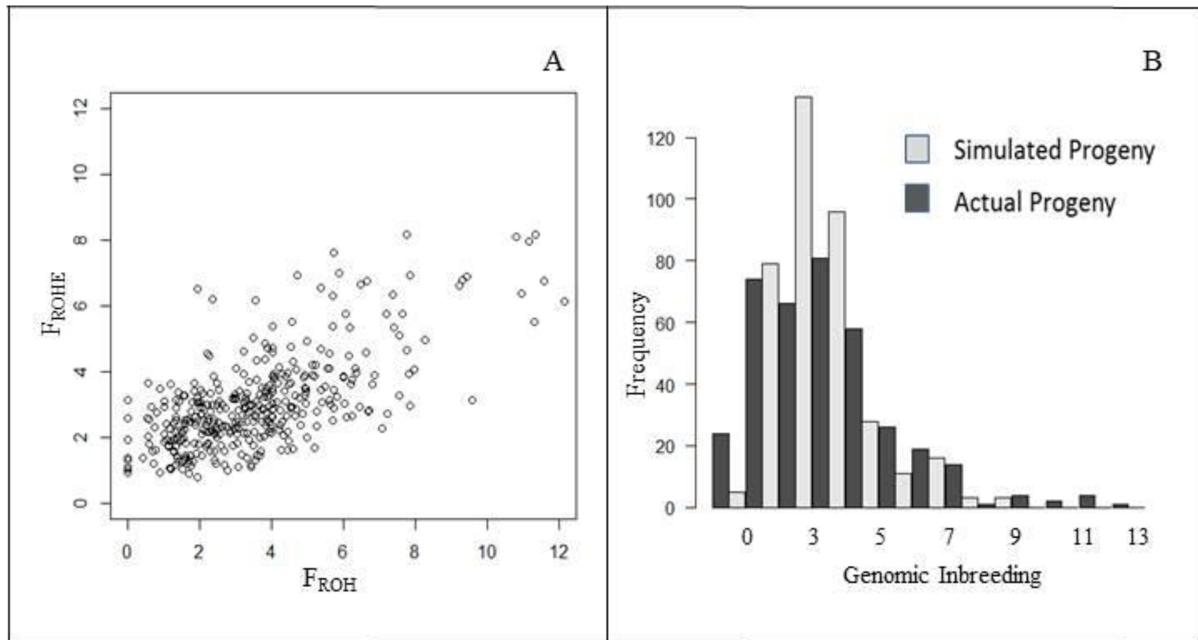


Figure 4. Effects of changes in recombination rate when simulated progeny on the average predicted F_{PH} (A), average predicted F_{ROH} (B), and the correlation between actual progeny F_{PH} and F_{ROH} and average predicted F_{PH} and F_{ROH} (C).

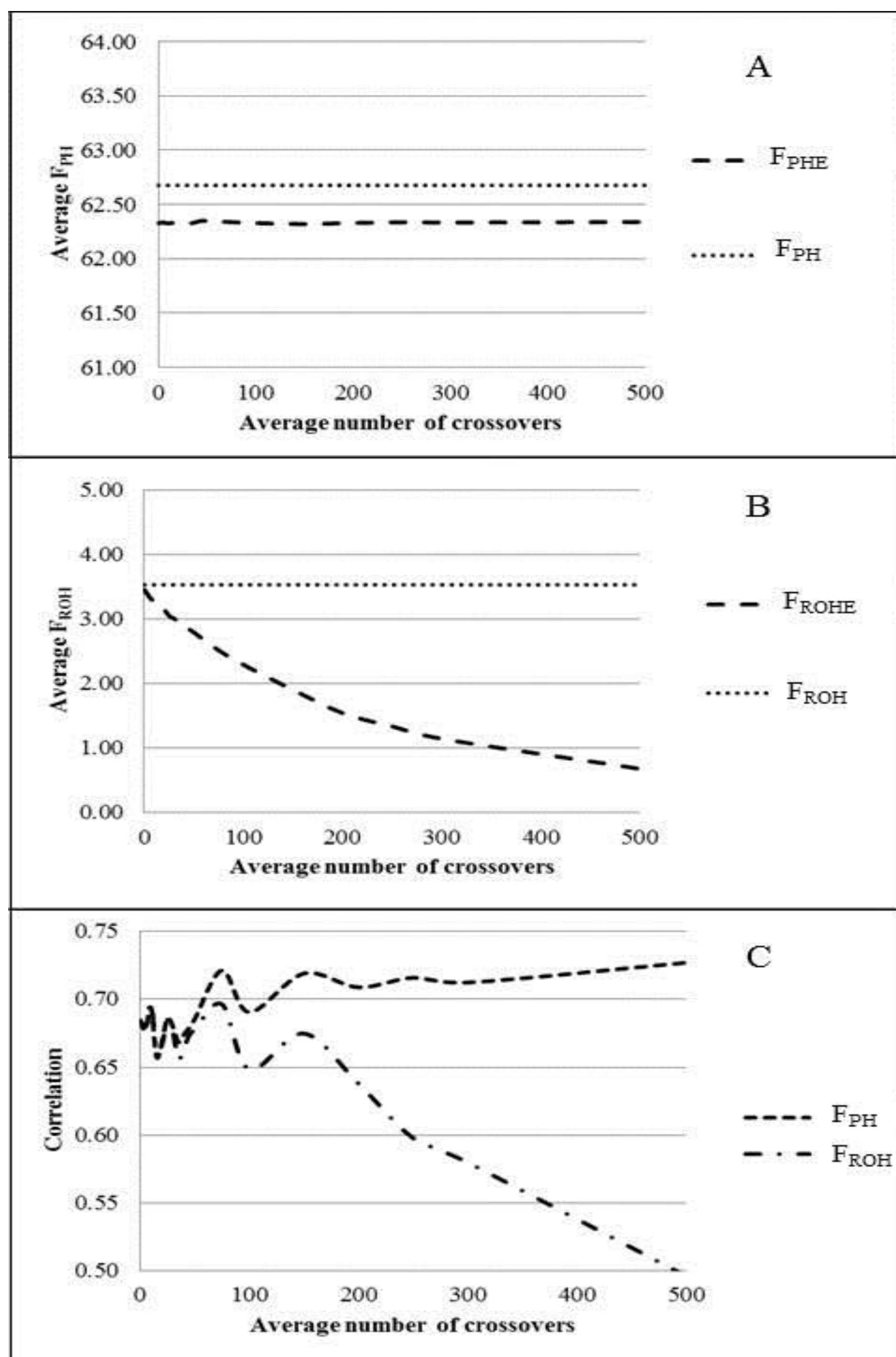
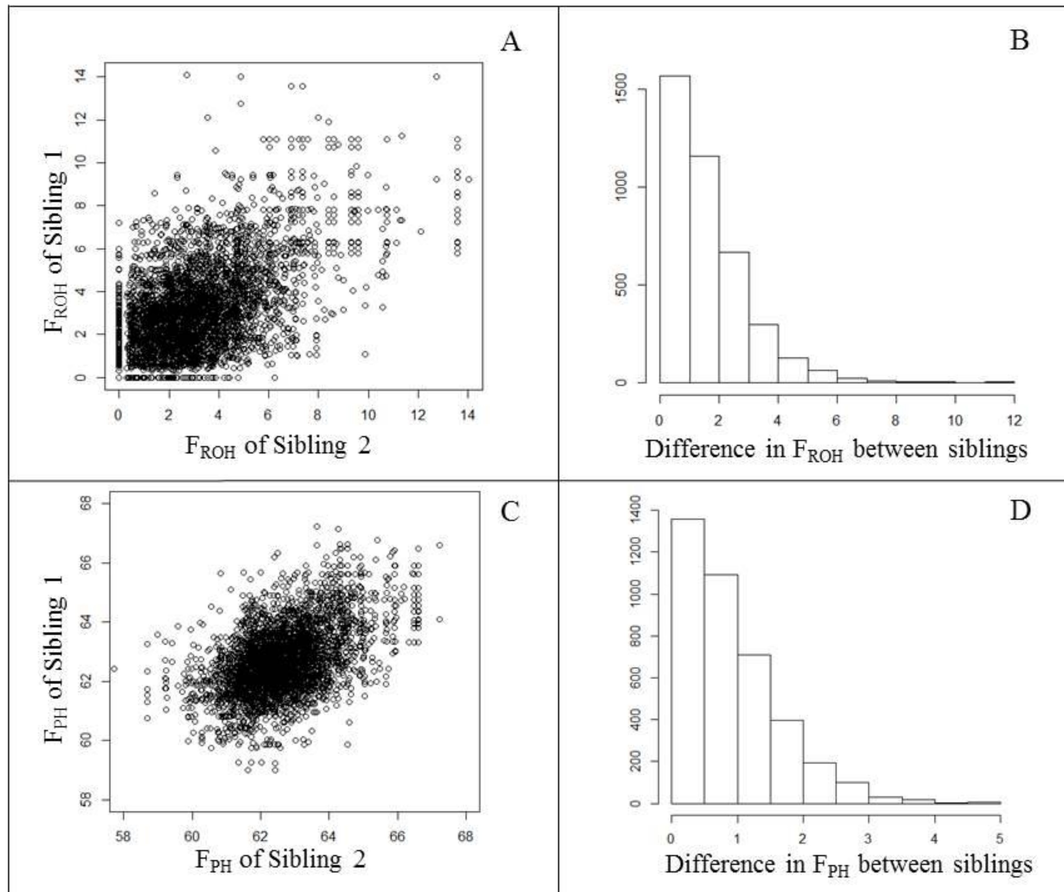


Figure 5. Scatterplot of F_{ROH} (A) and F_{PH} (C) between sibling pairs and the absolute values of the differences of F_{ROH} (B) and F_{PH} (D) between siblings.



CHAPTER 6

Overall Summary

The goal of the first research project presented in this manuscript (Chapter 3) was to determine the effects of increases in genomic inbreeding on economically important traits in Holstein cattle. Three measurements of genomic inbreeding were utilized in this study: the percentage of homozygous SNP in the animal's genome (F_{PH}), inbreeding derived from runs of homozygosity (F_{ROH}), and inbreeding derived from a genomic relationship matrix (F_{GRM}). Increases in all three measures of genomic inbreeding had negative effects on 205-d milk yield, average daily milk yield, days open, and calving ease. Linear type traits such as strength, front teat placement, teat length and udder tilt were also affected by increases in genomic inbreeding. Previous research has shown that when selection is being made with genomic information, limiting increases in genomic inbreeding, rather than pedigree based inbreeding is required. The current research suggests that any of these three measures of genomic inbreeding may be useful in limiting the accumulation of common alleles. F_{ROH} has been previously demonstrated to have the highest correlation with homozygous mutation load, as well as providing coefficients which are both more similar in distribution to pedigree inbreeding and easier to calculate with different base population. These qualities may suggest that utilizing F_{ROH} as the measure of genomic inbreeding may be most beneficial.

The second research project (Chapter 4) compares genomic inbreeding between Holstein, Jersey, Angus, and Nelore cattle, between genetically elite animals and commercial animals, and between measures derived from varying densities of SNP panels. Three measures of genomic inbreeding were calculated for each genetic group and SNP density: F_{PH}, F_{ROH} with a minimum ROH length of 4,000 kb (F_{ROH4}), and F_{ROH} with a minimum ROH length of 8,000 kb (F_{ROH8}). Genetically elite Holstein animals had higher genomic inbreeding values when compared to the commercial Holsteins, indicating that the intense selection place on the genetically elite animals

may result in reduced heterozygosity. When comparing breeds, Nelore animals had the highest F_{PH} but the lowest average F_{ROH4} and F_{ROH8}, suggesting that the actual inbreeding of these animals may be low, but as the SNP in the marker panel were selected to be polymorphic in *Bos taurus* breeds, the more highly polymorphic SNP in Nelore cattle are not present. Jersey cattle had the highest F_{ROH4} and F_{ROH8} values which is consistent with current pedigree information suggesting that the Jersey breed is highly inbred. Ascertainment bias in the selection of SNP for the lower density panels were evident as the equally spaced 6K panel and 50K panel had similar F_{PH} measurements, but increased F_{PH} values were seen in the low density SNP panel for all breeds except Nelore. As genomic information from Nelore cattle did not account for the selection of SNP in the low density SNP panel, F_{PH} for all three SNP panels were similar. The decreased density of the equally spaced 6K and low density SNP panels also made detecting ROH of shorter length very difficult. These results may suggest that only recent inbreeding may be accurately calculated with lower density SNP panels.

The third research project (Chapter 5) focused on studying genomic inbreeding in family based structures and had two main goals. The first was to compare expected genomic inbreeding calculated from genotypes in sire-dam mating pairs to the actual genomic inbreeding of their progeny. This was done to determine if high levels of homozygosity in the embryos have negative effects, causing the progeny that are actually born to less homozygotic than expected. The second goal was to compare genomic inbreeding between full sibling pairs to determine the variation in the measures caused by Mendelian sampling. Results indicated that actual genomic inbreeding measures from progeny were either similar or slightly exceeded expected genomic inbreeding derived from their parents, which is the opposite of what would be expected if genomic inbreeding had a large effect on embryo viability. When separating the expected

genomic inbreeding measures into thirds based on the lowest, median, and highest measures, differences in variation were observed. The larger variation for the actual genomic inbreeding in the highest expected genomic inbreeding groups may suggest noticeable negative effects are present only if the mating pair is closely related. Although these results may also be simply due to differences in Mendelian sampling and the chance of receiving common alleles in the actual progeny. Further results indicated that full siblings were moderately correlated when genomic inbreeding measures were compared. Differences in F_{ROH} between the siblings averaged 0.9 ± 0.7%, but differences up to 11.4% were observed between some full sibling pairs. As these animals would have the same pedigree inbreeding coefficient, results indicate that using genomic inbreeding would be more beneficial due to its ability to determine that actual inbred segments present in an animal.

Overall, results indicated that measures of genomic inbreeding can be calculated efficiently. Increases in these measures have been shown to have negative effects on economically important traits in dairy cattle. Differences between several domestic cattle breeds were also observed, as well as differences between genetic groups. Calculating expected genomic inbreeding from sire-dam mating pairs were also performed. Comparisons to actual genomic inbreeding did not indicate large effects due to increases in embryonic inbreeding.