

J. Dairy Sci. 98:4934–4944 http://dx.doi.org/10.3168/jds.2014-9014 © American Dairy Science Association<sup>®</sup>, 2015.

# Using a family-based structure to detect the effects of genomic inbreeding on embryo viability in Holstein cattle

D. W. Bjelland,\*<sup>1</sup> K. A. Weigel,\* A. D. Coburn,† and R. D. Wilson† \*Department of Dairy Science, University of Wisconsin, Madison 53706 †Genex Cooperative, Shawano, WI 54166

# ABSTRACT

Recent evidence has suggested that some of the decline in reproductive ability in dairy cattle has been caused by embryonic death. The current study compared expected genomic inbreeding from sire-dam mating pairs to genomic inbreeding from live progeny in an attempt to determine how embryonic inbreeding may affect fertility. A total of 11,484 Holstein cattle with 43,485 SNP markers and pedigree information were available for analysis. A total of 412 sire-dam-progeny trios in which all animals had reliable genotypes were discovered. After removal of trios because of parentage errors, 374 remained for analysis. Additionally, a total of 3,031 animals comprising 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 ( $F_{\text{ROHE}}$  and  $F_{\text{PHE}}$ ). Actual genomic inbreeding measures were calculated using the percent homozygosity of all SNP  $(F_{PH})$  and using runs of homozygosity (F\_{\rm ROH}). Average F\_{\rm PHE} values (62.8  $\pm$ 0.78%) were slightly lower than  $F_{PH}$  (63.1  $\pm$  1.12%), when considering each SNP independently. After phasing haplotypes,  $F_{PHE}$  (62.5  $\pm$  0.83%) was again slightly lower than  $F_{PH}$  (62.7  $\pm$  1.16%), and  $F_{ROHE}$  (3.46  $\pm$ 1.54%) was slightly lower than  $\mathrm{F_{ROH}}$  (3.53  $\pm$  2.17%). Results suggest increases in expected genomic inbreeding do not explain a large effect on embryo viability at average levels of expected inbreeding. Higher variation in F<sub>ROH</sub> values was present with sire–dam mating pairs exhibiting high F<sub>ROHE</sub>, which may suggest high levels of genomic inbreeding are required for a noticeable effect on overall embryo viability. Genomic inbreeding between full siblings was also compared with moderate correlations (0.47–0.52) present. Overall, expected genomic inbreeding measures were calculated, but results did not suggest a large effect of expected inbreeding on embryo viability.

**Key words:** genomic inbreeding, embryonic inbreeding, simulated progeny

# INTRODUCTION

Dairy cattle populations have seen a decline in reproductive ability over the past several decades. Washburn et al. (2002) analyzed the reproductive traits in Holstein and Jersev cows of the southeastern United States and reported an increase in days open from an average of  $122 \pm 2.8$  d in 1978 to  $152 \pm 2.8$  d in 1999 for Jerseys and from an average of 124  $\pm$  0.7 d in 1978 to 168  $\pm$ 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%, and 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 Jersevs (2.0 to 2.3 services), as did the average calving interval for Holsteins (410 to 422 d) and Jersevs (398 to 410 d).

Failures in reproduction can be caused by many aspects, such as increases in inbreeding, poor estrus detection, anestrus or abnormal luteal phases in highproducing 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 d 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 d 28 and 38 through 58 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 d 31 and 45 of gestation, with a total

Received October 24, 2014.

Accepted March 25, 2015.

<sup>&</sup>lt;sup>1</sup>Corresponding author: bjelland34@gmail.com

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, the chance of deleterious lethal disorders to be present is greater, such as bovine leukocyte adhesion deficiency (BLAD; Kehrli et al., 1990) or deficiency in uridine monophosphate synthase (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 whole-genome sequencing tools in cattle, 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 3 measures of genomic inbreeding, inbreeding derived from runs of homozygosity  $(\mathbf{F}_{ROH})$ , inbreeding derived from a genomic relationship matrix, and the overall percent homozygosity of the genome ( $\mathbf{F}_{\mathbf{PH}}$ ), and discovered negative effects on both milk production and reproductive traits with increases in all measures of genomic inbreeding. Increases in days open were between 1.06 and 1.76 d per 1-percentage-unit increase in inbreeding for the 3 measures of genomic inbreeding. Pryce et al. (2014) also discovered negative effects of increases in various measures of maternal genomic inbreeding on calving interval in Holstein and Jersey cows. They also examined specific areas of the genome and discovered several regions in which increases in homozygosity had a negative effect on fertility. Furthermore, increases in  $F_{ROH}$  and  $F_{PH}$  in human populations have been correlated with higher risk of disease (Keller et al., 2012; Simón-Sánchez et al., 2012) and decreases in fitness of quantitative traits (McQuillan et al., 2012).

The study herein attempted to determine whether the genomic inbreeding measures of progeny are appropriate when compared with what is expected from the parents. Deviations from what is expected may suggest that the highly inbred animals do not survive full term and that embryonic inbreeding has a large effect on the survival of the embryo.

#### Data

MATERIALS AND METHODS

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-Agricultural Research Service Animal Genomics and Improvement 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% parentprogeny 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). The SNP from the Genex data set were reduced to the 43,485 SNP used by the USDA-Agricultural Research Service data. Further editing of SNP was performed on the complete data set with the removal of SNP because of minor allele frequency (<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.

Sire-dam-progeny information was provided by Genex Cooperative. Up to 5 generation pedigrees were obtained from USDA-Agricultural Research Service Animal Genomics and Improvement for all genotyped animals. A total of 412 sire–dam–progeny trios were discovered using this pedigree for analysis, with 374 sire-dam-progeny trios remaining after removal of trios because of greater than 5% parent-progeny conflicts. In total 607 unique animals, consisting of 68 sires, 165 dams, and 374 progeny, were present in the 374 trios. Sires averaged 18 progeny, with the maximum of 39 and minimum of 1. Each dam averaged 5 progeny, with a maximum of 12 and a minimum of 1, whereas the largest full-sibling family consisted of 6 progeny, with 2 others containing 5 progeny. Additionally, 3,031 genotyped Holsteins within full-sibling families (without genotyped parents), composed of a total of 3,906 fullsibling pairs, were also discovered from the pedigree information. Pedigree inbreeding  $(\mathbf{F}_{ped})$  measures were derived from 5-generation pedigrees for all animals, as well.

#### Simple Method

To detect possible effects of genomic inbreeding on embryo viability, actual genomic inbreeding from live progeny were compared with genomic inbreeding that would be expected from the parents of these live progeny. Lower-than-expected actual genomic inbreeding could indicate recessive deleterious effects at higher levels of genomic inbreeding within this specific mating pair. The first method used 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 has the genotype AB and the dam has the genotype AB, 4 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 ( $\mathbf{F}_{PHE}$ ) is then calculated by the following formula,

$$\mathbf{F}_{\mathrm{PHE}} = \frac{\sum_{i}^{m} P\left(Hom_{i}\right)}{m}$$

where m = number of nonmissing markers and  $P(Hom_i)$ = probability of producing a homozygous marker at locus *i*, based on the probabilities in Table 1. The actual homozygosity (**F**<sub>PH</sub>) of the progeny is then calculated as

$$\mathbf{F}_{\rm PH} = \frac{N_{\rm AA} + N_{\rm BB}}{N_{\rm AA} + N_{\rm AB} + N_{\rm BB}},$$

where  $N_{AA}$ ,  $N_{AB}$ , and  $N_{BB}$  are the numbers of SNP that are classified as AA, AB, and BB, respectively. For each sire–dam–progeny trio,  $F_{PHE}$  and  $F_{PH}$  were then compared using a paired *t*-test.

#### Phased Haplotype Method

The second method used to determine expected inbreeding simulated possible 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 used in building a new hidden Markov model. Then in the sampling step, new haplotypes are sampled for each sire-dam-progeny trio based on the genotypic data and 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, 1 of the 2 chromosomal haplotypes was selected at random from each parent. After 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.

A method to include recombination events in the simulations was also evaluated. Simulations including 10, 20, 30, 40, 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 was formed for each parent, one was randomly selected from each parent and assembled into the 250 simulated progeny for that specific recombination rate. For this simulation, after the potential gametes were formed,  $(6 \times 6)^{29} = 1.4 \times$  $10^{45}$  possible progeny could be simulated from the 6 potential gametes for each parent.

Once the simulated progeny were produced, 2 separate methods were used to produce expected inbreeding coefficients. The first calculated the  $F_{PH}$  in the actual progeny with the same method outlined calculating  $F_{PH}$ for the simple method.  $F_{PHE}$  was also calculated using this formula for each of the 250 simulated progeny in each mating pair, and then averaged to create the

**Table 1.** All possible genotypes in progeny for given sire and dam genotypes, with probability the genotypeis homozygous in parentheses

Dam genotype	Sire genotype		
	AA	AB	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)

 $F_{\rm PHE}$  value to compare against  $F_{\rm PH}.$  The second method used PLINK (Purcell et al., 2007) to determine runs of homozygosity (**ROH**) for each simulated 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 passed down together, rather than a single SNP or marker. If inbreeding occurs, the large sections of DNA that 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. Linkage-disequilibrium pruning was also performed before 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 before phasing to eliminate any bias the phasing process may produce. A total of 6,452 SNP remained for ROH analysis after linkage-disequilibrium pruning. Because 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

$$F_{ROH}$$
 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**<sub>ROHE</sub> from the simulated progeny for each mating pair were then averaged to provide a coefficient to compare with F<sub>ROH</sub> from the actual progeny. Comparisons were made between F<sub>PH</sub> and F<sub>PHE</sub> and between F<sub>ROH</sub> and F<sub>ROHE</sub> using paired *t*-tests.

# **Full-Sibling Analysis**

In addition to the sire–dam–progeny trios, a total of 3,906 full-sibling pairs in which both siblings had genotype information were available for analysis. Two genomic measures of inbreeding previously outlined,  $F_{\rm ROH}$  and  $F_{\rm PH}$ , 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.

#### **RESULTS AND DISCUSSION**

#### Expected Versus Actual Genomic Inbreeding

Means of actual and expected genomic inbreeding measures are presented in Table 2. When using the simple method with SNP treated independently,  $F_{PHE}$  $(62.8 \pm 0.78\%)$  was slightly lower than  $F_{PH}$  (63.1  $\pm$ 1.12%). Similar results were present after simulation of progeny from phased haplotypes, where  $F_{PHE}$  (62.5)  $\pm$  0.83%) was slightly lower than F<sub>PH</sub> (62.7  $\pm$  1.16%), and  $F_{ROHE}$  (3.46  $\pm$  1.54%) was also slightly lower than  $F_{ROH}$  (3.53 ± 2.174 = %). These results do not suggest that expected inbreeding of progeny has a large effect on the viability of the embryos. To illustrate what the negative effect of increased genomic inbreeding on embryonic viability would look like, we performed a simple simulation. In the simulation, we had 400 mating pairs with the expected progeny inbreeding of 5.0% and a standard deviation of 1.0%. Running this simulation 10 times, mean of simulated progeny inbreeding was 4.98%. Next, to simulate the deleterious effects of high amounts of genomic inbreeding, we gave an increasing likelihood of removing the progeny (representing the embryo dying in utero) with an increase in the simulated genomic inbreeding, beginning with a 5% chance of dying with a value between 0 and 0.5 standard deviations above the mean to a 75% chance of dying with a value greater than 3 standard deviations from the mean. After running this simulation 10 times, we had an average of 28 progenv removed for each run, which is a loss of 7.0% of the possible progeny, and our mean progeny genomic inbreeding for the 10 runs was then 4.91%. Given that Moreira et al. (2001), Cartmill et al. (2001), and Chebel et al. (2004) had estimated embryonic losses of between 12.5 and 28.0%, we felt as though this was a conservative estimate for embryonic death in utero. As shown in this simulation, if the genomic inbreeding had a large effect on embryo viability, we should see a decrease in the actual progeny inbreeding when compared with the expected inbreeding from the mating pair.

The differences seen in the values of  $F_{PH}$  in Table 2 are present because for the simple method, calculations were made with unphased genotypes, whereas phased haplotypes were used 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.

Table 2. Homozygosity ( $F_{PHE}$  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<sup>1</sup>

Item	Mean	SD
Simple method		
FPHE	62.83	0.78
$F_{PH}$	63.14	1.12
Difference	-0.31	
Simulated progeny		
$\mathrm{F}_{\mathrm{PHE}}$	62.489	0.832
$\mathrm{F}_{\mathrm{PH}}$	62.672	1.16
Difference	-0.183	
$F_{ROHE}$	3.46	1.537
$F_{ROH}$	3.53	2.171
Difference	-0.069	

 $^1\mathrm{F}_{\mathrm{PHE}}=$  expected percent homozygosity;  $\mathrm{F}_{\mathrm{PH}}=$  actual percent homozygosity of live progeny;  $\mathrm{F}_{\mathrm{ROHE}}=$  expected inbreeding derived from runs of homozygosity in simulated progeny;  $\mathrm{F}_{\mathrm{ROH}}=$  inbreeding derived from runs of homozygosity in live progeny.

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. Studies that have focused on inbreeding, both pedigree (Smith et al., 1998) and genomic (Bjelland et al., 2013; Pryce et al., 2014), have found large negative effects of increases in maternal inbreeding on reproductive ability. Results here may suggest that other than avoiding large deleterious effects, the overall reproductive ability and viability of the embryo may be more affected by increases in maternal inbreeding rather than embryonic. Furthermore, 2 aspects may have limited the power of this study. The first is that only 374 sire-dam-progeny trios were available for analysis. Continuing this analysis by adding more trios may provide a more accurate effect of embryonic inbreeding. Second, because these matings were selected to avoid inbreeding, the majority of the mating pairs would have low expected inbreeding coefficients. To more fully realize the effect of high levels of expected inbreeding, half-sibling, full-cousin, or fullsibling matings would need to be performed. Although this may have been seen as a limitation in this study, the expected inbreeding levels would also be similar to those seen on a commercial dairy farm, which may actually provide more useful information for the dairy industry.

Further visualizations of the  $F_{PHE}$  compared against  $F_{PH}$  using the simple method are presented in Figure 1. A correlation between the 2 measures of 0.70 is shown in Figure 1A, suggesting that predicting the inbreeding of progeny using this method is acceptable and could be used in mate-selection programs. Histograms including both  $F_{PHE}$  and  $F_{PH}$  present in Figure 1B demonstrate the similarities in the overall mean and distribution of the 2 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 simulating progeny. Correlations between  $F_{PHE}$  and  $F_{PH}$ (0.68) and between  $F_{ROHE}$  and  $F_{ROH}$  (0.68) were also moderately high, suggesting that accurate prediction of genomic inbreeding in progeny can be made by using phased haplotype of potential mates. Figure 2A and Figure 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 measure of genomic inbreeding separately and split into the lowest, median, and highest predicted genomic inbreeding values. Standard deviations of  $F_{PH}$  varied greatly when the data were split into the 3 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%. The large variation in the high  $F_{PHE}$  group may correlate to negative effects of the increases in predicted inbreeding on embryo viability. Similar results are present when creating the 3 data sets for  $F_{ROHE}$ , with standard deviations of 1.27, 1.57, and 2.45%, for  $F_{ROH}$  in the lowest, median, and highest  $F_{ROHE}$  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 very high levels of expected inbreeding may cause a slight decrease in the viability of the embryo possibly because of an accumulation of small negative effects on fertility, such as those discovered by Khatib et al. (2009), although more data would be needed to accurately assess this hypothesis.

#### 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 because increasing the number of crossovers should be independent of whether homozygous or heterozygous SNP were selected. This also suggests that the  $F_{PHE}$  values calculated using the simple method, as discussed previously, should provide an adequate estimation of  $F_{PH}$  in possible progeny.

In contrast, the  $F_{ROHE}$  values simulated from progeny were highly affected by increases in the average recombination rate. This is also expected, as increasing the



Figure 1. Total expected homozygosity ( $F_{PHE}$ ) predicted using the simple method compared with actual overall percent homozygosity of the genome ( $F_{PH}$ ) 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).

average recombination rate, in this case, would be analogous to increasing the distance to a common ancestor in this individual's pedigree. For example, if the actual average recombination rate would be 30 recombination events per meiosis, using an average of 60 recombination events would essentially create 2 meiosis events and make the parents of the current individual into the grandparents of the current individual with respect to the number of recombination events. In this simulation, the recombination events are the only effects present



Figure 2. Total expected homozygosity ( $F_{PHE}$ ) predicted using simulated progeny to actual overall percent homozygosity of the genome ( $F_{PH}$ ) 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).

4940

BJELLAND ET AL.



Figure 3. The  $F_{ROHE}$  (expected inbreeding derived from runs of homozygosity in simulated progeny) predicted using simulated progeny compared with actual inbreeding derived from runs of homozygosity ( $F_{ROH}$ ) 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).

that can break up the potential ROH, but no possible events exist, such as other inbred matings, that can create new ROH. So, as shown in Figure 4B, as the average number of recombination events increases, the average  $F_{\rm ROHE}$  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 present in Figure 4C. Fluctuations when the average recombination events is less than 100 are most likely due to the relatively small (n = 250)number of progeny that 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. Because 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 higher recombination rates are used.

# **Full-Sibling Pairs**

Genomic inbreeding measures for the 3,031 animals with full siblings ranged from 0 to 14.1%  $F_{ROH}$ , with a

mean of  $3.1 \pm 1.9\%$ , and 57.7 to 67.2% for  $F_{PH}$ , with a mean of  $62.6 \pm 1.1\%$ . The 2 genomic inbreeding measures were moderately correlated with an  $R^2 = 0.76$ . Scatterplots with one full sibling plotted against the other full sibling are presented in Figure 5A for  $F_{ROH}$ and Figure 5C for  $F_{PH}$ , with a histogram of the absolute values of differences between siblings presented in Figure 5B for  $F_{ROH}$  and Figure 5D for  $F_{PH}$ . Moderate correlations between siblings was present for both  $F_{ROH}$  ( $R^2 = 0.51$ ) and  $F_{PH}$  ( $R^2 = 0.47$ ). 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  $0.9 \pm 0.7\%$ , and the mean difference for  $F_{PH}$  was  $1.6 \pm 1.3\%$ .

A separate analysis was performed by only looking at 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) only if one of the siblings had an  $F_{ROH}$  value between 6.0 and 7.0%. The average difference for the least inbred animals was  $1.7 \pm 1.3\%$ , and the average difference between the most inbred animals was  $2.4 \pm 1.6\%$ . The higher mean and variation present in the most inbred group suggests that if a sire and dam are closely related, there is a high probability of their progeny being highly inbred, but also chance of their progeny receiving only a small proportion of DNA that is inbred. In fact, one sibling



Figure 4. Effects of changes in recombination rate when simulating progeny on the average predicted overall percent homozygosity of the genome ( $F_{PH}$ ; A), average predicted inbreeding derived from runs of homozygosity ( $F_{ROH}$ ) and expected inbreeding derived from runs of homozygosity in simulated progeny ( $F_{ROHE}$ ; B), and the correlation between actual progeny  $F_{PH}$  and  $F_{ROH}$  and average predicted  $F_{PH}$  and  $F_{ROH}$  (C).

4942



Figure 5. Scatterplot of inbreeding derived from runs of homozygosity ( $F_{ROH}$ ; A) and average predicted overall percent homozygosity of the genome ( $F_{PH}$ ; C) between sibling pairs and the absolute values of the differences of  $F_{ROH}$  (B) and  $F_{PH}$  (D) between siblings.

in the highly inbred subset had an  $F_{ROH}$  value of 0, and 52 out of the 362 (14.3%) total siblings had an  $F_{ROH}$ 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%. In a product of double second 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%, and the product of a second-cousin mating, with an  $F_{ped}$  of 1.56%, has a standard deviation estimated of 1.1%.

To further study the differences in variation of inbreeding measures from close relatives, 200 full-sibling, half-sibling, full-first-cousin, and half-first-cousin pairs were discovered from the pedigree data. Potential progeny were then simulated from these pairs, and  $F_{ROHE}$ values were calculated. Results are presented in Figure 6, with the progeny simulated from full siblings having the highest mean  $F_{ROHE}$  and largest standard deviation (22.3 ± 2.4%), followed by progeny simulated from half siblings  $(12.3 \pm 1.9\%)$ , progeny simulated from full first cousins  $(7.6 \pm 1.6\%)$ , and progeny simulated from half first cousins (5.2  $\pm$  1.5%). The full first cousins (F<sub>ped</sub> = 6.25%) and half first cousins (F  $_{\rm ped}$  = 3.125%) have a slightly higher  $F_{ROHE}$  than expected  $F_{ped}$ , which is mostly likely due to other common ancestors further back in the pedigree than could be distinguished here, whereas the  $F_{ped}$  values are assuming no other common relatives. The lower  $F_{ROHE}$  than expected  $F_{ped}$  (25%) simulated from full sibling may be due to the process of calling ROH. The extremely long ROH that would be expected from full-sibling matings are much more easily broken up by a single false heterozygote or missing SNP. If this occurs in the middle of a ROH, both adjoining sections should still be called as a ROH, and little difference would be seen in the final  $F_{ROHE}$ . But if a false heterozygote or missing SNP occurs closer to the end of a true ROH, one of the adjoining sections may not meet the minimum length requirement to be called a ROH, which would artificially decrease the calculated  $F_{ROHE}$ . As the results presented in the current study and

in the study by Carothers et al. (2006) suggest, as the common ancestor in a pedigree is further removed from the current individuals, the inbreeding coefficients and the variation in the inbreeding coefficients decrease.

## CONCLUSIONS

Expected genomic inbreeding measures from siredam mating pairs were compared against genomic inbreeding values from actual live progeny to determine whether 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 was based on simulating progeny from phased haplotypes. Both methods generated sufficient measures of expected genomic inbreeding. Slight increases in the actual genomic inbreeding when compared with the expected were observed, which is the opposite of anticipated research findings, although high expected levels of genomic inbreeding



Figure 6. Frequency distributions of inbreeding derived from runs of homozygosity ( $F_{ROH}$ ) calculated from 200 simulated progeny from 250 full-sibling, half-sibling, full-first-cousin, and half-first-cousin mating pairs.

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.

## ACKNOWLEDGMENTS

D. W. Bjelland acknowledges financial support from USDA National Institute of Food and Agriculture National Needs Graduate Fellowship Grant Number 2010-38420-30477. D. W. Bjelland thanks the employees of Genex Cooperative/CRI (Shawano, WI), for their input and cooperation, in particular, Glenn Gilbert, Keith Heikes, Gamal Azim, and Ami Lasecki.

### REFERENCES

- Bjelland, D. W., K. A. Weigel, N. Vukasinovic, and 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.
- 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.
- 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. 51:1826–1836.
- Keller, M. C., M. A. Simonson, S. Ripke, B. M. Neale, P. V. Gejman, D. P. Howrigan, S. H. Lee, T. Lencz, D. F. Levinson, P. F. Sullivan, Schizophrenia Psychiatric Genome-Wide Association Study Consortium. 2012. Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. PLoS Genet. 8:e1002656. 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.
- 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.
- Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: Where will it end? J. Dairy Sci. 84:1277–1293.
- Mc Parland, S., J. F. Kearney, M. Rath, and D. P. Berry. 2007. Inbreeding effects on milk production, calving performance, fertil-

ity, and conformation in Irish Holstein-Friesians. J. Dairy Sci.  $90{:}4411{-}4419.$ 

- McQuillan, R., N. Eklund, N. Pirastu, M. Kuningas, B. P. McEvoy, T. Esko, T. Corre, G. Davies, M. Kaakinen, L. P. Lyytikäinen, K. Kristiansson, A. S. Havulinna, M. Gögele, V. Vitart, A. Tenesa, Y. Aulchenko, C. Hayward, A. Johansson, M. Boban, S. Ulivi, A. Robino, V. Boraska, W. Igl, S. H. Wild, L. Zgaga, N. Amin, E. Theodoratou, O. Polašek, G. Girotto, L. M. Lopez, C. Sala, J. Lahti, T. Laatikainen, I. Prokopenko, M. Kals, J. Viikari, J. Yang, A. Pouta, K. Estrada, A. Hofman, N. Freimer, N. G. Martin, M. Kähönen, L. Milani, M. Heliövaara, E. Vartiainen, K. Räikkönen, C. Masciullo, J. M. Starr, A. A. Hicks, L. Esposito, I. Kolčić, S. M. Farrington, B. Oostra, T. Zemunik, H. Campbell, M. Kirin, M. Pehlic, F. Faletra, D. Porteous, G. Pistis, E. Widén, V. Salomaa, S. Koskinen, K. Fischer, T. Lehtimäki, A. Heath, M. I. McCarthy, F. Rivadeneira, G. W. Montgomery, H. Tiemeier, A. L. Hartikainen, P. A. Madden, P. d'Adamo, N. D. Hastie, U. Gyllensten, A. F. Wright, C. M. van Duijn, M. Dunlop, I. Rudan, P. Gasparini, P. P. Pramstaller, I. J. Deary, D. Toniolo, J. G. Eriksson, A. Jula, O. T. Raitakari, A. Metspalu, M. Perola, M. R. Järvelin, A. Uitterlinden, P. M. Visscher, J. F. Wilson, and ROHgen Consortium. 2012. Evidence of inbreeding depression on human height. PLoS Genet. 8:e1002655. 10.1371/journal.pgen.1002655.
- 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.
- Pryce, J. E., M. Haile-Mariam, M. E. Goddard, and B. J. Hayes. 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genet. Sel. Evol. 46:71.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 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.
- Simón-Sánchez, J., L. L. Kilarski, M. A. Nalls, M. Martinez, C. Schulte, 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. 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.
- 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. 2009. 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. 92:3431–3436.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42.