

Human-Mediated Introgression of Haplotypes in a Modern Dairy Cattle Breed

Qianqian Zhang,^{*,†,1,2} Mario P. L. Calus,[†] Mirte Bosse,[†] Goutam Sahana,^{*} Mogens Sandø Lund,^{*} and Bernt Gulbrandsen^{*}

^{*}Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Denmark and

[†]Animal Breeding and Genomics, Wageningen University & Research, The Netherlands

ORCID IDs: 0000-0002-4423-1236 (Q.Z.); 0000-0002-3213-704X (M.P.L.C.); 0000-0001-7608-7577 (G.S.)

ABSTRACT Domestic animals can serve as model systems of adaptive introgression and their genomic signatures. In part, their usefulness as model systems is due to their well-known histories. Different breeding strategies such as introgression and artificial selection have generated numerous desirable phenotypes and superior performance in domestic animals. The modern Danish Red Dairy Cattle is studied as an example of an introgressed population. It originates from crossing the traditional Danish Red Dairy Cattle with the Holstein and Brown Swiss breeds, both known for high milk production. This crossing happened, among other things due to changes in the production system, to raise milk production and overall performance. The genomes of modern Danish Red Dairy Cattle are heavily influenced by regions introgressed from the Holstein and Brown Swiss breeds and under subsequent selection in the admixed population. The introgressed proportion of the genome was found to be highly variable across the genome. Haplotypes introgressed from Holstein and Brown Swiss contained or overlapped known genes affecting milk production, as well as protein and fat content (*CD14*, *ZNF215*, *BCL2L12*, and *THRSP* for Holstein origin and *ITPR2*, *BCAT1*, *LAP3*, and *MED28* for Brown Swiss origin). Genomic regions with high introgression signals also contained genes and enriched QTL associated with calving traits, body confirmation, feed efficiency, carcass, and fertility traits. These introgressed signals with relative identity-by-descent scores larger than the median showing Holstein or Brown Swiss introgression are mostly significantly correlated with the corresponding test statistics from signatures of selection analyses in modern Danish Red Dairy Cattle. Meanwhile, the putative significant introgressed signals have a significant dependency with the putative significant signals from signatures of selection analyses. Artificial selection has played an important role in the genomic footprints of introgression in the genome of modern Danish Red Dairy Cattle. Our study on a modern cattle breed contributes to an understanding of genomic consequences of selective introgression by demonstrating the extent to which adaptive effects contribute to shape the specific genomic consequences of introgression.

KEYWORDS selective introgression; signature of selection; high-yielding cattle breeds; modern dairy cattle breed

PROCESSES of adaptive introgression are complex, and their genomic signatures in human and other species have been studied extensively (Hasenkamp *et al.* 2015; Deschamps *et al.* 2016; Figueiró *et al.* 2017; Jagoda *et al.* 2017). Genome analysis has enabled an in-depth assessment of the genomic

consequences of different demographic processes including introgression, selection, and their interplay in the modern species (Deschamps *et al.* 2016; Figueiró *et al.* 2017). Domestic animals can serve as model organisms for these processes. They have several advantages in understanding the impact of introgression and selection on genomes: first, introgression and selection is known to occur between breeds, and the processes are often well documented by breeders; second, massive data are routinely collected before and after introgression and selection, such as parentage, genotypes, and phenotypes; and third, under controlled appropriate environmental conditions, a large part of genomic consequence is caused by human-mediated directional introgression and selection.

Copyright © 2018 by the Genetics Society of America

doi: <https://doi.org/10.1534/genetics.118.301143>

Manuscript received March 28, 2018; accepted for publication May 24, 2018; published Early Online May 30, 2018.

Available freely online through the author-supported open access option.

Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6383528>.

¹Present address: Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 3, 1870, Frederiksberg C, Copenhagen, Denmark

²Corresponding author: Aarhus University, Blichers alle 20, Tjele, 8830, Denmark
E-mail: Qianqian.zhang@mbg.au.dk

Artificial selection and different breeding strategies have enabled the generation of numerous desirable phenotypes in domestic animals such as cattle (Hartwig *et al.* 2015; Buzanskas *et al.* 2017; Davis *et al.* 2017), pigs (Bosse *et al.* 2014a,b; Ai *et al.* 2015), and dogs (Galov *et al.* 2015; von-Holdt *et al.* 2016). Strategies including crossbreeding and introgression have been very successful in improving productivity and performance in domestic animals. For example, Chinese pig breeds were imported to Europe to improve the productivity of European pigs in the late 18th and early 19th centuries (White 2011). The fertility-related traits have been largely improved by the crossbreeding and introgression from Asian pigs (Merks *et al.* 2012; Bosse *et al.* 2014, 2015). Similarly, in dairy cattle, crossbreeding and following introgression between local breeds with other breeds have been applied to achieve better productivity and performance (Davis *et al.* 2017). The genetic architecture of modern domestic animals, including dairy cattle, is shaped by the interplay between different forces, including the intentional introduction of favorable alleles from other breeds, subsequent selection for favorable introgressed alleles, and demographic processes.

Since the introduction of scientific breeding, the main breeding goal in cattle has been to improve milk yield, fat, and protein content, even for dual-purpose breeds in intensive farming systems, especially in Europe (Hartwig *et al.* 2015). By the crossing of high-yielding breeds with local breeds, the productivity of local breeds could rapidly be increased at the expense of the genetic distinctiveness of the local breeds. The high productivity of these admixed breeds was further improved by intense selection, resulting in increased frequencies or even fixation of favorable alleles. Many of the variants thus spreading in the population will have been of introgressed origin. With the great success of cattle breeding including crossbreeding and because of the availability of large-scale genomic data sets, analysis of admixed local cattle breeds represents an appealing model to identify the genomic consequences of admixture with the intention of improving traits of interest in local breeds. We expect that the introgressed genomic regions play an important role in improving productivity and performance in the crossbreeds. We hypothesize that: (1) the genome-wide introgressed regions are nonrandomly distributed across the genome with respect to their genomic locations; (2) the majority of highly introgressed regions affect production and performance traits; and (3) that these introduced haplotypes are or have been under selection in the admixed breed.

To test these hypotheses and identify specific genomic regions and genes of interest involved in the important traits from high-yield breeds such as Holstein (HOL) and Brown Swiss (BSW), we use the hybrid modern Danish Red Dairy Cattle (mRDC) breed as an example. Our analyses illustrate the patterns of introgressed and selected haplotypes in an admixed local breed. The hybrid mRDC originates from traditional Danish Red Dairy Cattle (tRDC). In recent decades, HOL and BSW have been used extensively to improve the milk

yield of mRDC (Kantanen *et al.* 2000). Years of crossbreeding and selection have led to the differentiation between mRDC and tRDC. The introgression of and selection for haplotypes from HOL and BSW have probably made a significant contribution to the increased milk production level of mRDC. The objective of this study was to examine genomic patterns of introgression from two high-yielding breeds (HOL and BSW) in a modern dairy breed (*i.e.*, the hybrid mRDC), and unravel the consequences of introgression and selection at the genome level using whole-genome sequence data.

Methods

SNP genotyping, sequencing, variant calling, and quality control

Whole-genome sequence data were available for 97 animals from four breeds (32 HOL, 20 BSW, 15 tRDC, and 30 mRDC). All individuals' genomes were sequenced to $\sim 10\times$ of depth or deeper using Illumina paired-end sequencing. Reads were aligned to the cattle genome assembly UMD3.1 (Zimin *et al.* 2009) using *bwa* (Li and Durbin 2009). Aligned sequences were converted to raw BAM files using *samtools* (Li *et al.* 2009). Duplicate reads were marked using the *samtools* *mdup* option (Li *et al.* 2009). The Genome Analysis Toolkit (McKenna *et al.* 2010) was used for local realignment around insertion/deletion (indel) regions and recalibration following the 1000 Bull Genome Project guidelines (Daetwyler *et al.* 2014) incorporating information from dbSNP (Sherry *et al.* 2001). Finally, variants were called using the Genome Analysis Toolkit's UnifiedGenotyper (McKenna *et al.* 2010), which simultaneously calls short indels and SNPs. Indels and variants on sex chromosomes were excluded from further analyses.

Population structure and admixture

Using *PLINK* (Chang *et al.* 2015), the sequence variants were pruned to remove markers with pairwise linkage disequilibrium (LD) > 0.1 with any other SNP within a 50-SNP sliding window (advancing by 10 SNPs at a time). The SNPs on all the autosomes were used to study the population structure. To get an overview of population structure of the genotyped animals from different breeds, the whole-genome sequence data were used for principal component (PC) estimation using *GCTA* (Yang *et al.* 2011). Admixture analyses were done using the program *Admixture* (Alexander *et al.* 2009) with values of K between 2 and 10. The K value with a low cross validation error was chosen.

To investigate the statistical significance of admixture among these cattle populations, *TreeMix* software was used to perform the three population (f_3) test (Pickrell and Pritchard 2012). In the f_3 test with the form of f_3 (A; B, C), an extreme negative f_3 statistic indicates that there has been significant gene flow to A from populations B and C. The combination of two of HOL, BSW, and tRDC (A and B) were used as source populations and mRDC were used as admixed

population (C) in the f_3 test, which results in three different combinations.

We also calculated the breed proportions for the sequenced mRDC individuals using pedigree. The full pedigrees of the sequenced mRDC animals were extracted and used to infer the breed proportions by coding the breed where the ancestors enter as the known parents of the traced animal.

Introgression mapping

Calculation of ancestry dosages in mRDC: A novel two-layer hidden Markov model was implemented in the method developed in Guan (2014) to infer the structure of local haplotypes introgressed from HOL, BSW, and tRDC in mRDC. The software package ELAI developed by Guan (2014) was used to infer the ancestry dosages of the haplotypes from three source populations in mRDC. The SNPs from sequence variants with minor allele frequency < 0.01 or with a missing proportion > 0.05 were removed from subsequent analyses. The option of three-way admixture with admixture generations of 10 was chosen, which approximates the history of admixture in the mRDC population. Thirty steps of an expectation-maximization (EM) algorithm were run to infer the ancestry dosages of HOL, tRDC, and BSW in mRDC.

Relative identity-by-descent scores: HOL, BSW, and tRDC have made large genetic contributions to the mRDC. Therefore, the sequenced mRDC, tRDC, BSW, and HOL were selected for introgression mapping analysis. The identity-by-descent (IBD) regions comparing between mRDC and tRDC were used as a reference to map the introgression regions from HOL and BSW using a pairwise comparison between these breeds. Following Bosse *et al.* (2014), sequences for 29 autosomes were first phased using Beagle fastIBD (version 3.3.2) (Browning and Browning 2007). Pairwise comparisons for detecting IBD were performed between mRDC and tRDC, mRDC and HOL, and mRDC and BSW. As recommended in the Beagle documentation (Browning and Browning 2007), 10 independent runs for phasing and pairwise IBD detection were performed. The identified IBD segments were combined from 10 runs and the threshold parameter compromising between power and false discovery rate was 10^{-10} for identifying the true shared IBD, as suggested by Browning and Browning (2007). We defined the IBD score as the proportion of the number of recorded true IBD haplotype segments over the total number of pairwise comparisons using a window of 10 kb. The IBD score was calculated for each pairwise comparison using a custom perl script. To quantify the relative proportion of introgressed genome from HOL or BSW, we calculated the relative IBD score (rIBD) as follows: IBD score (mRDC and tRDC) – IBD score (mRDC and HOL) or IBD score (mRDC and BSW). Thus, the rIBD has values in the range of -1 to 1 . $rIBD = 1$ signifies that all haplotypes in the target breed originate from the first source breed, while $rIBD = -1$ signifies 100% from the second source breed. The variance of rIBD score was calculated using a robust method (Williams 2000). The P -values of rIBD scores were derived

from the neutral hypothesis assuming that rIBD are normally distributed with an SD of the squared root of robust estimate of rIBD variance (Bosse *et al.* 2014). The significant introgressed haplotypes from HOL and BSW in mRDC were defined as the haplotypes with a corrected P -value using the Benjamini–Hochberg procedure (Benjamini and Yekutieli 2001) lower than 0.02 and 0.04, respectively. Peterson correlations between rIBD scores from HOL or BSW introgression larger than the median of rIBD scores, and the corresponding test statistics from signatures of selection analyses, were calculated. Moreover, the putative significant introgressed signals from HOL and BSW were extracted and compared with the putative significant signals from signatures of selection analyses using a χ^2 test with Monte Carlo simulation (Hope 1968) implemented in calculating P -values.

Detection of signature of selection

F_{st} analysis: The genetic differentiation between individuals from tRDC and mRDC was measured by pairwise F_{st} analysis following Weir and Cockerham (1984). Pairwise F_{st} was computed with Genepop 4.2 in bins of 10 kb over the full length of the genome (Weir and Cockerham 1984). The correlations between the F_{st} and rIBD scores for HOL and BSW introgression for the same bins of 10 kb were calculated. The R package FlexMix (Grün and Leisch 2008) was used to fit a series of finite normal mixture models to explore the underlying structure of the distribution of F_{st} presumably caused by different underlying evolutionary or demographic processes in the populations such as balancing, directional selection, and neutrality. This mixture model postulated that $\overline{Fst} : \sum_{k=1}^k \pi_k N(\overline{Fst}_l | \mu_k, \sigma_k^2)$, where k is the number of components of the mixture, l is the l th locus, π_k is the probability that \overline{Fst} belongs to cluster k , and μ_k and σ_k^2 are expectation and variance of normal distribution, respectively. The number of components of the mixture (k) was determined by the smallest Akaike's information criterion among different models with different numbers of components. Model parameters were estimated by maximum likelihood via the EM algorithm in FlexMix. At last, the P -value of the F_{st} combined with the k number of components was calculated as: $p = 1 - \sum_{k=1}^k \pi_k N(\overline{Fst}_l | \mu_k, \sigma_k^2)$. The regions with the cutoff of corrected P -values using Benjamini–Hochberg procedure (Benjamini and Yekutieli 2001) of < 0.05 were used for testing the dependency of putative significant signals between different test statistics.

Extended haplotype homozygosity test: The genome-wide scan for integrated haplotype score (iHS) for mRDC was performed using the R package *rehh* (Sabeti *et al.* 2002; Gautier and Vitalis 2012). The threshold of $|iHS|$ of 2.5 for SNPs was used as the cutoff to define signals of selection signatures (Voight *et al.* 2006; Gautier and Vitalis 2012). To compare iHS with other test statistics such as rIBD scores at the same scale, we calculated the proportion of SNPs with $|iHS| > 2$ in a window size of 10 kb and identified the

windows in the highest 1% of the empirical distribution for proportion of SNPs with $|iHS| > 2$ following Voight *et al.* (2006). The putative significant windows of proportion of SNPs with $|iHS|$ were used to test the dependency of putative significant signals between different test statistics.

Sharing of runs of homozygosity: Runs of homozygosity (ROH) were computed for the sequenced animals to detect shared ROH regions (minimum 10 kb considered) among individuals. For a description of the procedures used for the calculation of nucleotide diversity and detection of ROH, see Zhang *et al.* (2015a,b). The sharing of ROH regions was calculated as the number of individuals sharing the same ROH region on a particular segment using a window of 10 kb bin across the whole genome in mRDC. Regions of enrichment of shared ROH regions were defined as regions exceeding the 95th percentile of the empirical distribution of the number of individuals sharing the same ROH regions in any particular segment.

Gene annotations and enrichment of QTL

Genes in genomic regions showing significant introgression of haplotypes from HOL or BSW in mRDC were annotated. The cattle QTL were extracted from the Animal QTL Database (Hu *et al.* 2016). QTL on the X chromosome, or without locations and references in the list from the Animal QTL Database (Hu *et al.* 2016), were excluded. The remaining QTL were classified into six groups according to the associated traits: milk, reproduction, production, health, meat and carcass, and exterior. The QTL overlapping with genomic regions with high rIBD scores were identified. When two QTL had the same exact genomic interval and the same associated trait groups, they were counted as one QTL. Gene annotations in these regions were retrieved from the Ensembl Genes 89 Database using BioMart (Kinsella *et al.* 2011).

To test for enrichment of QTL in the candidate introgressed regions from HOL or BSW, we applied a permutation test. The candidate introgressed regions were randomly distributed across the whole genome. This permutation did not change the relative proportion and length of introgressed regions to preserve their correlation structure. We then computed the number of QTL and the number of QTL in the six groups of associated traits, which overlapped with the permuted, introgressed regions. In total, 10,000 permutations were performed. The distribution of numbers of QTL observed in the permuted regions were treated as the null distribution from which we computed the significance levels of the number of QTL observed in the real data.

Data availability

The whole-genome sequence data used in this study originated from the 1000 Bull Genome Project. Parts of the whole-genome sequence data of individual bulls of the 1000 Bull Genomes Project [Daetwyler *et al.* (2014)] are available at <https://doi.org/10.1038/s41588-018-0056-5> and National

Center for Biotechnology Information using Sequence Read Archive numbers SRP039339, SRR1188706, SRR1205973, SRR1205973, SRR1205992, SRR1262533, SRR1262536, SRR1262538, SRR1262539, SRR1262614, SRR1262659, SRR1262660, SRR1262788, SRR1262789, SRR1262846, and SRR1293227. Supplemental material available at Figshare: <https://doi.org/10.25386/genetics.6383528>.

Results and Discussion

Population structure and evidence of introgression

Population structures of the sampled cattle breeds in this study were analyzed using Admixture and PC analysis (PCA), as shown in Figure 1. mRDC had contribution from tRDC, HOL, and BSW, as observed in the Admixture analysis (Figure 1). Figure 1 clearly demonstrates the hybrid nature of mRDC cattle, which is consistent with recorded pedigree information of the introgression history of mRDC (Andersen *et al.* 2003). A large contribution to mRDC was observed from tRDC, the recipient population. It is notable that BSW and HOL are two mainstream breeds that each contributed heavily to the genomes of extant mRDC individuals. The proportion of introgression differs quite extensively among individuals. In the PCA analysis (Figure 1), PC1 (6.23% of variance) separated HOL, BSW, and tRDC. However, mRDC was dispersed between the other breeds demonstrating admixture of HOL, BSW, and tRDC, which have made contributions to mRDC. Similarly, PC2 separated HOL, BSW, and tRDC. mRDC had a wide range of PC2 values among the individuals from the other breeds.

Moreover, the breed proportions of mRDC individuals were derived using the recorded full pedigree and there is an average of 0.27 HOL ancestor, 0.17 BSW ancestor, and 0.29 tRDC ancestor in mRDC. The statistical significance of admixture in mRDC was measured using the combination of two populations chosen from HOL, BSW, and tRDC by f_3 test using the program threepop from TreeMix (Pickrell and Pritchard 2012). We observed extreme Z scores from the f_3 test for mRDC using any combination of HOL, BSW, and tRDC, *i.e.*, -14.97 , -11.72 , and -19.24 . These results are consistent with what we have observed from admixture and PCA analysis, and provide statistical significance for the admixture of mRDC from HOL, BSW, and tRDC. These results support that notion that mRDC is indeed a composite breed and can be studied further for introgression from HOL and BSW.

Introgression mapping

Introgression mapping was performed to identify regions in mRDC that contained an excess of introgressed haplotypes. We first examined the local structure of haplotypes in mRDC introgressed from HOL, BSW, and tRDC using a three-way admixture approach (Supplemental Material, Figure S1). We observed average ancestry dosages of 0.78, 0.63, and 0.59 with an SD of 0.24, 0.22, and 0.18 from HOL, tRDC, and BSW, respectively, to mRDC average across the whole genome

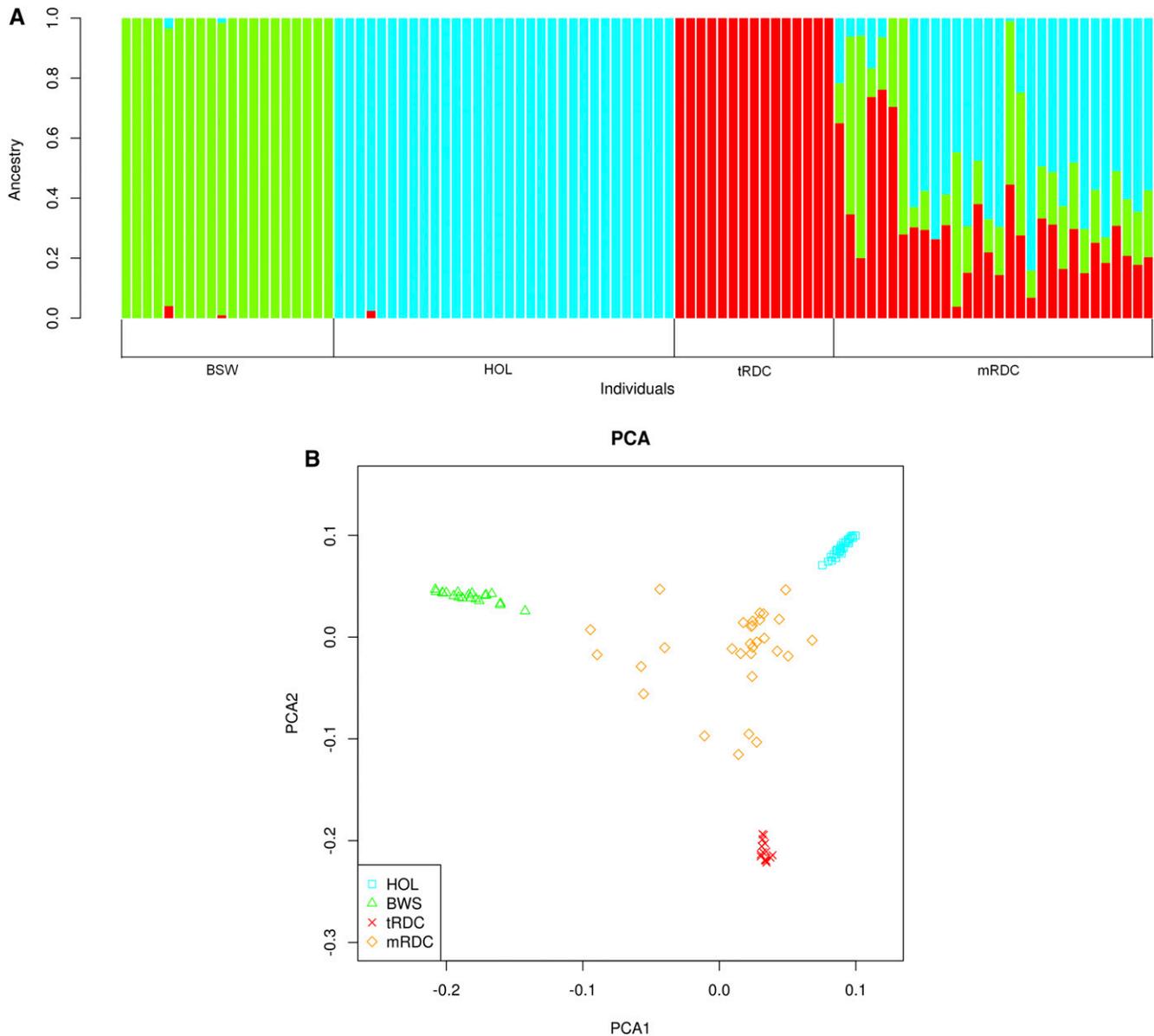


Figure 1 Population structure for four cattle breeds. (A) Admixture analysis of different cattle breeds with $k = 3$. (B) PCA plot among different cattle breeds (principal component 1 vs. principal component 2) Blue squares = HOL, green triangles = BWS, red crosses = tRDC, and yellow diamonds = mRDC. BSW, Brown Swiss; HOL, Holstein; mRDC, modern Red Dairy cattle; PCA, Principal component analysis; tRDC, traditional Red Dairy Cattle.

and all mRDC individuals. Three genomic regions with the ancestry dosage of two were observed from HOL to mRDC (Figure S1). This indicates complete replacement of tRDC genomic material by HOL genomic material. However, there is only one annotated gene (*SNORD116*) in these regions. Interestingly, the genomic region with the full HOL ancestry dosage on chromosome 21 (1,348,427–2,107,179 bp) is associated with gestation length and calving ease (Frischknecht *et al.* 2017). Other QTL overlapping regions with full HOL ancestry in mRDC were associated with bovine respiratory disease susceptibility, body weight, and udder swelling score (Saatchi *et al.* 2014; Keele *et al.* 2015; Michenet *et al.* 2016).

To address which admixed haplotypes have most influenced the existing mRDC breed, we inferred whether a genomic

region was introgressed from high-yield breeds in multiple individuals by examining the contributions proportional to the admixture fraction. The frequencies of all mRDC haplotypes that were of HOL, BSW, or tRDC origin were calculated across the whole genome. The relative fractions of HOL or BSW haplotypes vs. tRDC haplotypes in the mRDC group were calculated as rIBD scores. Shared haplotypes (*i.e.*, haplotypes with shared ancestry) were observed between mRDC on one hand and HOL, BSW, and tRDC on the other hand. These findings are in agreement with the results observed from population structure and admixture analysis that showed contributions from HOL, BSW, and tRDC to mRDC. In contrast to tRDC frequency, HOL haplotype and BSW haplotype frequency in the mRDC population for a given locus, (*i.e.*, rIBD score)

ranged from 0.73 to -0.74 and from 0.81 to -0.92 , where 1 indicates that all haplotypes were of HOL or BSW origin, while none were of tRDC origin. A value of -1 indicates that all haplotypes were tRDC-like (Figure 2 and Figure 3). The rIBD scores averaged across the genome were negative (-0.06 for HOL introgression and -0.09 for BSW introgression) (Figure 2a and Figure 3a), showing that the majority of the genome displayed more similarity with the tRDC than with either HOL or BSW. However, every chromosome contained genomic regions where the signal for the HOL or BSW haplotype was stronger than for tRDC.

The distribution of rIBD from the comparison between mRDC and HOL or BSW for IBD haplotypes resembled a normal distribution (Figure 2b and Figure 3b). By taking a cutoff of rIBD values, we were able to identify the regions that were likely to be of HOL or BSW origin. Across the whole genome, many known genes and QTL were located within regions with high rIBD for HOL origin (Table S1). We observed that the QTL are significantly enriched in the HOL-like haplotypes in mRDC ($P = 0.025$). The genes and QTL were associated with economic traits including milk-related traits such as milk yield, protein, fat yield, and percentage (*CD14*, *ZNF215*, *BCL2L12*, and *THRSP*) (Ashwell *et al.* 2004; Beecher *et al.* 2010; Magee *et al.* 2010; Cole *et al.* 2011; Cochran *et al.* 2013; Fontanesi *et al.* 2014; Capomaccio *et al.* 2015), calving traits (*MYH14*, *KCNC3*, *SYT3*, and *CTU1*) (Kolbehdari *et al.* 2008; Cole *et al.* 2011; Mao *et al.* 2016; Parker Gaddis *et al.* 2016; Abo-Ismael *et al.* 2017), feed efficiency-related traits (*LRR1Q3*, *ATP6V1B2*, and *CCKBR*) (Abo-Ismael *et al.* 2013, 2014), and carcass traits (*ZNF215*, *INTS4*, and *RPTOR*) (Magee *et al.* 2010; Sasago *et al.* 2017). The QTL-associated production and reproduction traits were significantly enriched in introgressed regions from HOL showing high rIBD scores ($P = 0.0004$ for production and $P = 0.034$ for reproduction traits). The longest continuous introgressed region (defined as the region with $rIBD > 0$) from HOL to mRDC was on chromosome 18 (56,320,000–61,350,000 bp) (Figure 2a). This region was previously found to be associated with calving traits and young stock survival in Nordic HOL cattle (Cole *et al.* 2011; Mao *et al.* 2016; Wu *et al.* 2017). The recombination rate of this region of chromosome 18 is low (Weng *et al.* 2014). The long genomic regions showing signals of introgression probably tend to occur in regions with low recombination rates. Moreover, introgressed haplotypes included numerous annotated genes due to genetic hitchhiking and a short time since introgression. The region with the highest rIBD score from HOL was located on chromosome 4 (120,540,000–120,810,000 bp, average $rIBD = 0.449$), which is downstream of the gene *VIPR2*. The *VIPR2* gene has been proposed to be a candidate gene affecting fat percentage and to play an important role in milk synthesis (Capomaccio *et al.* 2015).

Similarly, many known genes and QTL occurred in regions where mRDC shared haplotypes with BSW (Table S2). We observed an enrichment of QTL in the BSW-like haplotypes in mRDC that was close to significance ($P = 0.056$). These

genes and QTL mainly affected milk composition including fat and protein percentage and yield (*ITPR2*, *BCAT1*, *LAP3*, and *MED28*) (Cohen-Zinder *et al.* 2005; Pimentel *et al.* 2011; Zheng *et al.* 2011; Fang *et al.* 2014), growth and body conformation traits such as stature (*NCAPG*, *LCORL*, *PPP2R1A*, *IGFBP6*, and *CREBBP*) (Kolbehdari *et al.* 2008; Baeza *et al.* 2011; Lindholm-Perry *et al.* 2011; Cole *et al.* 2014; Sahana *et al.* 2015), calving and fertility traits (*EIF4G3*, *TGFA*, and *LAP3*) (Bongiorni *et al.* 2012; Hering *et al.* 2014; Höglund *et al.* 2015), and feed efficiency-related traits (*CLMP* and *PPP2R1A*) (Prakash *et al.* 2011; Cole *et al.* 2014). We observed a significant enrichment of QTL associated with milk, production, meat, and carcass traits in the introgressed regions from BSW in mRDC ($P = 0.045$ for milk, $P = 0.019$ for production, and $P = 0.017$ for meat and carcass traits). While the longest region introgressed from HOL in mRDC was ~ 4 Mb, there was no introgressed region longer than 1.5 Mb from BSW in mRDC. The highest value of rIBD signal was observed on chromosome 17 (35,630,000–35,780,000 bp, average $rIBD = 0.563$). This region is located downstream of the *IL2* gene. Also, there is one unannotated gene in this region. It has previously been shown that the *IL2* gene is associated with mastitis, milk yield, and lactation persistency (Alluwaimi *et al.* 2003; Prakash *et al.* 2011).

Regions of introgression and evidence for selection

We have shown that HOL or BSW haplotypes that were introgressed in mRDC often originated from genomic regions harboring genes associated with milk production, calving traits, feed efficiency, fertility or body conformation, and carcass traits. An introgression-introduced haplotype from HOL or BSW with continued survival in mRDC is not removed by negative selection or genetic drift, and might be a result of positive or balancing selection. The structure of an introgressed haplotype will change in such as LD, and the distribution of allele frequencies due to selection. The length of introgressed haplotype was affected by a combination of the local recombination rate and strength of selection. To further infer the introgressed regions that are under selection, we used three independent methods: iHS, F_{st} , and sharing of ROH among individuals to identify the regions that were under selection in the mRDC population. iHS identified regions with extended homozygosity in mRDC due to selection. Local high F_{st} statistics reflected the genomic regions that showed strong differentiation between mRDC and tRDC. The sharing of ROH among individuals could differentiate the genomic regions that have been fixed or were close to fixation in mRDC. We observed a significant dependency ($P < 0.01$) between the putative significant rIBD signals and significant signals from iHS, F_{st} , and sharing of ROH among individuals, except between the significant rIBD signals from HOL introgression and significant iHS signals. This supports the hypothesis that many introgressed regions from HOL or BSW are probably a result of selection.

We first observed significant positive correlations between F_{st} and rIBD scores larger than the median from both HOL

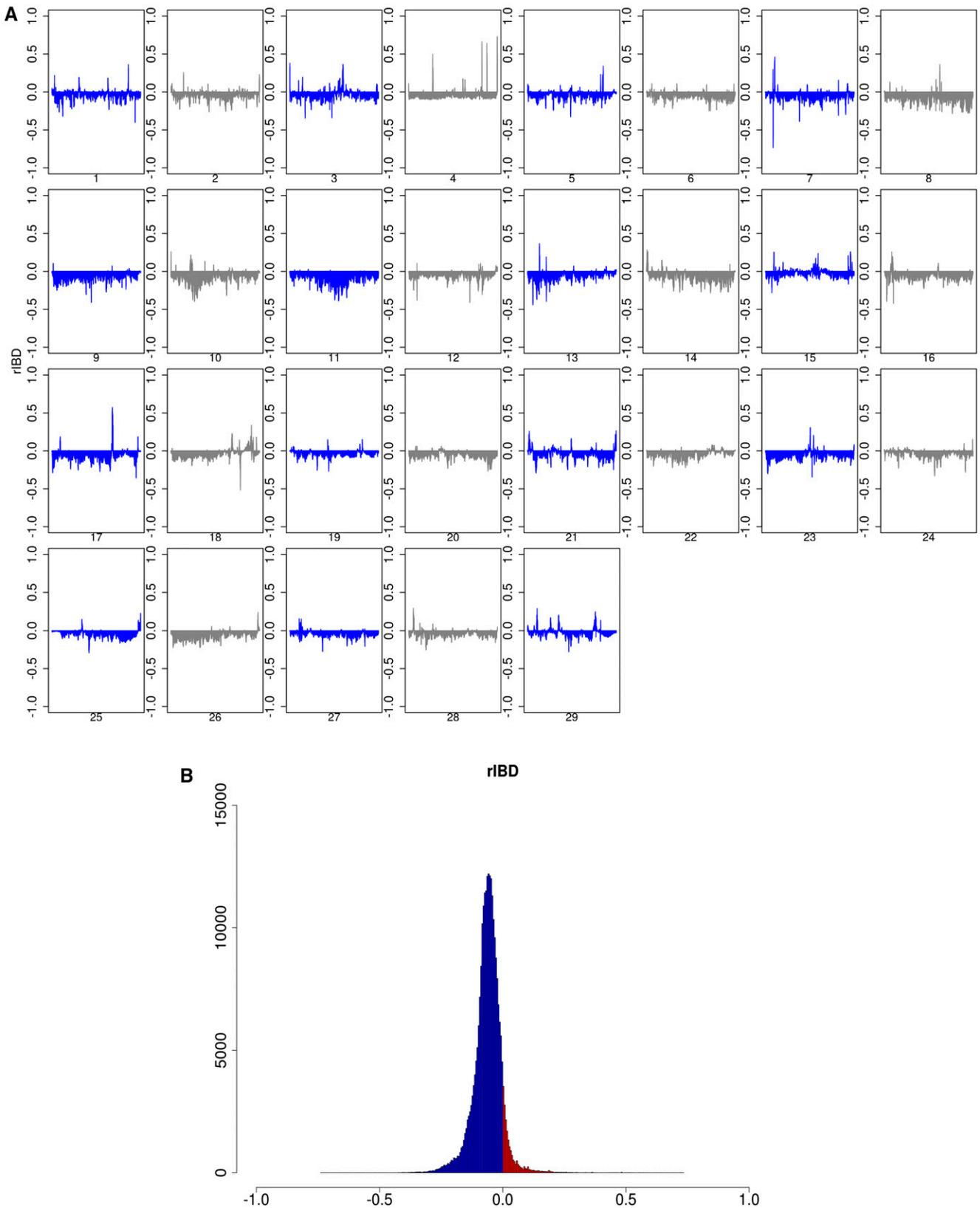


Figure 2 Genome-wide pattern of relative identity-by-descent (rIBD) score showing introgressed haplotypes from Holstein (HOL) in modern Danish Red dairy cattle (mRDC). (A) The rIBD score for all 29 autosomes: the positive scores show the signals where they are more HOL-like, whereas the negative scores show the signals where they are more traditional Danish red cattle (tRDC)-like. Chromosomes 1–29 are colored in blue and gray in order. (B) The distribution of rIBD scores: the positive scores show the signals where they are more HOL-like, whereas the negative scores show the signals where they are more tRDC-like.

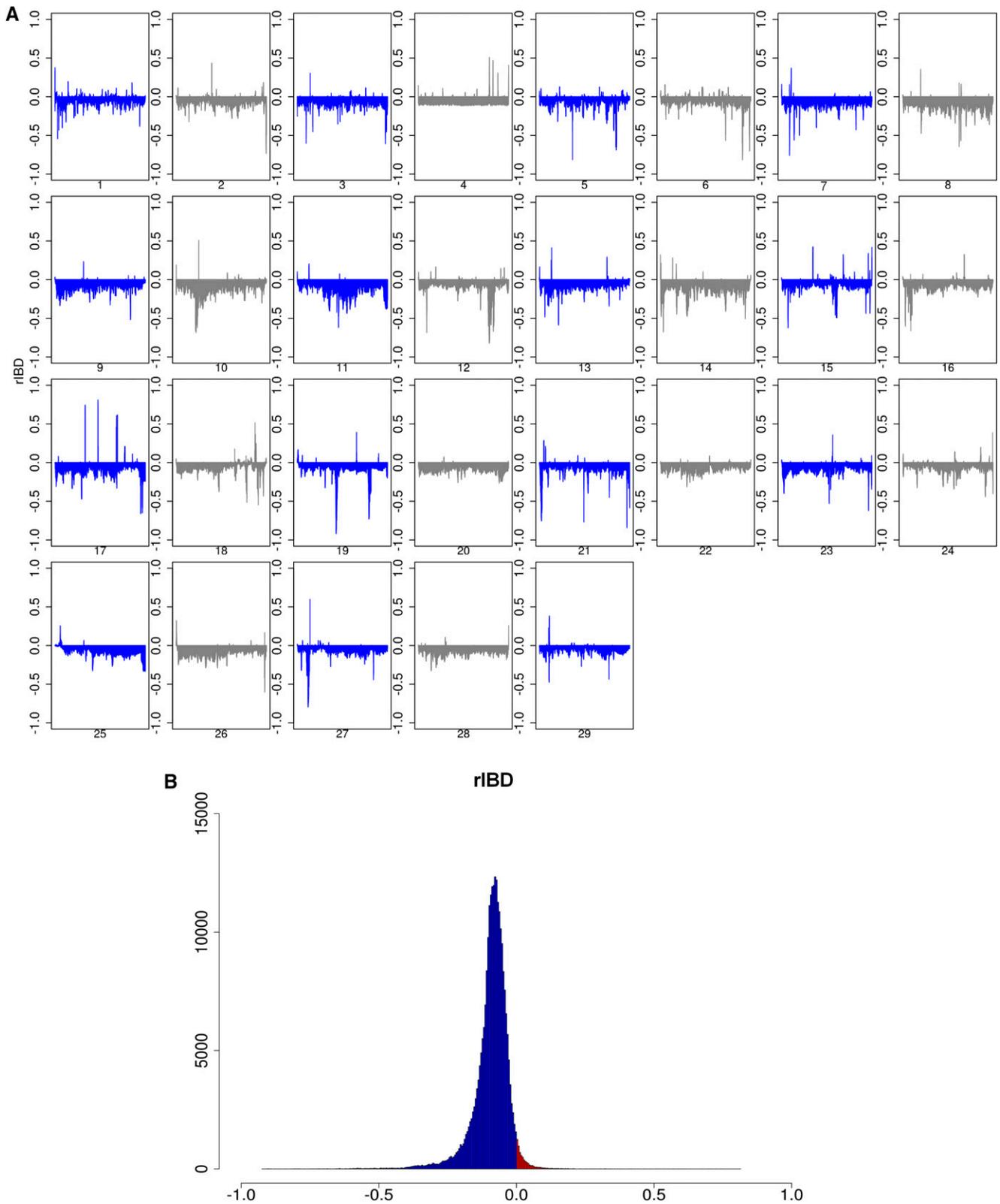


Figure 3 Genome-wide pattern of relative identity-by-descent (rIBD) score showing introgressed haplotypes from Brown Swiss (BSW) in modern Danish Red dairy cattle. (A) The rIBD score for all 29 autosomes: the positive scores show the signals where they are more BSW-like, whereas the negative scores show the signals where they are more traditional Danish red cattle (tRDC)-like. Chromosomes 1–29 are colored in blue and grey in order. (B) The distribution of rIBD scores: the positive scores show the signals where they are more HOL-like, whereas the negative scores show the signals where they are more tRDC-like.

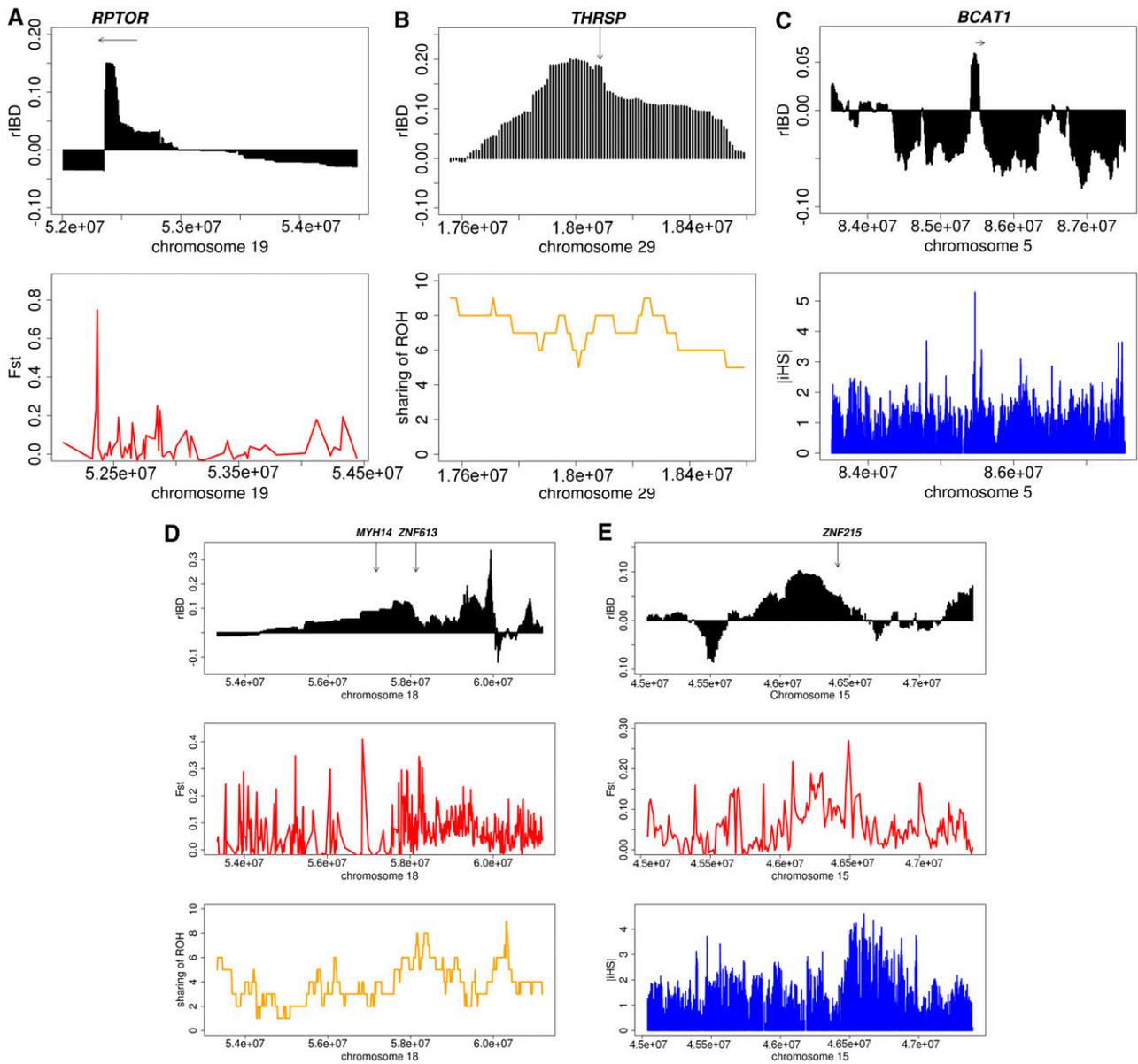


Figure 4 Examples of genomic regions with relative identity-by-descent (rIBD) scores, F_{st} , integrated haplotype score (iHS), and sharing of runs of homozygosity (ROH). (A) F_{st} and rIBD for the genomic region containing gene *RPTOR*; (B) rIBD and sharing of ROH for the genomic region containing gene *THRSP*; (C) rIBD and iHS for the genomic region containing gene *BCAT1*; (D) rIBD, F_{st} , and sharing of ROH for the genomic region containing genes *MYH14* and *ZNF613*; and (E) rIBD, F_{st} , and iHS for the genomic region containing gene *ZNF215*.

and BSW ($P < 0.001$), supporting the idea that at least some of the regions in mRDC showing differentiation from tRDC have been introgressed from HOL or BSW (Figures S2 and S3). Genes associated with milk yield, protein, and fat yield and percentage, such as *CD14* and *ZNF215* (Magee *et al.* 2010; Cochran *et al.* 2013), overlapped with haplotypes introgressed from HOL showing a high F_{st} (Figure 4e). The longest region introgressed from HOL in mRDC (chromosome 18: 56,320,000–61,350,000 bp) also showed high F_{st} values (Figure 4d). There were a number of genes included in this longest region with introgression from HOL and high F_{st} regions, such as *MYH14* and *ZNF613*, that were associated with calving ease and young stock survival (Abo-Ismael *et al.*

2017; Wu *et al.* 2017) (Figure 4d). The region on chromosome 19 (52,370,000–52,380,000 bp) with a very high F_{st} of 0.748 showed strong differentiation between tRDC and mRDC, which overlapped with a haplotype introgressed from HOL into mRDC. The *RPTOR* gene, associated with carcass traits in cattle (Sasago *et al.* 2017), is located here (Figure 4a). At the same time, the *RPTOR* gene was found to play an important role in regulating cell growth, energy homeostasis, apoptosis, and the immune response during adaptations (Sun *et al.* 2010). Similarly, we found that two BSW introgressed haplotypes overlapped with genes (*ITPR2* and *BCAT1*) associated with milk yield, fat, and protein yield and percentage (Pimentel *et al.* 2011; Fang *et al.* 2014) showed

high F_{st} (Figure 4c). Moreover, the region on chromosome 6 (38,730,000–38,780,000 bp) with an average F_{st} value of 0.251 overlapped with a BSW haplotype introgressed into mRDC. The *NCAPG* and *LCORL* genes, which are associated with body confirmation traits such as stature and feeding efficiency (Eberlein *et al.* 2009; Lindholm-Perry *et al.* 2011; Setoguchi *et al.* 2011; Xia *et al.* 2017), are located in this region.

Regions introgressed from HOL or BSW into mRDC and the significant regions from the iHS test were compared in Figures S4 and S5. There was a significant positive correlation between the rIBD score larger than the median for the introgressed region from BSW in mRDC and the iHS test ($P < 0.001$). However, there was no significant correlation between rIBD score for introgressed regions from HOL in mRDC and the iHS test. For example, the genomic region putatively under selection from the iHS test (38,510,000–38,540,000, 104 out of 108 SNPs with $|iHS| > 2$, highest $|iHS| = 3.8$) on chromosome 6 overlapped with an introgression signal from BSW into mRDC. This region is located upstream of the *LAP3* gene, which is associated with milk composition, including fat and protein percentage, and calving traits (Cohen-Zinder *et al.* 2005; Zheng *et al.* 2011). We observed that the significant SNPs with highest $|iHS|$ of 5.3 from the iHS test on chromosome 5 overlapped with a high introgression signal from BSW into mRDC. These significant SNPs overlap with the *BCAT1* gene, which is associated with milk yield and protein and fat percentage in milk (Pimentel *et al.* 2011) (Figure 4c). There were significant SNPs with a highest $|iHS|$ of 4.23 on chromosome 15 that overlapped with an introgressed region from BSW into mRDC. This region lies within the *ZNF215* gene affecting body confirmation and milk composition (Magee *et al.* 2010) (Figure 4e). Genomic regions showing signals both from the iHS test and introgression mapping, e.g., gene *BCAT1*, were introgressed from BSW, and probably under selection, but not yet fixed in the population due to low pressure of selection or recent introduction.

There was a significant positive correlation ($P < 0.001$) across genomic regions between the number of individuals containing overlapped ROH regions and rIBD scores larger than the median for both HOL and BSW (Figures S6 and S7). Short ROHs shared between individuals indicate selection events that have reached or are close to fixation. Interestingly, many small regions highly enriched for ROH hotspots overlapped with the longest region introgressed from HOL into mRDC, such as the region where the *ZNF613* gene associated with young stock survival (Wu *et al.* 2017) is located (Figure 4d). Similarly, we also observed that the region introgressed from HOL in mRDC, where the *THRSP* and *INTS4* genes are located, showed high enrichment of the ROH regions among individuals (Figure 4b). Studies have shown that the *THRSP* gene is associated with milk composition and involved in the regulation of mammary synthesis of milk fat (Fontanesi *et al.* 2014), while the *INTS4* gene is associated with a myristic acid content in carcass trait (Sasago *et al.* 2017). One region introgressed from BSW into mRDC containing the

TGFA gene, which is associated with sperm motility (Hering *et al.* 2014), also showed high levels of ROH sharing between individuals.

Conclusions

Together, the observed results demonstrate how crossbreeding followed by selection can shape the genomes of a modern breed on a genome-wide scale, using dairy cattle as an example. The well-documented breeding practice provides a robust model system for studying the consequences of adaptive introgression. Key observations were a highly uneven distribution across the genome of the proportions of genomic regions introgressed from the donor breeds. Highly introgressed regions contained genes and QTL known to affect traits of interest and to have been subjected to active selection by breeders. These traits include milk production, feed efficiency, calving traits, body confirmation, feed efficiency, carcass, and fertility traits. Artificial selection has played an important role on the genomic footprints from introgression on the genome of a modern dairy cattle breed. These findings contribute to our understanding of the genomic consequences of selective introgression in the genomes of modern species.

Acknowledgments

We are grateful to Viking Genetics (Randers, Denmark) for providing samples for sequencing. Q.Z. benefited from a joint grant from the European Commission within the framework of the Erasmus-Mundus joint doctorate European Graduate School in Animal Breeding and Genetics. This research was supported by the Center for Genomic Selection in Animals and Plants (GenSAP) funded by Innovation Fund Denmark (grant number 0603-00519B). M.C. acknowledges financial support from the Dutch Ministry of Economic Affairs, Agriculture, and Innovation (public-private partnership “Breed4Food” code BO-22.04-011-001-ASG-LR).

Author contributions: Q.Z. developed and planned the design of the study, coordinated the study, performed data analyses, and drafted the manuscript. B.G. participated in the design of the study, performed data analyses, and drafting of the manuscript. M.B., G.S., and M.C. participated in the design of the study and drafting of the manuscript. M.S.L. participated in drafting of the manuscript. All authors read and approved the final manuscript. The authors declare that they have no competing interests.

Literature Cited

- Abo-Ismael, M. K., M. J. Kelly, E. J. Squires, K. C. Swanson, S. Bauck *et al.*, 2013 Identification of single nucleotide polymorphisms in genes involved in digestive and metabolic processes associated with feed efficiency and performance traits in beef cattle. *J. Anim. Sci.* 91: 2512–2529. <https://doi.org/10.2527/jas.2012-5756>
- Abo-Ismael, M. K., G. Vander Voort, J. J. Squires, K. C. Swanson, I. B. Mandell *et al.*, 2014 Single nucleotide polymorphisms for

- feed efficiency and performance in crossbred beef cattle. *BMC Genet.* 15: 14. <https://doi.org/10.1186/1471-2156-15-14>
- Abo-Ismael, M. K., L. F. Brito, S. P. Miller, M. Sargolzaei, D. A. Grossi *et al.*, 2017 Genome-wide association studies and genomic prediction of breeding values for calving performance and body conformation traits in Holstein cattle. *Genet. Sel. Evol.* 49: 82. <https://doi.org/10.1186/s12711-017-0356-8>
- Ai, H. S., X. D. Fang, B. Yang, Z. Y. Huang, H. Chen *et al.*, 2015 Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* 47: 217–225. <https://doi.org/10.1038/ng.3199>
- Alexander, D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19: 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Alluwaimi, A. M., C. M. Leutenegger, T. B. Farver, P. V. Rossitto, W. L. Smith *et al.*, 2003 The cytokine markers in *Staphylococcus aureus* mastitis of bovine mammary gland. *J. Vet. Med. B Infect. Dis. Vet. Public Health* 50: 105–111. <https://doi.org/10.1046/j.1439-0450.2003.00628.x>
- Andersen, B., B. Jensen, A. Nielsen, L. G. Christensen, T. Liboriussen *et al.*, 2003 Malkerace-avlsmæssigt af kulturhistorisk belyst. Danmarks HordbrugsForskning, Denmark.
- Ashwell, M. S., D. W. Heyen, T. S. Sonstegard, C. P. Van Tassell, Y. Da *et al.*, 2004 Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J. Dairy Sci.* 87: 468–475. [https://doi.org/10.3168/jds.S0022-0302\(04\)73186-0](https://doi.org/10.3168/jds.S0022-0302(04)73186-0)
- Baeza, M. C., P. M. Corva, L. A. Soria, G. Rincon, J. F. Medrano *et al.*, 2011 Genetic markers of body composition and carcass quality in grazing Brangus steers. *Genet. Mol. Res.* 10: 3146–3156. <https://doi.org/10.4238/2011.December.19.3>
- Beecher, C., M. Daly, S. Childs, D. P. Berry, D. A. Magee *et al.*, 2010 Polymorphisms in bovine immune genes and their associations with somatic cell count and milk production in dairy cattle. *BMC Genet.* 11: 99. <https://doi.org/10.1186/1471-2156-11-99>
- Benjamini, Y., and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29: 1165–1188.
- Bongiorni, S., G. Mancini, G. Chillemi, L. Pariset, and A. Valentini, 2012 Identification of a short region on chromosome 6 affecting direct calving ease in Piedmontese cattle breed. *PLoS One* 7: e50137. <https://doi.org/10.1371/journal.pone.0050137>
- Bosse, M., H. J. Megens, L. A. F. Frantz, O. Madsen, G. Larson *et al.*, 2014a Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* 5: 4392. <https://doi.org/10.1038/ncomms5392>
- Bosse, M., H. J. Megens, O. Madsen, L. A. F. Frantz, Y. Paudel *et al.*, 2014b Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent *Sus scrofa* populations. *Mol. Ecol.* 23: 4089–4102. <https://doi.org/10.1111/mec.12807>
- Bosse, M., M. S. Lopes, O. Madsen, H. J. Megens, R. P. Crooijmans *et al.*, 2015 Artificial selection on introduced Asian haplotypes shaped the genetic architecture in European commercial pigs. *Proc. Biol. Sci.* 282: 20152019. <https://doi.org/10.1098/rspb.2015.2019>
- 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>
- Buzanskas, M. E., R. V. Ventura, T. C. S. Chud, P. A. Bernardes, D. J. D. Santos *et al.*, 2017 Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. *PLoS One* 12: e0171660. <https://doi.org/10.1371/journal.pone.0171660>
- Capomaccio, S., M. Milanese, L. Bomba, K. Cappelli, E. L. Nicolazzi *et al.*, 2015 Searching new signals for production traits through gene-based association analysis in three Italian cattle breeds. *Anim. Genet.* 46: 361–370. <https://doi.org/10.1111/age.12303>
- Chang, C. C., C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Cochran, S. D., J. B. Cole, D. J. Null, and P. J. Hansen, 2013 Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genet.* 14: 49. <https://doi.org/10.1186/1471-2156-14-49>
- Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Looor, A. Everts-van der Wind *et al.*, 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. <https://doi.org/10.1101/gr.3806705>
- Cole, J. B., G. R. Wiggans, L. Ma, T. S. Sonstegard, T. J. Lawlor, Jr. *et al.*, 2011 Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* 12: 408. <https://doi.org/10.1186/1471-2164-12-408>
- Cole, J. B., B. Waurich, M. Wensch-Dorendorf, D. M. Bickhart, and H. H. Swalve, 2014 A genome-wide association study of calf birth weight in Holstein cattle using single nucleotide polymorphisms and phenotypes predicted from auxiliary traits. *J. Dairy Sci.* 97: 3156–3172. <https://doi.org/10.3168/jds.2013-7409>
- Daetwyler, H. D., A. Capitan, H. Pausch, P. Stothard, R. Van Binsbergen *et al.*, 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>
- Davis, S. R., R. J. Spelman, and M. D. Littlejohn, 2017 BREEDING AND GENETICS SYMPOSIUM: Breeding heat tolerant dairy cattle: the case for introgression of the “slick” prolactin receptor variant into dairy breeds. *J. Anim. Sci.* 95: 1788–1800.
- Deschamps, M., G. Laval, M. Fagny, Y. Itan, L. Abel *et al.*, 2016 Genomic signatures of selective pressures and introgression from Archaic Hominins at human innate immunity genes. *Am. J. Hum. Genet.* 98: 5–21. <https://doi.org/10.1016/j.ajhg.2015.11.014>
- Eberlein, A., A. Takasuga, K. Setoguchi, R. Pfuhl, K. Flisikowski *et al.*, 2009 Dissection of genetic factors modulating fetal growth in cattle indicates a substantial role of the non-SMC condensin I complex, subunit G (NCAPG) gene. *Genetics* 183: 951–964. <https://doi.org/10.1534/genetics.109.106476>
- Fang, M., W. X. Fu, D. Jiang, Q. Zhang, D. X. Sun *et al.*, 2014 A multiple-SNP approach for genome-wide association study of milk production traits in Chinese Holstein cattle. *PLoS One* 9: e99544. <https://doi.org/10.1371/journal.pone.0099544>
- Figueiró, H. V., G. Li, F. J. Trindade, J. Assis, F. Pais *et al.*, 2017 Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci. Adv.* 3: e1700299. <https://doi.org/10.1126/sciadv.1700299>
- Fontanesi, L., D. G. Calo, G. Galimberti, R. Negrini, R. Marino *et al.*, 2014 A candidate gene association study for nine economically important traits in Italian Holstein cattle. *Anim. Genet.* 45: 576–580. <https://doi.org/10.1111/age.12164>
- Frischknecht, M., B. Bapst, F. R. Seefried, H. Signer-Hasler, D. Garrick *et al.*, 2017 Genome-wide association studies of fertility and calving traits in Brown Swiss cattle using imputed whole-genome sequences. *BMC Genomics* 18: 910. <https://doi.org/10.1186/s12864-017-4308-z>

- Galov, A., E. Fabbri, R. Caniglia, H. Arbanasic, S. Lapalombella *et al.*, 2015 First evidence of hybridization between golden jackal (*Canis aureus*) and domestic dog (*Canis familiaris*) as revealed by genetic markers. *R. Soc. Open Sci.* 2: 150450. <https://doi.org/10.1098/rsos.150450>
- Gautier, M., and R. Vitalis, 2012 rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28: 1176–1177. <https://doi.org/10.1093/bioinformatics/bts115>
- Grün, B., and F. Leisch, 2008 FlexMix Version 2: finite mixtures with concomitant variables and varying and constant parameters. *J. Stat. Softw.* 28: 1–35. <https://doi.org/10.18637/jss.v028.i04>
- Guan, Y. T., 2014 Detecting structure of haplotypes and local ancestry. *Genetics* 196: 625–642. <https://doi.org/10.1534/genetics.113.160697>
- Hartwig, S., R. Wellmann, R. Emmerling, H. Hamann, and J. Bennewitz, 2015 Short communication: importance of introgression for milk traits in the German Vorderwald and Hinterwald cattle. *J. Dairy Sci.* 98: 2033–2038. <https://doi.org/10.3168/jds.2014-8571>
- Hasenkamp, N., T. Solomon, and D. Tautz, 2015 Selective sweeps vs. introgression - population genetic dynamics of the murine leukemia virus receptor Xpr1 in wild populations of the house mouse (*Mus musculus*). *BMC Evol. Biol.* 15: 248. <https://doi.org/10.1186/s12862-015-0528-5>
- Hering, D. M., K. Olenski, and S. Kaminski, 2014 Genome-wide association study for poor sperm motility in Holstein-Friesian bulls. *Anim. Reprod. Sci.* 146: 89–97. <https://doi.org/10.1016/j.anireprosci.2014.01.012>
- Höglund, J. K., B. Buitenhuis, B. Guldbrandtsen, M. S. Lund, and G. Sahana, 2015 Genome-wide association study for female fertility in Nordic Red cattle. *BMC Genet.* 16: 110. <https://doi.org/10.1186/s12863-015-0269-x>
- Hope, A. C. A., 1968 A simplified Monte Carlo significance test procedure. *J. R. Stat. Soc. Series B Stat. Methodol.* 30: 582–598.
- Hu, Z. L., C. A. Park, and J. M. Reecy, 2016 Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res.* 44: D827–D833. <https://doi.org/10.1093/nar/gkv1233>
- Jagoda, E., D. J. Lawson, J. D. Wall, D. Lambert, C. Muller *et al.*, 2017 Disentangling immediate adaptive introgression from selection on standing introgressed variation in humans. *Mol. Biol. Evol.* 35: 623–630. <https://doi.org/10.1093/molbev/msx314>
- Kantanen, J., I. Olsaker, L. E. Holm, S. Lien, J. Vilkkil *et al.*, 2000 Genetic diversity and population structure of 20 North European cattle breeds. *J. Hered.* 91: 446–457. <https://doi.org/10.1093/jhered/91.6.446>
- Keele, J. W., L. A. Kuehn, T. G. McDaneld, R. G. Tait, S. A. Jones *et al.*, 2015 Genomewide association study of lung lesions in cattle using sample pooling. *J. Anim. Sci.* 93: 956–964. <https://doi.org/10.2527/jas.2014-8492>
- Kinsella, R. J., A. Kähäri, S. Haider, J. Zamora, G. Proctor *et al.*, 2011 Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database* 2011: bar030. <https://doi.org/10.1093/database/bar030>
- Kolbehdari, D., Z. Wang, J. R. Grant, B. Murdoch, A. Prasad *et al.*, 2008 A whole-genome scan to map quantitative trait loci for conformation and functional traits in Canadian Holstein Bulls. *J. Dairy Sci.* 91: 2844–2856. <https://doi.org/10.3168/jds.2007-0585>
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lindholm-Perry, A. K., A. K. Sexten, L. A. Kuehn, T. P. L. Smith, D. A. King *et al.*, 2011 Association, effects and validation of polymorphisms within the NCAPG-LCORL locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. *BMC Genet.* 12: 103. <https://doi.org/10.1186/1471-2156-12-103>
- Magee, D. A., K. M. Sikora, E. W. Berkowicz, D. P. Berry, D. J. Howard *et al.*, 2010 DNA sequence polymorphisms in a panel of eight candidate bovine imprinted genes and their association with performance traits in Irish Holstein-Friesian cattle. *BMC Genet.* 11: 93. <https://doi.org/10.1186/1471-2156-11-93>
- Mao, X., N. K. Kadri, J. R. Thomassen, D. J. De Koning, G. Sahana *et al.*, 2016 Fine mapping of a calving QTL on Bos taurus autosome 18 in Holstein cattle. *J. Anim. Breed. Genet.* 133: 207–218. <https://doi.org/10.1111/jbg.12187>
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20: 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Merks, J. W. M., P. K. Mathur, and E. F. Knol, 2012 New phenotypes for new breeding goals in pigs. *Animal* 6: 535–543. <https://doi.org/10.1017/S1751731111002266>
- Michenet, A., R. Saintilan, E. Venot, and F. Phocas, 2016 Insights into the genetic variation of maternal behavior and suckling performance of continental beef cows. *Genet. Sel. Evol.* 48: 45. <https://doi.org/10.1186/s12711-016-0223-z>
- Parker Gaddis, K. L., D. J. Null, and J. B. Cole, 2016 Explorations in genome-wide association studies and network analyses with dairy cattle fertility traits. *J. Dairy Sci.* 99: 6420–6435. <https://doi.org/10.3168/jds.2015-10444>
- Pickrell, J. K., and J. K. Pritchard, 2012 Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8: e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Pimentel, E. C. G., S. Bauersachs, M. Tietze, H. Simianer, J. Tetens *et al.*, 2011 Exploration of relationships between production and fertility traits in dairy cattle via association studies of SNPs within candidate genes derived by expression profiling. *Anim. Genet.* 42: 251–262. <https://doi.org/10.1111/j.1365-2052.2010.02148.x>
- Prakash, V., T. K. Bhattacharya, B. Jyotsana, and O. P. Pandey, 2011 Molecular cloning, characterization, polymorphism, and association study of the interleukin-2 gene in Indian Cross-bred cattle. *Biochem. Genet.* 49: 638–644. <https://doi.org/10.1007/s10528-011-9438-4>
- Saatchi, M., R. D. Schnabel, J. F. Taylor, and D. J. Garrick, 2014 Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* 15: 442. <https://doi.org/10.1186/1471-2164-15-442>
- Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419: 832–837. <https://doi.org/10.1038/nature01140>
- Sahana, G., J. K. Höglund, B. Guldbrandtsen, and M. S. Lund, 2015 Loci associated with adult stature also affect calf birth survival in cattle. *BMC Genet.* 16: 47. <https://doi.org/10.1186/s12863-015-0202-3>
- Sasago, N., T. Abe, H. Sakuma, T. Kojima, and Y. Uemoto, 2017 Genome-wide association study for carcass traits, fatty acid composition, chemical composition, sugar, and the effects of related candidate genes in Japanese Black cattle. *Anim. Sci. J.* 88: 33–44. <https://doi.org/10.1111/asj.12595>
- Setoguchi, K., T. Watanabe, R. Weikard, E. Albrecht, C. Kuhn *et al.*, 2011 The SNP c.1326T > G in the non-SMC condensin I complex, subunit G (NCAPG) gene encoding a p.Ile442Met variant is associated with an increase in body frame size at puberty in cattle. *Anim. Genet.* 42: 650–655. <https://doi.org/10.1111/j.1365-2052.2011.02196.x>

- Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan *et al.*, 2001 dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29: 308–311. <https://doi.org/10.1093/nar/29.1.308>
- Sun, C., C. Southard, D. B. Witonsky, R. Kittler, and A. Di Rienzo, 2010 Allele-specific down-regulation of RPTOR expression induced by retinoids contributes to climate adaptations. *PLoS Genet.* 6: e1001178. <https://doi.org/10.1371/journal.pgen.1001178>
- Voight, B. F., S. Kudravalli, X. Q. Wen, and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. *PLoS Biol.* 4: 446–458.
- vonHoldt, B. M., R. Kays, J. P. Pollinger, and R. K. Wayne, 2016 Admixture mapping identifies introgressed genomic regions in North American canids. *Mol. Ecol.* 25: 2443–2453. <https://doi.org/10.1111/mec.13667>
- Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population-structure. *Evolution* 38: 1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
- Weng, Z. Q., M. Saatchi, R. D. Schnabel, J. F. Taylor, and D. J. Garrick, 2014 Recombination locations and rates in beef cattle assessed from parent-offspring pairs. *Genet. Sel. Evol.* 46: 34. <https://doi.org/10.1186/1297-9686-46-34>
- White, S., 2011 From globalized pig breeds to capitalist pigs: a study in animal cultures and evolutionary history. *Environ. Hist.* 16: 94–120. <https://doi.org/10.1093/envhis/emq143>
- Williams, R. L., 2000 A note on robust variance estimation for cluster-correlated data. *Biometrics* 56: 645–646. <https://doi.org/10.1111/j.0006-341X.2000.00645.x>
- Wu, X. P., B. Guldbbrandtsen, U. S. Nielsen, M. S. Lund, and G. Sahana, 2017 Association analysis for young stock survival index with imputed whole-genome sequence variants in Nordic Holstein cattle. *J. Dairy Sci.* 100: 6356–6370. <https://doi.org/10.3168/jds.2017-12688>
- Xia, J. W., H. Z. Fan, T. P. Chang, L. Y. Xu, W. G. Zhang *et al.*, 2017 Searching for new loci and candidate genes for economically important traits through genebased association analysis of Simmental cattle. *Sci. Rep.* 7: 42048. <https://doi.org/10.1038/srep42048>
- Yang, J. A., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011 GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88: 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Zhang, Q. Q., M. P. L. Calus, B. Guldbbrandtsen, M. S. Lund, and G. Sahana, 2015a Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16: 88. <https://doi.org/10.1186/s12863-015-0227-7>
- Zhang, Q. Q., B. Guldbbrandtsen, M. Bosse, M. S. Lund, and G. Sahana, 2015b Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16: 542. <https://doi.org/10.1186/s12864-015-1715-x>
- Zheng, X., Z. H. Ju, J. Wang, Q. L. Li, J. M. Huang *et al.*, 2011 Single nucleotide polymorphisms, haplotypes and combined genotypes of LAP3 gene in bovine and their association with milk production traits. *Mol. Biol. Rep.* 38: 4053–4061. <https://doi.org/10.1007/s11033-010-0524-1>
- 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. <https://doi.org/10.1186/gb-2009-10-4-r42>

Communicating editor: D. J. de Koning