Haplotypes responsible for early embryonic lethality detected in Nordic Holsteins

Xiaoping Wu,^{1*} Md Mesbah-Uddin,^{1,2} Bernt Guldbrandtsen,¹ Mogens S. Lund,¹ and Goutam Sahana¹

¹Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark

²Animal Genetics and Integrative Biology, UMR 1313 GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

ABSTRACT

Widespread use of a limited number of elite sires in dairy cattle breeding increases the risk of some deleterious allelic variants spreading in the population. Genomic data are being used to detect relatively common (frequency >1%) haplotypes that never occur in the homozygous state in live animals. Such haplotypes likely include recessive lethal or semilethal alleles. The aim of this study was to detect such haplotypes in the Nordic Holstein population and to identify causal genetic factors underlying these haplotypes. Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) genotypes for 26,312 Nordic Holstein animals were phased to construct haplotypes. Haplotypes that are common in the population but never observed as homozygous were identified. Two such haplotypes overlapped with previously identified recessive lethal mutations in Holsteins—namely, structural maintenance of chromosomes 2 (HH3) and brachyspina. In addition, we identified 9 novel putative recessive lethal-carrying haplotypes, with 26 to 36 homozygous individuals expected among the genotyped animals but only 0 to 3 homozygotes observed. For 2 out of 9 homozygous-deficient haplotypes, insemination records of at-risk mating (carrier bull with daughter of carrier sire) showed reduced insemination success compared with not-at-risk mating (noncarrier bull with daughter of noncarrier sire), supporting early embryonic mortality. To detect the causative variant underlying each homozygous-deficient haplotype, data from the 1000 Bull Genome Project were used. However, no variants or deletions identified in the chromosome regions covered by the haplotypes showed concordance with haplotype carrier status. The carrier status of detected haplotypes could be used to select bulls to reduce the frequency of the latent lethal mutations in the population. If desired, at-risk matings could be avoided.

Key words: recessive lethal region, recessive lethal haplotype, at-risk mating, early embryonic mortality

INTRODUCTION

Recessive lethal alleles in a homozygous state cause the death of an organism. Widespread use of a limited number of elite sires for AI in dairy cattle has improved productivity but at the same time increased the risk of deleterious allelic variants spreading in the population and becoming homozygous in individuals in subsequent generations. Hundreds of haplotypes or mutations related to inherited disorders have been recorded in cattle (University of Sydney, 2019).

The traditional approach to identifying genetic factors causing defects or death is to trace the common ancestors of the affected animals using pedigree information. However, this requires information from both affected and nonaffected animals based on phenotypic identification. This approach is not able to identify deleterious genetic mutation in cases where the phenotype is not detectable, such as early embryonic deaths. Instead, genomic data can be used to identify haplotypes that are common in the population but never occur in a homozygous state in live animals (VanRaden et al., 2011). With this approach, we can identify haplotypes putatively carrying lethal variants based on genomic information collected only from live animals, and no phenotypic records are required. Several recessive lethal-carrying haplotypes have been identified in cattle using this approach (VanRaden et al., 2011; Fritz et al., 2013; Sahana et al., 2013). Subsequently, the underlying factors causing embryonic mortality in cattle have been identified for some of the reported lethal haplotypes, such as structural maintenance of chromosomes 2 p.Phe1135Ser (HH3; McClure et al., 2014) and phosphoribosylglycinamide formyltransferase (GART) p.Asn290Thr (HH4; Fritz et al., 2013) in Holsteins, RNASEH2B 660-Kb deletion in Nordic Red dairy cattle (Kadri et al., 2014), tubulin delta 1 (TUBD1) p.His210Arg (BH2) in Brown Swiss (Schwarzenbacher et al., 2016), and CWC15 p.Arg55* (JH1) in Jerseys

Received March 19, 2019.

Accepted August 9, 2019.

*Corresponding author: xiaoping.wu@mbg.au.dk

(Sonstegard et al., 2013). The aim of this study was to identify recessive lethal haplotypes for prenatal death in the Nordic Holstein cattle population and to search for causal genetic factors underlying the recessive lethal effect of these haplotypes.

MATERIALS AND METHODS

Genotyping and Phasing

A total of 54,323 SNP for 26,312 Nordic Holsteins were obtained using Illumina BovineSNP50 BeadChip version 1 and 2 (Illumina Inc., San Diego, CA). A total of 45,105 SNP remained on 29 autosomes after removing SNP with minor allele frequencies below 1% or deviated from Hardy-Weinberg proportions ($P < 10^{-6}$). Then, Beagle 4.0 (Browning and Browning, 2007) was used to impute sporadic missing genotypes and chromosome-wise haplotype phasing. The SNP positions within a chromosome were based on the *Bos taurus* genome assembly UMD3.1 (Zimin et al., 2009). To examine the imputation accuracy, we randomly dropped 3,000 markers for 2,000 cattle and imputed those missing genotypes with Beagle using the rest of the genotyped animals as reference. The imputation accuracy was computed as the mean correlation between the actual and imputed genotypes for these 3,000 markers in the 2,000 cattle.

Genetic Analysis

To detect the embryonic lethal factors using genotype data, we first examined the haplotypes that showed no homotypic individuals among the live animals. Next, we estimated the haplotype effect on insemination success in at-risk matings (matings between carrier bull and daughter of carrier sire) compared with other matings following Kadri et al. (2014). Finally, to detect the causal factors underlying detected homozygous-deficient haplotypes, whole-genome sequence data from *run6* of the 1000 Bull Genomes Project (Daetwyler et al., 2014) were analyzed for concordance between the carrier status of the homozygous-deficient haplotype and sequence variants.

Homozygous-Deficient Haplotype Detection.

To find the genomic region responsible for embryonic recessive lethality, we carried out a 2-step approach using a BovineSNP50 BeadChip. In the first step, 25 consecutive markers were used to construct haplotypes with a sliding window for each autosome. For each haplotype label in a 25-marker window, the observed and expected numbers of homozygotes were compared. The expected number of homozygous individuals for any particular haplotype was calculated by multiplying

the number of genotyped animals by the squared haplotype frequency assuming random mating. Thereafter, we used a chi-squared test to assess the significance of the difference between the number of observed and expected haplotypes. To define a recessive lethal haplotype region, we set the following conditions: (1) the frequency of the haplotype must be higher than 1%, (2) the number of expected homozygous individuals for the haplotype must be higher than the observed homozygotes, (3) the difference between observed and expected haplotypes must be significant after Bonferroni multiple testing correction ($P < 0.05/M = 1.13 \times 10^{-6}$, where $M = 44,409$ is the number of haplotypes with a frequency > 0.01), and (4) the observed number of homozygotes must be 3 or less. Finally, all the haplotypes satisfying the conditions were collected. We grouped these homozygous-deficient haplotypes into homozygous-deficient regions (**HDR**) based on their relative proximity.

In the second step, each HDR was further analyzed to refine the tag haplotype (**TH**) for an underlying causative variant. A sliding window with variable haplotype length was moved along each HDR to refine the TH. Within each sliding window, more than 1 TH was selected to be sure to meet conditions 1 through 3 of step 1 mentioned above. The haplotype completely missing at the homozygous state was selected as the TH; if there were none, we selected the haplotype with the lowest P -value for the difference between observed and expected haplotype frequency as TH.

Effect of Selected Haplotypes on Insemination

Results. We further analyzed the 35-, 56-, 100-, and 150-d nonreturn rate (**NRR**), which are based on the observation that a mated cow has not returned to service within a defined number of days. The NRR was recorded as 1 in cases of insemination success (nonreturn, conception) and 0 in cases where the cow returned to service. The haplotype status for most of the dams was unknown because they were not genotyped. Therefore, we compared the mating outcomes based on the carrier status of the bull and the sire of the cow. Four classes of matings were defined according to carrier status for each TH: (1) noncarrier bull mating with cow of noncarrier sire ($NC_b \times NC_{cs}$), (2) noncarrier bull mating with cow of carrier sire ($NC_b \times C_{cs}$), (3) carrier bull mating with cow of noncarrier sire ($C_b \times NC_{cs}$), and (4) carrier bull mating with cow of carrier sire ($C_b \times C_{cs}$; at-risk mating). If both the bull and sire of the cow are carriers in a mating, then there is a 12.5% risk of the conceptus being homozygous for the haplotype in question. Therefore, we expect $C_b \times C_{cs}$ matings to lower insemination success. Some level of extra return to service will be observed for the $C_b \times NC_{cs}$ mating because the dam may inherit the recessive lethal muta-

tion from its dam. For the other 2 mating types ($NC_b \times C_{cs}$ and $NC_b \times NC_{cs}$), no extra increase in return to service is expected compared with the background level. A total of 2,402,533 records of NRR at 35, 56, 100, and 150 d were analyzed for each identified TH to compare the effect of at-risk matings with mating type 1. The logistic regression model with mating type of each TH as fixed effect was analyzed:

$$\boldsymbol{\eta} = \mathbf{X}_p \mathbf{p} + \mathbf{X}_t \mathbf{t} + \mathbf{X}_m \mathbf{m} + \mathbf{Z} \mathbf{h},$$

where $\boldsymbol{\eta}$ was the logit transformation of $\boldsymbol{\theta}$ with $\eta_i = \ln\left(\frac{\theta_i}{1-\theta_i}\right)$, assuming $\mathbf{y} \sim \text{bin}(n, \boldsymbol{\theta})$ with $n = 1$; $\mathbf{y}$ was a vector of NRR (1 in case of success and 0 in case of failure). We analyzed NRR at 35, 56, 100, and 150 d; θ_i was the probability of nonreturn for the i th observation. Then, the i th element in $\mathbf{y}$ followed a Bernoulli distribution $y_i \sim \text{Be}(\theta_i)$. In addition, $\mathbf{p}$ was a vector of effects of parity; $\mathbf{X}_p$ was an incidence matrix relating phenotypes to parity effects; $\mathbf{X}_t$ was an incidence matrix relating phenotypes to month-year effect; $\mathbf{t}$ was a vector of effects of period-month of insemination (4 periods: yr ≤ 2000 , yr 2001–2005, yr 2006–2010, and yr > 2010); $\mathbf{m}$ was a vector of mating type effects; $\mathbf{X}_m$ was an incidence matrix relating phenotypes to mating type effect for each TH; $\mathbf{h}$ was a vector of random effects of herds; and $\mathbf{Z}$ was an incidence matrix relating phenotypes to random herd effects. The latent variable in the logistic regression model was unobservable, so the residual variance was not estimable. Analyses were performed using the DMU package (Madsen et al., 2014).

Effects of mating type were investigated by comparing the elements (m_1 and m_4) of $\mathbf{m}$ corresponding to mating types 1 (noncarrier by noncarrier) and 4 (carrier by carrier). Under the null hypothesis the haplotype has no effect on insemination success, so $m_1 - m_4 = 0$, and the contrast $\mathbf{k} = (1, 0, 0, -1)'$ was formed. A t -test was performed using the test statistic $t = \frac{\mathbf{k}'\hat{\mathbf{m}}}{\sqrt{\mathbf{k}'\hat{\mathbf{V}}\mathbf{k}}}$, where $\hat{\mathbf{m}}$ was the vector of mating type effects and $\hat{\mathbf{V}}$ was the estimated error covariance matrix for the effects of mating types. Under the null hypothesis, the test statistic t follows a Student's t distribution with 1 df.

Identification of TH-Associated Variants. The causative factors that underlie detected TH could be a point mutation, an indel, or a structural variant such as a large deletion. In case there is a recessive genetic factor that causes extra conception failures, we expected (1) that the variants will never be observed in the homozygous state; (2) that the causal factor will be located within or in close proximity (± 1 Mb) to the TH;

and (3) that if the TH tags the causal variant perfectly, then the animals carrying the TH will be heterozygous for the causal variant, whereas animals not carrying the TH will be homozygous of the reference allele for the causal variant. Therefore, we checked whether any SNP or indel within detected TH meeting the above conditions using 2,333 animals (130 Holstein animals overlap 50K genotype data) in *run6* data of the 1000 Bull Genomes Project (Daetwyler et al., 2014).

Large chromosomal deletion can also be responsible for lethality. The SNP within a deletion often deviate from Hardy-Weinberg proportion and show runs of homozygosity. To identify deletions, we looked for such features on whole-genome sequence SNP within each HDR, along with 1 Mb on both sides. In addition, we used the population-scale structural variation discovery tool “Genome STRucture in Populations” (GenomeStrip-2.00.1678; Handsaker et al., 2011) for identifying large deletions ($100 \text{ bp} \leq \text{size} \leq 1 \text{ Mb}$) in the 1000 Bull Genomes Project data of 67 Nordic Holstein samples (60 bulls + 7 cows). Genotype calling for discovered deletions was performed using GenomeStrip’s “SVGenotyper” module, and low-quality deletion calls were filtered (for details, see Mesbah-Uddin et al., 2017). We then checked the concordance between the TH carrier status and deletion carrier status for 55 animals with both sequence and 50K genotype data. Here, a deletion was selected as concordant with the TH only when all the carrier animals of the TH were also the carrier for the deletion and vice versa.

RESULTS

The proportion of genotypes per individual with nonmissing data was 0.95. These sporadic missing genotypes were imputed using Beagle software. We examined the imputation accuracy by dropping 3,000 makers from 2,000 animals and imputed those using the rest of the genotyped animals as reference. The imputation accuracy was 96%. Based on detected homozygous-deficient haplotypes, we grouped 29 HDR based on their relative proximity (Supplemental Table S1, <https://doi.org/10.3168/jds.2019-16651>). Then, we shortlisted 29 TH with 26 to 69 expected homozygotes based on the haplotype frequencies, but only 0 to 3 homozygotes were observed. The carrier frequencies of these TH varied from 3.12 to 5.11%. Of the 29 detected TH, 2 TH on BTA8 and BTA21 have been previously reported. Association analyses on 4 NRR phenotypes validated 2 TH that showed a harmful effect with at-risk mating (Table 1); 7 TH exhibited a complete absence of homozygotes (Table 2); and at-risk mating of the remaining 18 TH did not significantly lower the insemination success, and at least some homozygotes

were observed. Moreover, among 130 Holstein cattle common both in genotype data and the 1000 Bull Genomes Project, 0 to 15 carriers were detected in the 29 TH. However, we could not validate any point mutation or indel concordance with TH carrier status. We also checked for deviations from Hardy-Weinberg proportion among all SNP in the pre-edited 50K genotype data within each HDR. The minimum P -value for SNP within 29 HDR was 3.81×10^{-30} on BTA21. Comparing animals carrying detected deletions using GenomeStrip with TH carrier, we validated only the previously identified deletion causing brachyspina (Charlier et al., 2012). No other large deletions inside the HDR were concordant with TH carrier status.

2 Known Recessive Lethal Haplotypes Segregating in Nordic Holsteins

HH3. An HDR at 94.6 to 96.3 Mb on BTA8 was detected in Nordic Holsteins. The HDR was overlapped with HH3, which was first reported by VanRaden et al. (2011) for early embryonic losses. Subsequently, McClure et al. (2014) identified a nonsynonymous SNP rs456206907 (g.95410507T>C in UMD 3.1) in the *SMC2* gene as the likely causative mutation. With a carrier frequency of 3.42%, the TH at 95,003,606–95,498,189 bp on BTA8 exhibited a complete absence of homozygotes. Association analysis for insemination records showed that at-risk matings had significantly lower NRR at 35, 56, 100, and 150 d compared with mating type 1 (Figure 1A). In the 1000 Bull Genome Project data, 130 Holstein animals were also present in the genotyped animals. Of these, 7 were carriers and 123 were noncarriers for the TH. If the TH had tagged the causal variant perfectly, we would expect all 7 carriers to be heterozygous for the TH and none of 123 noncarriers to carry the variant allele. Indeed, the 7 TH carriers were heterozygotes for rs456206907, whereas only 1 animal of the 123 noncarriers was a heterozygote. It is worth noting that we also detected another haplotype at 95,563,701–96,166,978 bp close to rs456206907 with 0 homozygotes observed, whereas 30 were expected (lowest P -value for this HDR). At-risk matings had significantly lower NRR than mating type 1. However, carrier status of this haplotype was not concordant with the causal variant (rs456206907) for HH3. The carrier correlation between this 2 adjacent TH (95,003,606–95,498,189 bp and 95,563,701–96,166,97 bp) was -0.035 . We tend to believe that 95,563,701–96,166,97 bp was a false-positive TH.

Brachyspina. A previous report has shown that a 3.3-kb deletion in the *FANCI* gene is responsible for the brachyspina syndrome causing fetal death in Holstein cattle (Charlier et al., 2012). On BTA21, we

Table 1. Recessive lethal haplotypes for embryonic loss

BTA	Position (bp)		Frequency (%)	Homozygous (no.)		P -value	t -test ¹ (P -value)				Carriers (no.)
	Start	End		Observed	Expected		NRR35	NRR56	NRR100	NRR150	
2	35,478,963	36,142,577	3.39	3	30	8.24E-07	1.66E-02*	6.81E-02	7.63E-02	2.17E-01	0
5	6,942,103	7,832,521	3.35	2	29	5.34E-07	2.38E-02*	4.45E-04**	1.18E-06**	8.29E-06**	9

¹ t -test for nonreturn rate (NRR) at 35, 56, 100, and 150 d.

* $P < 0.05$; ** $P < 0.01$.

Table 2. Recessive lethal haplotypes with complete absence of homozygotes

BTA	Position (bp)		Frequency (%)	Homozygous (no.)		P-value	t-test ¹ (P-value)				Carriers (no.)
	Start	End		Observed	Expected		NRR35	NRR56	NRR100	NRR150	
3	70,592,505	71,077,825	3.36	0	30	4.32E-08	1.97E-01	3.88E-01	5.49E-01	6.03E-01	0
6	496,589	818,864	3.68	0	36	1.97E-09	9.79E-01	5.71E-01	9.08E-01	8.91E-01	5
9	24,127,805	25,323,880	3.19	0	27	2.03E-07	1.60E-02*	1.97E-02*	1.05E-03**	9.92E-04**	15
9	58,507,530	59,472,770	3.55	0	33	9.22E-09	4.32E-01	1.51E-01	4.40E-01	3.91E-01	7
10	26,479,738	27,644,999	3.12	0	26	3.41E-07	5.44E-01	7.02E-01	6.40E-01	5.22E-01	1
15	47,739,031	48,068,620	3.16	0	26	3.41E-07	6.76E-01	7.01E-01	6.12E-01	6.15E-01	5
17	32,964,041	33,317,888	3.28	0	28	1.21E-07	6.74E-07**	2.24E-03**	5.44E-02	2.72E-01	8

¹t-test for nonreturn rate (NRR) at 35, 56, 100, and 150 d.

*P < 0.05; **P < 0.01.

detected an HDR at 19.8 to 21.2 Mb in Nordic Holsteins. With variable haplotype length, the haplotype at 20,248,628–21,117,869 bp on BTA21 with a carrier frequency of 3.11% exhibited a complete absence of homozygotes. At-risk matings of this haplotype had significantly lower NRR at 56, 100, and 150 d compared with mating type 1 (Figure 1B). We detected 7 animals carrying the deletion at 21,184,871–21,188,198 bp by GenomeStrip. Among the 55 animals that overlapped with our genotyped animals, 4 deletion carriers were also TH carriers, and 51 deletion noncarriers included 1 TH carrier and 50 TH noncarriers.

Figure 1. At-risk mating of 2 known lethal haplotypes, structural maintenance of chromosomes (Chr) 2 (HH3) and brachyspina, segregating in Nordic Holsteins. (A) The haplotype Chr8: 95,003,606–95,498,189 bp. (B) The haplotype Chr21: 20,248,628–21,117,869 bp. The y-axis shows the relative difference between at-risk mating and mating type 1 (noncarrier bull mated to a daughter of a noncarrier sire), and the x-axis shows the nonreturn rate (NRR) at 35, 56, 100, and 150 d postinsemination.

2 Novel Lethal Haplotypes Negatively Affecting NRR

Two putative novel TH for recessive lethal haplotypes besides HH3 and brachyspina were detected in Nordic Holsteins, causing early embryonic death (Table 1). The frequencies of the 2 haplotypes were 3.39 and 3.35%. We compared the effect of at-risk matings with mating type 1, which lead to different probabilities of reproductive failure using the logistic regression model. At-risk mating of these 2 TH displayed harmful effects with significantly lower ($P < 0.05$) NRR, supporting the segregation of embryonic lethal alleles (Table 1; Figure 2). The TH at 6,942,103–7,832,521 bp on BTA5 displayed significant harmful effect on NRR at 35, 56, 100, and 150 d. The TH at 35,478,963–36,142,577 bp on BTA2 displayed a significant harmful effect only on early NRR. We found hundreds of functional variants within these 2 TH in the whole-genome sequence data. However, none of the variants were in perfect concordance with the corresponding TH carrier status.

7 Novel Lethal Haplotypes Displaying Complete Absence of Homozygotes

An additional 7 TH displayed a complete absence of homozygotes in Nordic Holsteins (Table 2). The average frequency of these haplotypes was 3.34%. Comparing the effect of at-risk matings with mating type 1 produced no evidence of increased early embryo death. Of the 29 HDR, we detected 2 adjacent large deletions using sequence read depth—one at Chr14: 64,696,254 to 64,700,546 (4,291 bp) and the other at Chr14: 70,227,428 to 70,235,281 (7,851 bp)—but neither were concordant with corresponding TH carrier status.

DISCUSSION

In dairy cattle, the inbreeding rate has increased due to intense use of a limited number of elite bulls for AI. Subsequently, the frequency of some deleterious alleles has increased, thus favoring the expression of recessive harmful mutations, including embryonic lethal alleles. Therefore, it is valuable for cattle selection programs to identify alleles or haplotypes that are associated with mortality traits. Previously, case-control association analysis was used widely to locate the target region on the genome containing the recessive lethal variant (Pollott, 2018). However, the phenomenon of embryo loss is difficult to detect, and the genotype for affected embryos is almost impossible to collect. Because more and more SNP data are available following genomic prediction, we have the chance to identify alleles or haplotypes that are associated with embryo loss using large genomic data. In this study, using genotype

Figure 2. At-risk mating of 2 novel lethal haplotypes segregating in Nordic Holsteins. (A) The haplotype Chr2: 35,478,963–36,142,577 bp. (B) The haplotype Chr5: 6,942,103–7,832,521 bp. The y-axis shows the relative difference between at-risk mating and mating type 1 (noncarrier bull mated to a daughter of a noncarrier sire), and the x-axis shows the nonreturn rate (NRR) at 35, 56, 100, and 150 d post-insemination. Chr = chromosome.

data, we confirmed the segregation of 2 known recessive lethal mutations, HH3 (VanRaden et al., 2011) and brachyspina syndrome (Charlier et al., 2012), in Nordic Holsteins.

We identified 27 novel TH that exhibited significant depletion of homozygous haplotype. Of the 27 new TH, 2 TH were close to genes related to prenatal death in mice. The *Atp6v1c1* gene causes death before the appearance of tooth buds and organogenesis in mice (Dickinson et al., 2016). Although *Atp6v1c1* is not located in the TH Chr14: 69,880,924–71,228,051 bp, it is within HDR Chr14: 53,082,514–71,256,147 bp. Another previously reported gene, *Ilk*, can cause embryonic lethality before organogenesis and preweaning lethality

in mouse (Dickinson et al., 2016). This gene is located within HDR Chr15: 46,443,525–48,068,620 bp. Using 50K genotype data for 7,937 Nordic Holstein bulls, Sahana et al. (2013) identified 17 homozygote-deficient haplotypes and loosely clustered them into 8 genomic regions harboring possible recessive lethal alleles. Only 1 genomic region on BTA8 (85–89 Mb) overlapped with our study. In this study, we had a larger population size and used both bull and cow records. It is indicated that sample size has a big effect on the haplotype detection method (VanRaden et al., 2011; Mesbah-Uddin et al., 2019).

We compared the NRR among at-risk matings with mating type 1 (see Materials and Methods) to ascertain whether at-risk mating increases the chance that the cow returned to estrus, an indication of early embryonic death. Considering the TH carrier status of the mating bull and cow in a logistic regression model, we expected that the at-risk matings of TH carriers would show harmful effects on NRR compared with mating type 1. Of the 27 new TH, 2 TH showed significant harmful effects on NRR, and 25 did not significantly lower insemination success. Therefore, it is not convincing that these TH were responsible for early embryonic losses (Table 2; Supplemental Table S2, <https://doi.org/10.3168/jds.2019-16651>). However, not all TH effects can be reflected in NRR due to factors such as frequency of at-risk matings and timing of embryo losses. In our data set, the distribution of frequencies of 4 mating types was very skewed. Among 2,402,533 insemination records analyzed, on average, 88.13% were of mating type 1, 5.80% were of mating type 2, 5.61% were of mating type 3, and only 0.44% were of mating type 4 (i.e., at-risk mating). The average NRR was 0.73 at 35 d, 0.59 at 56 d, 0.49 at 100 d, and 0.46 at 150 d after insemination. On the other hand, the timing of embryonic loss is critical. For example, if the embryo is lost between 35 and 56 d after insemination, then at-risk matings will show a harmful effect on NRR only at time points later than 35 d after insemination. If embryo loss happens more than 150 d after insemination, there will be no effect on the NRR at the time points observed. However, reproduction would still be affected.

Seven out of 27 novel TH (Table 2) exhibited strong signals. Homozygotes for the haplotypes were absent. However, no evidence of early embryonic loss among at-risk matings for these haplotypes was found. Presumably, these haplotypes affect cattle mortality later than 150 d postinsemination.

In this study, we checked homozygotes missing for all the variants within 2 novel TH that negatively affect NRR (Table 1) and 7 novel TH that displayed a complete absence of homozygotes in the 1000 Bull Genome Project data. None of these variants were concordant

with the corresponding TH carrier status. One possible reason is that detecting the causal variant underlying each TH relies strongly on the linkage disequilibrium pattern between TH and the causal variant. But complete linkage disequilibrium between the causal variant and the TH is very unlikely in real situations if 2 different copies of the haplotype, one with the causal mutation and the other with the ancestral allele, are segregating in the population. Another possible reason for failure to detect causative variants is that we have fewer carriers due to the limited number of sequenced animals in the 1000 Bull Genomes Project that were also included among the genotyped animals. Therefore, additional analyses are needed to detect the short sequence variants or large deletions within HDR regions.

CONCLUSIONS

Using genomic data, we detected 29 HDR of putative recessive lethal haplotypes based on significant depletion of homozygotes, including 2 known ones (HH3 and brachyspina). Out of the novel ones, 2 TH had harmful effects on NRR and 7 TH showed a complete absence of homozygotes. Haplotype carrier status could be used to select bulls to reduce the frequencies of latent lethal mutations in the population. If desired, at-risk matings could be avoided. Hundreds of putative causal variants were located within these identified TH, but none were in perfect concordance with carrier status for the corresponding TH. To facilitate identification of the causal factor within each TH, a sizable number of bulls carrying the TH would need to be whole-genome sequenced.

ACKNOWLEDGMENTS

We are grateful to NAV (Aarhus, Denmark) for providing the phenotypic data used in this study and Viking Genetics (Randers, Denmark) for providing blood and semen samples for genotyping. This work was supported by the research project “Identification and control of recessive mutations” funded by the Milk Levy Fund (Aarhus, Denmark) and the GUDP project “LiveCalf” (no. 34009-16-1101) from the Ministry of Environment and Food of Denmark (Copenhagen). Md Mesbah-Uddin acknowledges the European Commission’s Erasmus-Mundus joint doctorate “EGS-ABG” program and the Center for Genomic Selection in Animals and Plants (GenSAP) funded by Innovation Fund Denmark (Copenhagen; grant 0603-00519B). The 1000 Bull Genomes Project is kindly acknowledged for sharing the whole-genome sequence data. The authors declare no competing interests. GS, XW, BG, and MSL conceived and designed the study; XW and MMU analyzed the data; XW wrote the paper; and MSL, GS,

and BG contributed materials and analysis tools. All authors read, revised, and approved the final manuscript.

REFERENCES

- Browning, S. R., and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–1097. <https://doi.org/10.1086/521987>.
- Charlier, C., J. S. Agerholm, W. Coppieters, P. Karlsskov-Mortensen, W. Li, G. de Jong, C. Fasquelle, L. Karim, S. Cirera, N. Cambisano, N. Ahariz, E. Mullaart, M. Georges, and M. Fredholm. 2012. A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachypina. *PLoS One* 7:e43085. <https://doi.org/10.1371/journal.pone.0043085>.
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. van Binsbergen, R. F. Brøndum, X. Liao, A. Djari, S. C. Rodriguez, C. Grohs, D. Esquerré, O. Bouchez, M.-N. Rossignol, C. Klopp, D. Rocha, S. Fritz, A. Eggen, P. J. Bowman, D. Coote, A. J. Chamberlain, C. Anderson, C. P. VanTassell, I. Hulsege, M. E. Goddard, B. Guldbrandtsen, M. S. Lund, R. F. Veerkamp, D. A. Boichard, R. Fries, and B. J. Hayes. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* 46:858–865. <https://doi.org/10.1038/ng.3034>.
- Dickinson, M. E., A. M. Flenniken, X. Ji, L. Teboul, M. D. Wong, J. K. White, T. F. Meehan, W. J. Weninger, H. Westerberg, H. Adissu, C. N. Baker, L. Bower, J. M. Brown, L. B. Caddle, F. Chiani, D. Clary, J. Cleak, M. J. Daly, J. M. Denegre, B. Doe, M. E. Dolan, S. M. Edie, H. Fuchs, V. Gailus-Durner, A. Galli, A. Gambadoro, J. Gallegos, S. Guo, N. R. Horner, C. W. Hsu, S. J. Johnson, S. Kalaga, L. C. Keith, L. Lanoue, T. N. Lawson, M. Lek, M. Mark, S. Marshall, J. Mason, M. L. McElwee, S. Newbidding, L. M. J. Nutter, K. A. Peterson, R. Ramirez-Solis, D. J. Rowland, E. Ryder, K. E. Samocha, J. R. Seavitt, M. Selloum, Z. Szoke-Kovacs, M. Tamura, A. G. Trainor, I. Tudose, S. Wakana, J. Warren, O. Wendling, D. B. West, L. Wong, A. Yoshiki, International Mouse Phenotyping Consortium. Jackson Laboratory, Infrastructure Nationale PHENOMIN, Institut Clinique de la Souris (ICS), Charles River Laboratories, M. R. C. Harwell, Toronto Centre for Phenogenomics, Wellcome Trust Sanger Institute, RIKEN BioResource Center, D. G. MacArthur, G. P. Tocchini-Valentini, X. Gao, P. Flicek, A. Bradley, W. C. Skarnes, M. J. Justice, H. E. Parkinson, M. Moore, S. Wells, R. E. Braun, K. L. Svenson, M. H. de Angelis, Y. Herault, T. Mohun, A. M. Mallon, R. Mark Henkelman, S. D. M. Brown, D. J. Adams, K. C. Kent Lloyd, C. McKelrie, A. L. Beaudet, M. Bucan, and S. A. Murray. 2016. High-throughput discovery of novel developmental phenotypes. *Nature* 537:508–514. <https://doi.org/10.1038/nature19356>.
- Fritz, S., A. Capitan, A. Djari, S. C. Rodriguez, A. Barbat, A. Baur, C. Grohs, B. Weiss, M. Boussaha, D. Esquerré, C. Klopp, D. Rocha, and D. Boichard. 2013. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS One* 8:e65550. <https://doi.org/10.1371/journal.pone.0065550>.
- Handsaker, R. E., J. M. Korn, J. Nemesh, and S. A. McCarroll. 2011. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nat. Genet.* 43:269–276. <https://doi.org/10.1038/ng.768>.
- Kadri, N. K., G. Sahana, C. Charlier, T. Iso-Touru, B. Guldbrandtsen, L. Karim, U. S. Nielsen, F. Panitz, G. P. Aamand, N. Schulman, M. Georges, J. Vilkkil, M. S. Lund, and T. Druet. 2014. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: Additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genet.* 10:e1004049. <https://doi.org/10.1371/journal.pgen.1004049>.
- Madsen, P., J. Jensen, R. Labouriau, O. F. Christensen, and G. Sahana. 2014. DMU—A package for analyzing multivariate mixed models in quantitative genetics and genomics. Page 699 in *Proc. 10th World Congress on Genetics Applied to Livestock Production*, vol. Methods and Tools: Statistical methods—Linear and nonlinear models (Posters). Vancouver, BC, Canada.
- McClure, M. C., D. Bickhart, D. Null, P. Vanraden, L. Xu, G. Wiggins, G. Liu, S. Schroeder, J. Glasscock, J. Armstrong, J. B. Cole, C. P. Van Tassell, and T. S. Sonstegard. 2014. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in SMC2 for HH3. *PLoS One* 9:e92769. <https://doi.org/10.1371/journal.pone.0092769>.
- Mesbah-Uddin, M., C. Hoze, P. Michot, A. Barbat, R. Lefebvre, M. Boussaha, G. Sahana, S. Fritz, D. Boichard, and A. Captian. 2019. A missense mutation (p. Tyr452Cys) in CAD gene compromises reproductive success in French Normande cattle. *J. Dairy Sci.* 102:6340–6356. <https://doi.org/10.3168/jds.2018-16100>.
- Mesbah-Uddin, M., B. Guldbrandtsen, T. Iso-Touru, J. Vilkkil, D. J. De Koning, D. Boichard, M. S. Lund, and G. Sahana. 2017. Genome-wide mapping of large deletions and their population-genetic properties in dairy cattle. *DNA Res.* 25:49–59. <https://doi.org/10.1093/dnares/dsx037>.
- Pollott, G. E. 2018. Invited review: Bioinformatic methods to discover the likely causal variant of a new autosomal recessive genetic condition using genome-wide data. *Animal* 12:2221–2234. <https://doi.org/10.1017/S1751731118001970>.
- Sahana, G., U. S. Nielsen, G. P. Aamand, M. S. Lund, and B. Guldbrandtsen. 2013. Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. *PLoS One* 8:e82909. <https://doi.org/10.1371/journal.pone.0082909>.
- Schwarzenbacher, H., J. Burgstaller, F. R. Seefried, C. Wurmser, M. Hilbe, S. Jung, C. Fuerst, N. Dinopl, and H. Weissenböck. 2016. A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. *BMC Genomics* 17:400. <https://doi.org/10.1186/s12864-016-2742-y>.
- Sonstegard, T. S., J. B. Cole, P. M. VanRaden, C. P. Van Tassell, D. J. Null, S. G. Schroeder, D. Bickhart, and M. C. McClure. 2013. Identification of a nonsense mutation in CWC15 associated with decreased reproductive efficiency in Jersey cattle. *PLoS One* 8:e54872. <https://doi.org/10.1371/journal.pone.0054872>.
- University of Sydney. 2019. Online Mendelian inheritance in animals. Accessed Jan. 8, 2019. <http://omia.org/>.
- 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. <https://doi.org/10.3168/jds.2011-4624>.
- 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. <https://doi.org/10.1186/gb-2009-10-4-r42>.

ORCID

- Xiaoping Wu <https://orcid.org/0000-0002-7844-1397>
Md Mesbah-Uddin <https://orcid.org/0000-0003-1846-0411>
Bernt Guldbrandtsen <https://orcid.org/0000-0003-1764-135X>
Goutam Sahana <https://orcid.org/0000-0001-7608-7577>