Genomic analyses suggest adaptive differentiation of Northern European native cattle breeds

Stronen, Astrid V.; Pertoldi, Cino; Iacolina, Laura; Kadarmideen, Haja; Kristensen, Torsten N.

Published in:
Evolutionary Applications

Link to article, DOI:
10.1111/eva.12783

Publication date:
2019

Document Version
Peer reviewed version

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Genomic analyses suggest adaptive differentiation of Northern European native cattle breeds

Astrid V. Stronen\textsuperscript{1,2,3}, Cino Pertoldi\textsuperscript{1,4}, Laura Iacolina\textsuperscript{1,4}, Haja N. Kadarmideen\textsuperscript{5}, Torsten N. Kristensen\textsuperscript{1}

\textsuperscript{1}Section of Biology and Environmental Science, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, DK-9220 Aalborg, Denmark, cp@bio.aau.dk, lauraiacolina@gmail.com, tnk@bio.aau.dk

\textsuperscript{2}Department of Biology, Biotechnical Faculty, University of Ljubljana, Večna pot 111, SI-1000 Ljubljana, Slovenia, astrid.stronen@gmail.com

\textsuperscript{3}Department of Biotechnology and Life Sciences, Insubria University, via J.H. Dunant 3, 21100 Varese, Italy

\textsuperscript{4}Aalborg Zoo, Mølleparkvej 63, DK-9000 Aalborg, Denmark

\textsuperscript{5}Quantitative Genomics, Bioinformatics and Computational Biology Group, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kongens Lyngby, Denmark, hajak@dtu.dk

\textbf{Running head}: Genomic analyses of North European native cattle

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/eva.12783

This article is protected by copyright. All rights reserved.
Abstract

Native domestic breeds represent important cultural heritage and genetic diversity relevant for production traits, environmental adaptation and food security. However, risks associated with low effective population size, such as inbreeding and genetic drift, have elevated concerns over whether unique within-breed lineages should be kept separate or managed as one population. As a conservation genomic case study of the genetic diversity represented by native breeds, we examined native and commercial cattle (Bos taurus) breeds including the threatened Danish Jutland cattle. We examined population structure and genetic diversity within breeds and lineages genotyped across 770K single nucleotide polymorphism loci to determine (a) the amount and distribution of genetic diversity in native breeds, and (b) the role of genetic drift vs. selection. We further investigated the presence of outlier loci to detect (c) signatures of environmental selection in native vs. commercial breeds, and (d) native breed adaptation to various landscapes. Moreover, we included older cryopreserved samples to determine (e) if cryopreservation allows (re)introduction of original genetic diversity. We investigated a final set of 195 individuals and 677K autosomal loci for genetic diversity.
within and among breeds, examined population structure with principal component analyses and a maximum likelihood approach, and searched for outlier loci suggesting artificial or natural selection. Our findings demonstrate the potential of genomics for identifying the uniqueness of native domestic breeds, and for maintaining their genetic diversity and long-term evolutionary potential through conservation plans balancing inbreeding with carefully designed outcrossing. One promising opportunity is the use of cryopreserved samples, which can provide important genetic diversity for populations with few individuals, while helping to preserve their traditional genetic characteristics. Outlier tests for native vs. commercial breeds identified genes associated with climate adaptation, immunity and metabolism, and native breeds may carry genetic variation important for animal health and robustness in a changing climate.

**Key words:** Animal health, artificial selection, *Bos taurus*, climate adaptation, conservation genomics, environmental selection, production traits, single nucleotide polymorphism

**Introduction**

Domestication of plants and animal species has permitted significant human population growth and the development of modern human societies (Larson & Burger, 2013; Larson & Fuller, 2014; Marshall et al., 2014). Domestication is considered to have occurred along three major pathways, termed commensal, prey or directed (reviewed in Larson & Burger, 2013; Larson & Fuller, 2014). They describe the commensal pathway as centered on animal habituation to a human niche, whereas the prey pathway involved animals that humans initially preyed upon and later started to manage. In contrast, Larson & Burger (2013) and Larson & Fuller (2014) note that the more recent directed pathway has been the only deliberate route to domestication, which bypassed the habituation and management phases.
There is evidence that sheep (Ovis sp.), goats (Capra sp.) and cattle (Bos sp.) were domesticated 10,500-10,000 years before present via the prey pathway, through various stages of intensive breeding of captive animals and the subsequent development of distinct breeds (Larson & Burger, 2013; Larson & Fuller, 2014). Conservation of native domestic animals and plants is now receiving growing attention. Populations of conservation concern may encompass important cultural heritage and potentially genetic diversity relevant for modern breeding and future food security such as for production traits, adaptation to harsh environments and climate change (Hoffmann, 2013; Kantanen et al., 2015; Kristensen et al., 2015; Iacolina et al., 2016).

*Genetic diversity within and among native breeds*

Native domestic breeds are known to share several conservation concerns with populations of wild species at risk, such as low effective population size (N_E), which in turn reduces the effectiveness of selection and increases the impacts of genetic drift and inbreeding (Kantanen et al., 2000; Taberlet et al., 2008; Leroy et al., 2013; Pertoldi et al., 2014). Numerous livestock breeds have gone extinct or are threatened (FAO, 2007; 2015). Genetic variation within and between breeds is rapidly lost and, for many breeds, we have little information about levels of genetic variation, N_E, and adaptation to past and present local environmental conditions. Past selection for conditions such as the ability to survive on food with limited nutritional content could be important for adaptation to climate change, and for the use of native breeds in habitat management. To maintain sustainable populations in the short term an N_E of at least 50 is sometimes recommended in conservation genetics to keep inbreeding rates at acceptable levels and an N_E > 500 is needed to allow maintenance of evolutionary potential over time (i.e., across hundreds of generations) (Frankham et al., 2014; Hoffmann et al., 2017). Because domestic animal breeds typically have N_E < 100 (Leroy et al., 2013)
current breeding practices raise long-term concerns for the evolutionary potential of many of these populations.

Genomic methods and reproductive techniques are rapidly developing and can help provide answers to many important questions in conservation genetics. These include levels of inbreeding and genetic drift, genetic uniqueness, identification of genomic regions under selection, and creation of genetic rescue programs or breeding schemes to maintain adaptive genetic variation in domestic animals more efficiently than pedigree-based breeding (e.g., Porto-Neto et al., 2014; Kantanen et al., 2015; Williams et al., 2015; Kukučková et al., 2017). Improved communication between the fields of research and management concerning experimental results on genetic rescue and other conservation actions is therefore important for an efficient management of populations with small and declining \( N_E \) (Kristensen et al., 2015; Hoffmann et al., 2016).

An example of a native cattle (\textit{Bos taurus}) breed under threat is the Danish Jutland cattle, which includes four contemporary lineages (within-breed subpopulations maintained in (relative) isolation from other such groups). These are the Westergaard, Vesterbølle, Oregaard and Kortegaard lineages (Figure 1). An earlier microsatellite genetic study of cattle breeds reported the Jutland cattle to be genetically unique (Brüniche-Olsen et al., 2012). Over the past centuries the breed has experienced a severe decline in population size (Brüniche-Olsen et al., 2012) and a 2004 estimate indicated an overall \( N_E \) for the Jutland cattle breed of around 40 (http://www.fao.org/dad-is/browse-by-country-and-species/en/), suggesting strong drift and high rates of inbreeding. The low \( N_E \) for the Jutland cattle breed was supported by Pertoldi et al. (2014) who analysed genome-wide profiles with 50K single nucleotide polymorphism (SNP) markers in the Jutland Kortegaard-lineage. This native livestock breed thus offer an informative case study of how genome-wide profiles can inform conservation genetic management of small populations at risk.
Genetic drift and selection in native breeds

Breeders and managers working with small populations at risk are often concerned with the genetic uniqueness of these populations (e.g., Kantanen et al., 2000; Ginja et al., 2010; Withen et al., 2011). This issue is relevant for wild and domestic species, including native livestock breeds where survival has been influenced by local environmental conditions. Genomic profiles from such breeds can advance evolutionary research and conservation by improving our understanding of i) the role of genetic drift vs. selection in contemporary within-breed lineages, ii) signatures of selection in native vs. commercial breeds, and iii) adaptation to different landscapes in native breeds.

For the Jutland cattle, a long-standing discussion among farmers and managers in Denmark has been whether lineages should be kept separate or managed as one population. Analyses of genomic profiles from all four contemporary Jutland cattle lineages will thus allow us to determine the amount and distribution of diversity within the breed and among individual lineages. This issue has relevance across livestock breeds where managers acknowledge that rapid conservation actions may be needed to preserve small and declining populations, yet they have concerns that admixture of within-breed lineages may risk further loss of unique genetic variation, with potential negative implications for locally-adapted traits (Taberlet et al., 2008; Hoffmann, 2013). Vital management considerations include the costs and benefits of maintaining separate lineages with few remaining individuals, and preservation of characteristics potentially limiting productivity (e.g., milk, wool, meat) but augmenting survival in harsh environments such as areas with extreme temperatures or precipitation levels, or in habitats with poor-quality food sources (Pariset et al., 2009; Hoffmann, 2013).
Preservation of semi-natural and cultural landscapes has been recognized as a key priority in Europe (Halada et al., 2011; Timmermann et al., 2014), and native breeds appear well-suited to extensive agriculture and biodiversity maintenance in the form of grazing, which can simultaneously offer suitable conditions for in situ natural selection by means of exposure to changing environmental conditions (Hoffmann, 2013). Northern European countries bordering the Atlantic Ocean have cool and humid climates, which may exert selective pressures on domestic species (Pariset et al., 2009). Yet within this region there are substantial differences in landscape form and terrain ruggedness. In this study we compare the Jutland cattle from Denmark, dominated by a relatively flat terrain, to native breeds from rugged landscapes in Norway and the Faroe Islands, enabling an investigation of adaption to local conditions (Raziq et al., 2011; Bailey et al., 2015).

The potential role of cryopreservation in preserving native breeds

The carrying capacity of a population can be increased artificially (i.e., without expanding the in situ population) by supplying genetic material from less related earlier generations of animals. This strategy can be applied to native breeds of cattle and other species by cryopreservation of spermatozoa and oocytes in a gene bank (Curry, 2000; Su et al., 2012). An increasing number of species will, in the near future, be managed with the help of cryopreservation techniques (Charlton et al., 2018), including initiatives such as The Frozen Ark Project (https://www.frozenark.org/). In Denmark, cryopreservation is now used for native livestock breeds including the Jutland cattle, and these collections can provide genetic material for ongoing conservation efforts. Specifically, Hertz et al. (2016) simulated a supplementation in the Jutland Kortegaard-lineage where one male was added to the population every five years, representing cryopreserved semen from previous generations.
The simulation suggested that such supplementation can postpone the time to extinction and reduce inbreeding levels (Hertz et al., 2016).

Emerging genomic methods provide exciting prospects for identification of genetic variation and selection in native breeds, yet additional efforts are needed for native breed conservation management to benefit fully from these new approaches (Bruford et al., 2015). The objective of our study was to use within-breed lineages of the Danish Jutland cattle; native breeds from neighbouring countries that may have been subject to different selective pressures (artificial and natural selection), and commercial dairy breeds selected for high productivity, to investigate: (a) the amount and distribution of genetic diversity in a native livestock breed with several lineages, which over recent decades have declined in numbers and are believed to have experienced bottlenecks and extensive genetic drift; (b) the role of genetic drift vs. selection in contemporary lineages within native breeds; (c) signatures of selection in native vs. commercial breeds; (d) native breed adaptation to various landscapes and environmental conditions (e.g., terrain ruggedness); and (e) whether cryopreserved semen samples conserved over the past decades offer possibilities for infusing native breeds with new variability from past generations.

Materials and methods

Genetic diversity within and among native breeds

We compared genomes from the Danish Jutland cattle to native breeds from other northern European countries and to commercial breeds originating from northern Europe (Table 1). The Jutland cattle is an indigenous breed that descends from original black- and grey pied cattle from the 16th-18th century, and although the Jutland cattle was once widespread in Denmark it declined following competition and near-replacement with larger more
productive breeds (Kantanen et al., 2000; Brùniche-Olsen et al., 2012; Sørensen & Nielsen, 2015). In 1949 the breeding association for the Jutland cattle decided to accept Dutch and German black pied cattle (Friesian) into the stud book, and the resulting highly successful crosses formed the Danish black pied cattle that within a decade had replaced almost all original Jutland cattle (Brùniche-Olsen et al., 2012, Sørensen & Nielsen, 2015). Subsequently, genetic material from Holstein cattle in North America, where this European breed had been imported and further developed, was introduced into the global population of black pied cattle, and this breed constitute the modern Holstein (also named Holstein-Frisian) breed (Sørensen & Nielsen, 2015). This process also occurred in Denmark, however; a Friesian population with limited contribution of Holstein genes has been preserved and this lineage is named SDM-1965 (DAD-IS, 2017a).

To compare genetic diversity, population structure and selection in Jutland cattle with other native breeds, we sampled the SDM-1965, and native breeds from rugged landscapes in Norway and the Faroe Islands (Table 1) that could exhibit signs of selection for different traits. Additionally, for comparison of native breeds from extensive agriculture with commercial dairy breeds, we sampled Holstein and Jersey cattle that both originate from regions of Northern Europe with a cool and humid climate, and have a history of intense selection for high productivity. Within the Jutland breed we examined the four existing lineages (Figure 1). Moreover, we investigated cryopreserved material of Danish black pied cattle and SDM-1965, referred to as old bulls. These cryopreserved samples were divided into pre-1980 and post-1980 groups to allow temporal comparison of genetic diversity (Table 1), where the samples collected 1960-1980 may be expected to exhibit higher genetic diversity than that of modern samples. The 19 pre-1980 cryopreserved samples are listed in the stud book as SDM-1965. For the post-1980 samples, four are SDM-1965 and 14 are classified as Jutland cattle (Supplemental Table S1). For contemporary individuals we
collected ear tissue samples from live animals using standard methods with the nextGen tissue sampling unit by AllFlex (http://www.allflexusa.com/, following the manufacturer’s recommendations) with the help of cattle owners and managers, and obtained semen samples from Viking Genetics (http://www.vikinggenetics.com/). Samples were genotyped with the Illumina BovineHD array with 776,665 SNPs at Genoskan A/S (Aarhus, Denmark) following the manufacturer’s protocol and based on the UMD 3.1 bovine genome assembly. Duplicate samples of two individuals, one old bull post-1980 and one from the Faroe Islands, were included and allowed comparison of genotyping consistency. The data was filtered in PLINK 1.9 (Purcell et al., 2007) for individual genotyping success of at least 90%, SNP genotyping success of minimum 98%, and minor allele frequency (MAF) of 1%. To obtain equalised sample sizes, we then subsampled the more numerous Kortegaard and Oregaard-lineages by including the first 20 individuals from a list of sample IDs, without any additional knowledge of the individuals, for further analyses. Accordingly, the analyses were performed with reduced sample size for the Kortegaard and Oregaard-lineages, except for one comparative principal component analysis (PCA) to evaluate the data for these two groups without downsampling (Supplemental Figure S1).

We pruned the data for loci in linkage disequilibrium (LD), and because the Jutland cattle lines have small populations and low genetic diversity where LD is expected to be high, we filtered the data to remove highly linked SNPs. For filtering we used the PLINK formula (--indep 50 5 2), where 50 is the size of the sliding window (i.e., 50 bp at a time are examined for linked loci), 5 is the number of SNPs shifted in each step, and 2 is the variance inflation factor (see further details at http://zzz.bwh.harvard.edu/plink/summary.shtml). We then examined genetic structure in the data with PCA in the adegenet package (Jombart, 2008; Jombart & Ahmed, 2011) in R 2.14.2 (R Development Core Team, 2012) and with ADMIXTURE (Alexander et al., 2009). In ADMIXTURE, the number of population clusters...
(K) is determined with a cross-validation procedure, where the optimal K-value has the lowest cross-validation error relative to alternate K-values (Alexander et al., 2015). We examined a range of K-values from 1-15, and used 20 cross-validations for each K-value and 1000 bootstrap replicates. Subsequently, to provide another measure of differentiation among breeds and lineages, we calculated pairwise $F_{ST}$ in Genepop (4.6) (Raymond & Rousset, 1995; Rousset, 2008). For each pairwise comparison, we evaluated the statistical significance of population differentiation by permutations in GenoDive v.2.0b23 (Meirmans & van Tienderen, 2004) with 50,000 randomly selected SNP loci. We applied the Bonferroni correction for multiple comparisons (Rice, 1989). We used TreeMix 1.12 (Pickrell & Pritchard, 2012; http://gensoft.pasteur.fr/docs/treemix/1.12/treemix_manual_10_1_2012.pdf sections 4.1 and 4.4) to build a maximum likelihood tree and examine signs of possible admixture among populations. We analysed 195 individuals and 595 025 SNPs pruned for MAF and genotyping success in PLINK, as outlined above, and plotted a maximum likelihood tree with the R-script plottingfuncs.R provided in the source code for TreeMix (https://bitbucket.org/nygcresearch/treemix/wiki/Home). To measure the amount of genetic diversity within breeds, and within Jutland cattle lineages, we calculated polymorphism ($P\%$), observed ($H_{O}$) and expected heterozygosity ($H_{E}$) with standard error, and pairwise identity by descent (IBD) between individuals within breeds/lineages in PLINK. Furthermore, we calculated $N_{E}$ with 95% confidence intervals in NeEstimator v2.1 (Do et al., 2014) with the linkage disequilibrium method (Waples & Do, 2008) and the data set pruned for LD. The same data set was used to calculate the number of private alleles ($P_{A}$) per population in R with a script (https://johnbborne.wordpress.com/2017/07/12/identifying-private-snps-in-r/) for the HierfStat package (Goudet, 2005).
To evaluate whether genomic profiles of different breeds showed signs of selection we performed tests of outlier loci in the pcadapt package (Luu et al., 2017) in R. The program incorporates a false discovery rate (FDR) approach to account for multiple testing that permits users to select a specific alpha-level (Luu et al., 2017, https://bcm-uga.github.io/pcadapt/articles/pcadapt.html). We used equalised population sizes and an alpha-level (q-value) of 0.01 for detecting loci that deviated significantly from the neutral distribution, whereby loci with q-values < 0.01 were considered outliers. We performed seven separate tests for outlier loci (Table 2). These included six tests for signs of possible selection linked to local environmental conditions and one comparison of two Jutland cattle lineages that show signs of genetic structure believed to have been produced by genetic drift.

Although the latter test does not represent any true control, we intended it to provide context toward understanding how the outlier test performs in the event of (presumed) genetic drift only. Firstly, (T1) examined all population clusters discovered (K = 9, see Results). The subsequent T2-T7 were pairwise tests, each time with two relevant populations, to help identify unique variants or results shared across two or more populations (Table 2). The (T2) tested the Kortegaard vs. the Oregaard-lineages of Jutland cattle, two relatively recent lineages where observed genetic divergence (Brüniche-Olsen et al., 2012) is expected to be based on drift with no known differences in selection regimes. Subsequently, (T3) compared 1960-1980 cryopreserved Danish SDM-1965 with Holstein cattle (test of native vs. commercial breeds), (T4) tested 1960-1980 cryopreserved Danish SDM-1965 vs. Western Norwegian Red-polled cattle (comparing two native breeds), and (T5) compared the Western Norwegian Red-polled cattle with Western Norwegian Fjord cattle (comparing two native breeds, where the Western Norwegian Red-polled cattle are larger and polled, i.e., without horns). Finally, (T6) compared Jutland cattle, represented by old Jutland bulls post-1980,
with Holstein cattle (test of native vs. commercial breeds), and (T7) tested the old Jutland 
bulls post-1980 vs. Western Norwegian Red-polled cattle (comparing two native breeds). For 
brevity we subsequently refer to these tests as T1-T7 (Table 2).

For SNPs identified as outliers under divergent selection we screened the 
flanking region of the bovine genome in the NCBI Map Viewer, annotation release 104 
the approach in Porto-Neto et al. (2014) in their study of climate adaptation in cattle and 
examined 3000 base-pairs (bp) on each side of outlier SNPs to identify functional genes 
under known or potential selection. We noted genes found within the 3000 bp flanking 
regions of the outliers, and examined the genes with the NCBI Gene web site 
(https://www.ncbi.nlm.nih.gov/gene/) for information on gene function in cattle and other 
species (primarily records from Homo sapiens and Mus musculus). We centered on genes 
relevant to selection in cattle and our investigation of native breeds (see Supplemental Table 
S2 for a list of all genes), and chose genes for in-depth investigation (henceforth focal genes) 
using the bibliography from NCBI Gene and additional references obtained from literature 
searches.

For focal genes, we examined possible enrichment in gProfiler (Reimand et al., 
2016) with the Gene Ontology (GO) Biological Processes and Human Phenotype Ontology 
databases (https://biit.cs.ut.ee/gprofiler/index.cgi). We followed the approach of Caniglia et 
al. (2018) and limited the size of functional categories to maximum 500 terms, to focus our 
analyses on more specific genome regions. We chose the gProfiler g:SCS significance 
threshold for multiple testing as recommended by the program authors (Reimand et al., 
2016).
We examined and compared runs of homozygosity (ROH) in PLINK for representative groups including two native breeds from different environments and one commercial breed. These were the native Jutland cattle (old bulls post-1980), the native Western Norwegian Red-pollled cattle, and the commercial Holstein breed. ROH represent long homozygous segments in individuals where both parents have transmitted identical haplotypes, and can indicate various processes that include selection and recent inbreeding (Purfield et al., 2012). We used data for autosomal loci filtered for MAF and genotyping success, and PLINK functions --homozyg and --homozyg-group with default parameters. These parameters permit one heterozygote and five missing genotypes within windows of 5 Mb and 50 SNPs. To compare results from different methods to detect possible selection, we plotted the distribution of focal genes detected in outlier flanking regions and ROH across all autosomal chromosomes based on the approach in Stronen et al. (2017), comparing Jutland vs. Holstein cattle, and Jutland vs. Western Norwegian Red-pollled cattle.

Results

*Genetic diversity within and among native breeds*

Quality screening of the data produced 710,471 SNPs. After equalising sample sizes (Table 1) we obtained a final sample of 195 individuals and 677,311 autosomal SNPs filtered for genotyping success (as defined above per individual and per SNP), and 88,190 SNPs after linkage pruning. Principal Component Analysis plots (Figure 2, Supplemental Figure S2) showed Norwegian and Faroe Island cattle, commercial breeds, old bulls, SDM-1965 and contemporary Jutland cattle lineages distributed along PC1. The contemporary four Jutland lineages were divergent from old bulls pre-1980 and most SDM-1965 individuals on PC1, with the more recent samples from old bulls post-1980 broadly scattered. PC2 showed separation among Jutland cattle lineages, where the Kortegaard and Westergaard lineages...
emerged as most divergent. The third PC separated old bulls pre-1980, and SDM-1965; PC4 highlighted the Jutland Westergaard-lineage, and PC5 highlighted differentiation between Western Norwegian Red-polled and Jersey cattle (Supplemental Figure S2).

The ADMIXTURE results showed the highest support for $K = 9$, although it should be noted that the rate of reduction in the CV error showed a marked decline at $K = 5$ (Figure 3). The ADMIXTURE results for $K = 2$ exhibited differentiation between Danish native cattle and other breeds. New units that emerged at $K$-values of 3-5 were the Westergaard-lineage, the Kortegaard-lineage, the old bulls pre-1980 (i.e., SDM-1965) and modern SDM-1965 (Figure 4). The $K = 4$ value separated Danish native cattle, i.e. the four Jutland cattle lineages and SDM-1965 with Jutland ancestry, from all other commercial and foreign breeds (for a detailed discussion of Jutland lineages see Supplemental Note S1). At $K = 5$ the Westergaard-lineage constituted a separate group within the Jutland cattle whereas the Kortegaard-lineage and a portion of the old bulls post-1980 were assigned to one cluster, to which individuals from the Oregaard-lineage and the Vesterbølle-lineage also showed partial affinity. SDM-1965, old bulls post-1980 and some of the old bulls pre-1980 formed another cluster. Other non-Danish breeds emerged with subsequent increases in $K$-values. The results for $K = 9$ provided additional resolution including differentiation among the four Jutland cattle lineages, with individuals from the Vesterbølle-lineage showing partial membership in the cluster representing the Kortegaard-lineage. Modern SDM-1965 and the old bulls pre-1980 samples of SDM-1965 showed substantial overlap, whereas the old bulls post-1980 emerged as a highly admixed group with representation primarily in the Kortegaard- and Vesterbølle-lineages. Each of the Norwegian breeds and the Jersey cattle appeared as separate clusters, and there was no indication of introgression into the Jutland cattle. In contrast, there was no obvious distinction in ADMIXTURE between the Faroe Island and
Holstein individuals. The two individuals with duplicate samples showed consistent genomic profiles.

Genetic differentiation measured by pairwise $F_{ST}$ values varied between native and commercial breeds, and between the native breeds and lineages (Table 3). For Jutland cattle, pairwise comparisons between lineages all showed $F_{ST} > 0.1$ except for Oregaard-Vesterbølle. Moreover, pairwise comparison with Norwegian breeds and Faroe cattle all showed $F_{ST} > 0.1$ and suggested clearly differentiated populations. Although ADMIXTURE did not separate Faroe Island and Holstein cattle, the pairwise $F_{ST}$ values of 0.1131 showed some differentiation. The TreeMix results indicated that the old bulls post-1980 had experienced admixture with the Kortegaard-lineage and with SDM-1965 (Figure 5). Analyses of genetic drift between native Jutland Kortegaard and Oregaard-lineages were supported by PCA (Figure 2, Supplemental Figure S1 and S2), ADMIXTURE (Figure 4) and the $F_{ST}$ value of 0.1034. Among Jutland cattle lineages, Westergaard was shown to have experienced the highest level of genetic drift, followed by Kortegaard.

Results for $H_O$ and $H_E$ showed clear differences among breeds and lineages, also within the Jutland breed (Table 4). Genetic diversity was generally lower for Jutland cattle than for other groups, with the lowest values observed for the Westergaard-lineage. In contrast, old bulls pre-1980 had some of the highest values observed. Although the Holstein, Jersey and Faroe cattle had the lowest numbers of individuals tested, the Holstein and Faroe cattle showed comparatively high $H_E$ and $H_O$ values. Polymorphism also varied among populations, with the lowest value (67.8%) found in the Westergaard-lineage and the highest value (84.1%) observed in the Western Norwegian Fjord cattle (Table 4). Values for IBD differed among populations and in general showed an inverse pattern to that of genetic diversity, with the highest mean value observed in the Westergaard-lineage and the lowest mean value in the old bulls post-1980 followed by the Western Norwegian Fjord cattle (Table
4, Supplemental Table S3 and Supplemental Figure S3). The results for $N_E$ exhibited relatively broad variation with the smallest value found for the Westergaard-lineage and the highest estimate observed in the Holstein cattle (Table 4). The values for $P_A$ ranged from none in the old-bulls post-1980 to 221 being found in the Western Norwegian Red-poll cattle (Table 4). Within Jutland lineages, Kortegaard had the lowest number ($n = 2$) and Oregaard the highest ($n = 32$), with intermediate values for Vesterbølle ($n = 5$) and Westergaard ($n = 9$).

**Genetic drift and selection in native breeds**

Analyses of SNPs under potential selection identified 1985 loci, some of which were found in multiple tests (Supplemental Table S2). The number of observed outliers differed among tests, with T1 producing $n = 229$ loci, T2: $n = 62$, T3: $n = 336$, T4: $n = 199$, T5: $n = 235$, T6: $n = 425$ and T7: $n = 766$ outliers. After examination of all outliers (Supplemental Table S2) we selected $n = 146$ focal genes and genome regions for further examination based on their relevance for livestock and native breeds (Table 5, Supplemental Note S2). We categorized these findings as production traits (growth and meat quality, $n = 46$ genes), milk production ($n = 11$), reproduction ($n = 35$), physical appearance ($n = 2$), climate adaptation ($n = 6$), behaviour and cognition ($n = 10$), hormones ($n = 2$), infection and immunity ($n = 11$), metabolism ($n = 15$), and sensory including olfaction ($n = 3$), vision ($n = 3$) and hearing ($n = 2$) (Table 5, Supplemental Note S2, Supplemental Table S4), and these traits may be influenced by artificial and/or natural selection.

Comparisons of the outlier results show that the majority of genes identified were associated with production traits (growth/meat) and reproduction. The distribution of results was uneven, ranging from T2 where we identified one flanking gene (IGF1) linked to production traits, milk and reproduction, to T7 where we detected numerous genes associated
with production traits (n = 27), reproduction (n = 11) and other traits (Table 5). Genes linked to milk production were identified in all comparisons, with the majority identified in T3 (n = 4). In contrast, genes associated with physical appearance were found only in comparisons involving the Western Norwegian Red-polled cattle, and all results were related to the polled phenotype exhibited by this breed. Enrichment analyses identified GO categories (n = 19) for production traits that included cellular response to growth factors, muscle development, and bone formation (Supplemental Table S5), and for milk we observed categories (n = 6) primarily linked to transmembrane transport. For behaviour and cognition we identified GO categories (n = 11) including locomotory behaviour and synaptic signalling (Supplemental Table S5), whereas for hearing we detected categories (n = 3) for sensory perception of sound and mechanical stimulus, including equilibrioception (the sense of balance). Genes linked to other types of traits produced no significant GO results. Examination of ROHs found that none were shared among all individuals within breeds. We plotted ROH shared among at least six individuals per breed, and focal genes found within outlier flanking regions, across all autosomal chromosomes (Figure 6a,b).

The potential role of cryopreservation in preserving native breeds

Cryopreserved samples exhibited high genetic diversity, with seven P_A found in the pre-1980 sample of SDM-1965 cattle (Table 4). As expected, we also observed higher genetic variability in the 15 pre-1980 samples of SDM-1965 cattle (H_0 = 0.3034, H_E = 0.2860) than in the 20 modern samples of SDM-1965 (H_0 = 0.2842, H_E = 0.2838). Mean IBD was also higher (0.216 vs. 0.160) and P% very similar (81.3 vs. 81.2). The post-1980 samples of Jutland cattle did not exhibit P_A, but the 14 sampled bulls showed higher H_E and lower IBD than all contemporary Jutland lineages, and higher H_0 than all but one of them (the Vesterbølle-lineage, Table 4).
Discussion

Our genome-wide study of the amount and distribution of genetic diversity in native breeds documented distinct native and commercial cattle breeds. The Jutland cattle exhibited clear substructure, where increased $N_E$ may be needed for long-term conservation management, particularly given the reduced genetic diversity found within the lineages with the lowest census population sizes. Comparison of commercial and native breeds identified genes in the native breeds that merit further study concerning trade-offs between commercial output vs. animal health, robustness in low-input habitats, and adaptation to changing environments, such as immune function and metabolism. Native breeds may also comprise important genetic diversity for local adaptive variation, including heat tolerance, spatial learning and memory. Importantly, cryopreserved samples exhibit high and unique genetic variation, and present opportunities to increase variation and restore diversity from earlier generations. Finally, we demonstrated how genomics can help toward solving questions of broad interest within conservation and evolutionary genetics that also have practical relevance for breed managers.

Genetic diversity within and among native breeds

Our results suggest the Danish Jutland cattle offer an informative case study on the threats facing many native breeds around the world (FAO, 2015), and the resources these breeds represent for long-term preservation of genetic variation in domestic animals (Hoffmann, 2013). Our findings on genetic structure and diversity agree with earlier results (Brüniche-Olsen et al., 2012) showing Jutland cattle as a distinct breed (see also Supplemental Note S1). Although earlier microsatellite analyses produced mixed results for the Kortegaard-lineage, (Brüniche-Olsen et al., 2012) our study identified all four contemporary lineages as separate population clusters. The Jutland cattle exhibited genetic similarities to SDM-1965, as...
expected from shared ancestry. The two Norwegian native breeds had genetic profiles and diversity that appear consistent with their different phenotypes and history (Kantanen et al., 2000), whereas Faroe Island cattle and Holstein did not emerge as unequivocally distinct genetic units in our study. The genetic background of the Faroe Island cattle is complex (Li et al., 2005) and further genome-wide analyses with a broader range of breeds may be needed to resolve their history. Jutland cattle genetic diversity seems consistent with earlier microsatellite data suggesting comparatively high variability (Brüniche-Olsen et al., 2012), and our findings of reduced $H_E$, $H_O$ and polymorphism for the Westergaard-lineage are concurrent with the lower levels of mtDNA haplotype variation (Brüniche-Olsen et al., 2012) found for this lineage. However, a direct comparison between the two studies is not possible and differences could, at least in part, be explained by the properties of different marker types with microsatellites typically exhibiting rapid mutation rates and a bias toward highly polymorphic loci (Brandström & Ellegren, 2008). An earlier study based on data from a lower-density chip with 50K SNPs reported lower genetic diversity in the Kortegaard-lineage than for a range of commercial breeds (Pertoldi et al., 2014). Moreover, Pertoldi et al. (2014) reported $H_O = 0.266$ and $H_E = 0.259$, and we observed 0.262 and 0.247, respectively; thus similar levels of heterozygosity were found in the two studies.

Within the Jutland cattle we observed the highest IBD-values for the Westergaard-lineage, followed by Kortegaard, Oregaard and Vesterbølle, whereas Brüniche-Olsen et al. (2012) found strong signatures of inbreeding (significant deviation from Hardy-Weinberg equilibrium due to heterozygosity deficiency) for all groups except Oregaard. Inclusion of different marker types, sample sizes and individuals may have contributed to discrepancies among studies (Supplemental Note S1). However, the range of observed IBD-values suggests there are opportunities to carefully select individuals to minimize future inbreeding in ongoing conservation management. Additional considerations for selection of
breeding individuals are long-term conservation of SNPs under potential selection, and preservation of such variants may be considered a key aim of breeding strategies for native breeds. We observed relatively low polymorphism (67.8%) and high IBD (0.384) in the 16 Westergaard-individuals compared to that found in nine Danish Holstein cattle (69.9% and 0.327, respectively). However, in a broader perspective, the polymorphism and IBD values for the investigated native breeds suggest these comprise important diversity relative to their considerably smaller census population sizes than those of commercial breeds (Table 1). Altogether 1019 native cattle breeds have been reported globally, with 369 in Europe and the Caucasus (FAO, 2015). Understanding and preserving the genetic diversity of native breeds, especially those with small population sizes, is thus a priority, including assessments of epigenetic processes linked to environmental factors (e.g., diet) (FAO, 2015). Adaptation to changing environmental conditions will become increasingly important (Hoffmann, 2013), and the existing diversity within the many native cattle breeds could help facilitate more rapid adaptation.

Diverse patterns of variation may emerge from the use of different markers (Brandström & Ellegren, 2008), but we believe the detailed SNP profiles to be representative for contemporary Jutland cattle breed lineages and to reflect genome-wide variability and structure within and among breeds in our study. Overall, the results from this and earlier research on Jutland cattle indicate a distinct breed with multiple lineages. Moreover, the $P_A$ values suggest that considerable variation may be found in smaller populations of native breeds. The old bulls post-1980 (Jutland bulls) was the only group that did not exhibit any $P_A$. This is likely explained by its variants being shared among contemporary lineages, which seems consistent with the TreeMix results of admixture involving this group. In contrast, the old bulls pre-1980 (SDM-1965) exhibited seven $P_A$, which seems consistent with the group’s high levels of heterozygosity. The two Norwegian breeds also showed a high number of $P_A$.\[This article is protected by copyright. All rights reserved.\]
despite modest values for $N_E$. The small remaining populations of Vesterbølle and Westergaard, which represent the upper and lower $N_E$ estimates for Jutland lineages (i.e., subpopulations) with four and 21 individuals, respectively, illustrate the difficulties inherent in managing such lineages as separate units. As suggested for isolated populations of wild species with small $N_E$ (Laikre et al., 2016), long-term management of native livestock breeds might benefit from a metapopulation approach where lineages are considered as metapopulations. Earlier simulations have also suggested that augmenting the breeding pool with individuals of both sexes could help increase the probability of persistence (Hertz et al., 2016).

Genetic drift and selection in native breeds

Our analyses of outliers indicated that the breeds have evolved in consequence of both artificial and natural selection. The relatively high number of genes found associated with production traits and reproduction is likely influenced by the importance of these features for commercial breeds and thus for development of the SNP panel used in our study. The enrichment results also seem to suggest strong research focus on genomic regions important for livestock production (meat and milk) and potentially also for other areas within biology and evolution (hearing, behaviour and cognition). The overall distribution of results among outlier tests appears to support our expectations of broad genome-wide differences among the investigated populations and the evolutionary uniqueness of native breeds. Danish native cattle exhibited differentiation from commercial breeds, which has also been reported from other regions (e.g., Kawahara-Miki et al., 2011; Iso-Touru et al., 2016; Lim et al., 2016). Pairwise tests between native breeds also showed signs of divergent selection, which is in accordance with findings from other parts of the world (Iso-Touru et al., 2016; Stucki et al., 2017). Certain characteristics, such as olfaction (McRae et al., 2013), hormonal cues
concerning reproduction (Jiang et al., 2006), and cognitive features including visual map development (Xu et al., 2011) and spatial memory (Qiu et al., 2010; Bailey et al., 2015) (see also Supplemental Tables S4 and S5) could be vital for conserving native livestock and their role in extensive agriculture, where native breeds may be subject to in situ natural selection while contributing to grazing and biodiversity maintenance priorities (Halada et al., 2011; Timmermann et al., 2014; Bailey et al., 2015).

Our results suggest native breeds may carry genetic variants important for adaptation to a rapidly changing environment due to climate change and other anthropogenic activities, and such variants could be critical for their preservation and contribute to increased environmental tolerance in commercial breeds via modern genomic techniques (O’Neill et al., 2010; Rauw & Gomez-Raya, 2015). Such advances could affect livestock survival and development related to important environmental stressors including infectious disease and parasite resistance (O’Neill et al., 2010; Kadarmideen et al., 2011; Porto-Neto et al., 2014). Our findings may also indicate trade-offs between production gains and animal health (O’Neill et al., 2010; Rauw & Gomez-Raya, 2015; Takasuga et al., 2015). For example, genes related to metabolism and ketosis suggests potential conflicts between human-induced artificial selection toward high milk production and the optimal energy balance of individual animals (Mulligan & Doherty, 2008). Moreover, natural selection for more robust individuals where a higher proportion of the food intake is allocated toward energy reserve maintenance and survival (and consequently lowers productive output) merits further attention for animal welfare and long-term evolutionary potential.

For ROHs shared among six or more individuals per breed there was limited overlap with focal genes flanking outlier SNP loci. Results supported by multiple analytical approaches will typically have higher support, although different methods with various underlying approaches can also provide diverse results (Narum & Hess, 2011; François et al.,

This article is protected by copyright. All rights reserved.
The influences of selection, drift and inbreeding on ROH can be difficult to resolve (Purfield et al., 2012), in particular for small populations such as Jutland cattle. We cannot, based on our investigation of outliers and ROH, show evidence for selection of any particular trait in Jutland cattle or other native breeds. Yet taken together, the broad range of genes associated with traits under natural or human-induced selection indicate that Jutland cattle and other native breeds represent genomic resources relevant for farming practices concerned with sustainability, animal welfare and adaptation to climate change.

There are limiting factors that should be considered for interpretation of our results, including potential ascertainment bias in the bovine array developed with a focus on commercial breeds (Supplemental Note S1). Moreover, we cannot exclude the possibility that small sample sizes may have affected our results, including the analyses of outlier loci. Reports from earlier simulations with pcadapt indicated benefits from increasing sample sizes from 20 to 60 individuals (Luu et al., 2016). However, as certain Jutland cattle lineages now have few remaining individuals (e.g., we obtained only 16 from Westergaard and 20 from Vesterbølle) we focused on tests with equalised sample sizes and a high number of LD-pruned loci. Accordingly, for these small populations we are confident of having included most of the genomic variation (included on the SNP array) that is still in existence.

The potential role of cryopreservation in preserving native breeds

A possibility emerging from our findings, consistent with recommendations from earlier studies (Charlton et al., 2018), is the potential to infuse variability into genetically depauperate lineages by means of original genetic diversity from older cryopreserved samples. These older semen samples could represent some of the existing variability prior to dramatic declines in census size, and may be considered for inclusion in conservation breeding plans. For the Jutland cattle, this would appear especially relevant for the...
Westergaard and Vesterbølle-lineages, where it is difficult to see management of genetically closed populations as a long-term viable option.

Conclusion
Our analyses demonstrate the potential use of genomics for investigating the genetic structure and unique variation of native domestic breeds, and how these may differ from commercial breeds in traits relevant for production and environmental tolerance. Our results, using Northern European extensive native cattle as a case study, also underline how these differences may have implications for human-induced artificial and natural selection of both native and commercial breeds. We show the potential for addressing basic evolutionary inquiries and applied conservation questions that can help managers create conservation plans to preserve genetic uniqueness while maintaining in situ selection in native domestic breeds, so these can continue to be shaped by their local environmental conditions even if the $N_E$ is low (and thus selection has limited ability to act on the frequency of SNPs under potential selection). Furthermore, the implementation of an appropriate breeding plan will help increase the $N_E$, which will in turn reduce the fluctuations in allelic frequencies across generations. At the same time, assortative mating strategies may be chosen for individuals that carry genes of interest, to augment the relevant allelic frequencies and reduce the risk of their loss from the population.

Acknowledgements
who contributed samples and information toward this study. We also thank reviewers for their helpful comments. AVS received funding from the Danish Natural Science Research Council (grant 1337-00007) and a senior postdoctoral fellowship from Insubria University in Varese, Italy. LI received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Action (grant 656697). The research was supported by the Ministry of Environment and Food of Denmark and the Aalborg Zoo Conservation Foundation (AZCF, grant 7-2017).

Data archiving statement:
Data for this study are available at: https://doi.org/10.5061/dryad.hr5d4g8. Data include the PLINK files (bed, bim, fam) for 195 individual cattle and 677,311 autosomal single nucleotide polymorphism loci based on the UMD 3.1 bovine genome assembly (https://www.ncbi.nlm.nih.gov/genome?term=bos%20taurus) and filtered in PLINK for genotyping success of minimum 98%. Additionally, we include an Excel file with information on each of the 195 individuals, including PLINK family ID (FID) and individual ID (IID), breed/lineage and, if relevant, identity codes such as Danish stud book number.

References


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


Table 1. The populations included in the study, with cattle breed, number of animals sampled, and the estimated number of existing animals per breed at present. Where relevant, we have noted where a lineage or subsampled population belongs to the Jutland cattle by adding this name in brackets. For Jutland cattle lineages Kortegaard and Oregaard, we subsampled n=20 individuals from each group to equalise sample sizes among groups (details in Methods).

<table>
<thead>
<tr>
<th>Cattle breed</th>
<th>Number sampled</th>
<th>Estimated number in existence</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jutland cattle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregaard-lineage [Jutland]</td>
<td>131</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vesterbølle-lineage [Jutland]</td>
<td>186</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Westergaard-lineage [Jutland]</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old bulls pre-1980 ^1</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old bulls post-1980 ^1 [n = 14 Jutland]</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM-1965</td>
<td>20</td>
<td>212 (2015)</td>
<td>Denmark</td>
</tr>
<tr>
<td>Western Norwegian Fjord cattle</td>
<td>21</td>
<td>692 (2015)</td>
<td>Norway</td>
</tr>
<tr>
<td>Western Norwegian Red-poled cattle</td>
<td>19</td>
<td>139 (2015)</td>
<td>Norway</td>
</tr>
<tr>
<td>Faroe cattle</td>
<td>8</td>
<td>Circa 40 ^6</td>
<td>Faroe Islands</td>
</tr>
<tr>
<td>Holstein</td>
<td>9</td>
<td>904,045 (2016) ^7</td>
<td>Denmark ^5</td>
</tr>
<tr>
<td>Jersey</td>
<td>9</td>
<td>142,179 (2016) ^8</td>
<td>Denmark ^8</td>
</tr>
</tbody>
</table>

^1DAD-IS (2017b). Numbers within the various Danish Jutland cattle lineages were not available.
^2Old bulls pre-1980 are cryopreserved semen samples of SDM-1965 included to evaluate temporal changes in genetic diversity and structure.
^3Old bulls post-1980 are cryopreserved semen samples of n = 14 Danish Jutland cattle (one sampled in duplicate) and n = 4 SDM-1965.
^4DAD-IS (2017a)
^5Sæther & Rehnberg (2016)
^7DAD-IS (2017d). This report refers to Danish Holstein and thus the Danish population of this international breed originating from The Netherlands.
^8DAD-IS (2017e). This report refers to Danish Jersey and thus the Danish population of this international breed originating from the island of Jersey in the United Kingdom.
Table 2. Tests for the presence of outlier loci in cattle breeds, including one test across all breeds (T1), and one test used as a measure of control (T2) where differentiation is expected to be explained by recent genetic drift without any known history of selection. A description of cattle breeds and lineages is provided in Table 1.

<table>
<thead>
<tr>
<th>Test</th>
<th>Cattle included</th>
<th>Test type</th>
<th>Performed to evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>All</td>
<td>Across all groups (n=9)</td>
<td>Outliers among all groups</td>
</tr>
<tr>
<td>T2</td>
<td>Jutland Kortegaard vs. Oregard lineages</td>
<td>Pairwise</td>
<td>Genetic drift in native breed lineages (few/no results expected)</td>
</tr>
<tr>
<td>T3</td>
<td>Old bulls pre-1980 (cryopreserved SDM-1965) vs. Holstein cattle</td>
<td>Pairwise</td>
<td>Native vs. commercial breed; selection for local environmental conditions in native cattle</td>
</tr>
<tr>
<td>T4</td>
<td>Old bulls pre-1980 (cryopreserved SDM-1965) vs. Western Norwegian Red-pollled cattle</td>
<td>Pairwise</td>
<td>Two native breeds; selection for different landscape type (rugged terrain in Western Norway, gentle terrain in Denmark)</td>
</tr>
<tr>
<td>T5</td>
<td>Western Norwegian Red-pollled cattle vs. Western Norwegian Fjord cattle</td>
<td>Pairwise</td>
<td>Two native breeds; selection for polled phenotype (without horns) in Red-pollled cattle</td>
</tr>
<tr>
<td>T6</td>
<td>Old bulls post-1980 (cryopreserved Jutland) vs. Holstein cattle</td>
<td>Pairwise</td>
<td>Native vs. commercial breed; selection for local environmental conditions in native cattle</td>
</tr>
<tr>
<td>T7</td>
<td>Old bulls post-1980 (cryopreserved Jutland) vs. Western Norwegian Red-pollled cattle</td>
<td>Pairwise</td>
<td>Two native breeds; selection for different landscapes type (rugged terrain in Western Norway, gentle terrain in Denmark)</td>
</tr>
</tbody>
</table>
Table 3. Estimates of genetic differentiation calculated as $F_{ST}$ for all pairs of cattle breeds/lineages. All pairwise comparisons were significant at $p < 0.001$, except the values marked with 1 ($p = 0.463$), 2 ($p = 0.004$) and 3 ($p = 0.002$) after overall Bonferroni correction for multiple comparisons (Rice, 1989) ($K = 66$).

<table>
<thead>
<tr>
<th>Breed/lineage</th>
<th>Kortegaard (n=20)</th>
<th>Oregaard (n=20)</th>
<th>Vesterbølle (n=20)</th>
<th>Westergaard (n=16)</th>
<th>Old bulls pre-1980 (n=15)</th>
<th>Old bulls post-1980 (n=14*)</th>
<th>SDM-1965 (n=20)</th>
<th>Vestlandsk Fjordfe (n=21)</th>
<th>Vestlandsk Raudkolle (n=19)</th>
<th>Faroe cattle (n=8)</th>
<th>Danish Holstein (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jutland</td>
<td>0.1034</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oregaard-</td>
<td>0.1047</td>
<td>0.0780</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lineage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jutland Vesterbølle-lineage</td>
<td>0.1822</td>
<td>0.1457</td>
<td>0.1238</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jutland Westergaard-lineage</td>
<td>0.1121</td>
<td>0.0959</td>
<td>0.0709</td>
<td>0.1403</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old bulls pre-1980</td>
<td>0.0609</td>
<td>0.0487</td>
<td>-0.0009¹</td>
<td>0.1046</td>
<td>0.0486</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old bulls post-1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=14*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDM-1965</td>
<td>0.1192</td>
<td>0.0995</td>
<td>0.0701</td>
<td>0.1403</td>
<td>0.0163²</td>
<td>0.0491</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Norwegian Fjord cattle</td>
<td>0.1371</td>
<td>0.1159</td>
<td>0.1016</td>
<td>0.1618</td>
<td>0.0833</td>
<td>0.0801</td>
<td>0.0919</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western Norwegian Red-polled cattle</td>
<td>0.1589</td>
<td>0.1358</td>
<td>0.1217</td>
<td>0.1838</td>
<td>0.1048</td>
<td>0.1007</td>
<td>0.1125</td>
<td>0.0713</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faroe cattle</td>
<td>0.1742</td>
<td>0.1488</td>
<td>0.1301</td>
<td>0.2030</td>
<td>0.1061</td>
<td>0.1069</td>
<td>0.1148</td>
<td>0.0849</td>
<td>0.1142</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>(n=8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved.
<table>
<thead>
<tr>
<th></th>
<th>Holstein (n=9)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1565</td>
<td>0.1326</td>
<td>0.1096</td>
<td>0.1866</td>
<td>0.0872</td>
<td>0.0866</td>
<td>0.0898</td>
<td>0.0900</td>
<td>0.1116</td>
</tr>
<tr>
<td>Jersey (n=9)</td>
<td>0.2180</td>
<td>0.1912</td>
<td>0.1749</td>
<td>0.2474</td>
<td>0.1547</td>
<td>0.1529</td>
<td>0.1614</td>
<td>0.1242</td>
<td>0.1524</td>
</tr>
</tbody>
</table>

*Including only cryopreserved semen samples of n = 14 Danish Jutland cattle recognized in the stud book.
Table 4. The number of samples analysed (No), observed ($H_O$) and expected heterozygosity ($H_E$) values with standard error (S.E.), percent polymorphic loci (P%), identity by descent (IBD) per cattle group, effective population size with 95% confidence interval ($N_E$), and number of private alleles ($P_A$). $N_E$ and $P_A$ were calculated with LD-pruned data. Further IBD details are provided in Supplemental Table S3 and Supplemental Figure S3.

<table>
<thead>
<tr>
<th>Cattle breed/lineage</th>
<th>No</th>
<th>$H_O$ (S.E.)</th>
<th>$H_E$ (S.E.)</th>
<th>P%</th>
<th>IBD Mean (range)</th>
<th>$N_E$</th>
<th>$P_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jutland Kortegaard-lineage</td>
<td>20</td>
<td>0.2618 (0.00026)</td>
<td>0.2474 (0.00023)</td>
<td>73.1</td>
<td>0.296 (0.642)</td>
<td>9.1 (9.1-9.2)</td>
<td>2</td>
</tr>
<tr>
<td>Jutland Oregaard-lineage</td>
<td>20</td>
<td>0.2776 (0.00026)</td>
<td>0.2647 (0.00023)</td>
<td>75.8</td>
<td>0.261 (0.677)</td>
<td>10.0 (10.0-10.0)</td>
<td>32</td>
</tr>
<tr>
<td>Jutland Vesterbølle-lineage</td>
<td>20</td>
<td>0.2833 (0.00026)</td>
<td>0.2752 (0.00022)</td>
<td>78.8</td>
<td>0.218 (0.620)</td>
<td>21.6 (21.5-21.7)</td>
<td>5</td>
</tr>
<tr>
<td>Jutland Westergaard-lineage</td>
<td>16</td>
<td>0.2532 (0.00028)</td>
<td>0.2267 (0.00023)</td>
<td>67.8</td>
<td>0.384 (0.187-0.724)</td>
<td>4.0 (4.0-4.1)</td>
<td>9</td>
</tr>
<tr>
<td>Old bulls pre-1980</td>
<td>15</td>
<td>0.3034 (0.00026)</td>
<td>0.2860 (0.00022)</td>
<td>81.3</td>
<td>0.216 (0.591)</td>
<td>24.8 (24.7-24.9)</td>
<td>7</td>
</tr>
<tr>
<td>Old bulls post-1980*</td>
<td>14</td>
<td>0.2800 (0.00024)</td>
<td>0.2857 (0.00022)</td>
<td>80.7</td>
<td>0.134 (0.685)</td>
<td>11.8 (11.7-11.8)</td>
<td>-</td>
</tr>
<tr>
<td>SDM-1965</td>
<td>20</td>
<td>0.2842 (0.00024)</td>
<td>0.2838 (0.00022)</td>
<td>81.2</td>
<td>0.160 (0.592)</td>
<td>18.0 (17.9-18.0)</td>
<td>9</td>
</tr>
<tr>
<td>Western Norwegian Fjord cattle</td>
<td>21</td>
<td>0.3035 (0.00024)</td>
<td>0.2983 (0.00021)</td>
<td>84.1</td>
<td>0.151 (0.566)</td>
<td>35.0 (34.9-35.1)</td>
<td>216</td>
</tr>
<tr>
<td>Western Norwegian Red-polled cattle</td>
<td>19</td>
<td>0.2980 (0.00025)</td>
<td>0.2849 (0.00022)</td>
<td>80.6</td>
<td>0.216 (0.609)</td>
<td>18.0 (18.0-18.1)</td>
<td>221</td>
</tr>
<tr>
<td>Faroe cattle</td>
<td>8</td>
<td>0.2997 (0.00029)</td>
<td>0.2707 (0.00023)</td>
<td>75.0</td>
<td>0.293 (0.196-0.646)</td>
<td>12.2 (12.2-12.3)</td>
<td>55</td>
</tr>
<tr>
<td>Holstein</td>
<td>9</td>
<td>0.3053 (0.00028)</td>
<td>0.2836 (0.00022)</td>
<td>77.0</td>
<td>0.241 (0.209-0.341)</td>
<td>92.0 (89.6-94.4)</td>
<td>45</td>
</tr>
<tr>
<td>Jersey</td>
<td>9</td>
<td>0.2705 (0.00029)</td>
<td>0.2483 (0.00023)</td>
<td>69.9</td>
<td>0.327 (0.272-0.506)</td>
<td>28.9 (28.6-29.2)</td>
<td>112</td>
</tr>
</tbody>
</table>

*Including only cryopreserved semen samples of n = 14 Danish Jutland cattle recognized in the stud book.
Table 5. Focal genes found in the 3000 bp flanking regions of outlier single nucleotide polymorphism loci in native and commercial cattle.

Details on tests T1-T7 and populations are given in Table 2; gene information and references are provided in Supplemental Note S2 and Supplemental Table S4.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Production traits (growth, meat)</td>
<td>ME1, TRPV4, LRP2</td>
<td>LRP2, CTTNBP2NL, MECOM, PTPN1</td>
<td>PRDM16, ADAMTSL3</td>
<td>FLT1</td>
<td>ARHGEF3, BMP7, CEP128, CYP2J2, ERCC6L2, ESRRG, FBN1, FG3, FLVCR1, IARS, MYOM3, PALM2, PHLDB2, ROR2, SLC8A1, ZNFX1</td>
<td>AKT3, BBX, BMP7, COL5A1, DCHS2, DIS3L2, DSCAM, ESRRG, FAP, FBLN7, FBN1, FG3, FLVCR1, F10, GLRA1, GUCY1A1, HPS5, IARS, MYBPC1, PDE1A, PHLDB2, RIN3, ROR2, RPS6KA4, RUNX2, SMURF1, TENM4, TTN, ITGA6, SLC5A1</td>
<td></td>
</tr>
<tr>
<td>Milk production</td>
<td>BTC, ITGA6, PDE4D</td>
<td>IGF1</td>
<td>ABCG2, CTNNB1, TCF7L2, CUX1, LRP2, GABRA4, LHX6, LRRCD34, NR3C1, ABCG2, CTTNBP2NL, RXFP1/LGR7</td>
<td>SLC2A8/GLUT1, ATP1A1, ELMO1</td>
<td>PDE4D</td>
<td>ITGA6</td>
<td>AVEN, CD9, EBF1, GSTM1, SPEGF2, NPHP4, SPATA31A3, STAU2, UNC5C, ZNF462</td>
</tr>
<tr>
<td>Reproduction</td>
<td>ITGA6, HSD17B12, SOX5, DPH6/ATPBD4, LRP2, BIRC5</td>
<td>IGF1</td>
<td>LRP2, GABRA4, LHX6, LRRCD34, NR3C1, ABCG2, CTTNBP2NL, RXFP1/LGR7</td>
<td>SLC2A8/GLUT1, ATP1A1, ELMO1</td>
<td>PDE4D</td>
<td>ITGA6</td>
<td>AVEN, CD9, EBF1, GSTM1, SPEGF2, NPHP4, SPATA31A3, STAU2, UNC5C, ZNF462</td>
</tr>
<tr>
<td>Physical appearance</td>
<td>UBE2E3, WSB2, ZP2, KRTAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Climate adaptation</td>
<td>ACBD6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behaviour and cognition</td>
<td>DAB1, DSCAM, NETO1, TMEM132D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infection and immunity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory - olfaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory - vision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory - hearing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>NUDCD3</strong></th>
<th><strong>HSPBP1, SLC18A1/VMAT-1, CERS6, GRID2/LOC53636</strong></th>
<th><strong>SLC18A1/VMAT-1, ATP1A1, RPTOR CHRN62</strong></th>
<th><strong>LSAMP</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SLC26A4</strong></td>
<td><strong>SAF862</strong></td>
<td><strong>TRAPPC9, NFATC2IP</strong></td>
<td><strong>TRAPPC9, SPAG11B</strong></td>
</tr>
<tr>
<td><strong>DMKN, TRAPPC9</strong></td>
<td><strong>NDUFA10</strong></td>
<td><strong>OXCT1</strong></td>
<td><strong>NDUFA10</strong></td>
</tr>
<tr>
<td><strong>LOC783998/OR11G2</strong></td>
<td><strong>LIPF/HGL, PANK1</strong></td>
<td><strong>NDUFA10</strong></td>
<td><strong>NDUFA10</strong></td>
</tr>
<tr>
<td><strong>RORA, BTC, GAPDHS</strong></td>
<td><strong>CPNE3, (Ig)M/CADM2, PLCXD3, SIGIRR, EBF1, HACE1, KLHL32,</strong></td>
<td><strong>OXCT1</strong></td>
<td><strong>NDUFA10</strong></td>
</tr>
<tr>
<td><strong>LOC783998/OR11G2</strong></td>
<td><strong>NDUFA10</strong></td>
<td><strong>OXCT1</strong></td>
<td><strong>NDUFA10</strong></td>
</tr>
<tr>
<td><strong>CTNND2, NPHP4</strong></td>
<td><strong>CTNND2, TENM3</strong></td>
<td><strong>PCDH15, TSPEAR</strong></td>
<td><strong>CTNND2, TENM3</strong></td>
</tr>
</tbody>
</table>
Figure legends

Figure 1. Proposed timeline for founding of the four contemporary lineages of the Danish Jutland cattle breed. The Westergaard-lineage is deemed to be the oldest, although the precise time of its origin is unknown.

Figure 2. Principal Component Analyses (PCA) with 195 individuals showing the first and second PC axes. Cattle breeds/lineages (denoted in figure legend as Pop) are FI: Faroe Island cattle, HOL: Holstein, JER: Jersey, JK: Jutland cattle Kortegaard-lineage, JO: Jutland cattle Oregaard-lineage, JV: Jutland cattle Vesterbølle-lineage, JW: Jutland cattle Westergaard-lineage, OB1: Old bulls pre-1980 (cryopreserved semen samples from SDM-1965 cattle), OB2: Old bulls post-1980 (cryopreserved from n=14 Jutland and n=4 SDM-1965 cattle), SDM: SDM-1965 cattle, WNF: Western Norwegian Fjord cattle, WNR: Western Norwegian Red-pollled cattle.

Figure 3. ADMIXTURE analyses of cattle with cross-validation (CV) error plot for K-values from 2-9 with 195 individuals and 88,190 single nucleotide polymorphism loci. The CV error is markedly reduced with each increase in K until K = 5. Hereafter it declines more slowly toward the minimal value at K = 9. Increases in K beyond this level are not supported.

Figure 4. ADMIXTURE plots for K-values from 2-9 clusters with 195 individuals and 88,190 single nucleotide polymorphism loci. Each vertical bar represents one individual, and the Y-axis shows individual ancestry (range 0-1). Cattle breeds/lineages are FI: Faroe Island cattle, HOL: Holstein, JER: Jersey, JK: Jutland cattle Kortegaard-lineage, JO: Jutland cattle Oregaard-lineage, JV: Jutland cattle Vesterbølle-lineage, JW: Jutland cattle Westergaard-lineage, OB1: Old bulls pre-1980 (cryopreserved semen samples from SDM-1965 cattle),

**Figure 5.** Maximum likelihood tree for 12 cattle populations with 195 individuals and 595,025 SNPs pruned for minor allele frequency of 1% and genotyping success of 98%. The scale bar on the horizontal axis shows 10x the average standard error of the sample covariance matrix, and the length of horizontal branches are proportional to the amount of genetic drift the populations have experienced. Cattle breeds/lineages are FI: Faroe Island cattle, HOL: Holstein, JER: Jersey, JK: Jutland cattle Kortegaard-lineage, JO: Jutland cattle Oregaard-lineage, JV: Jutland cattle Vesterbølle-lineage, JW: Jutland cattle Westergaard-lineage, OB1: Old bulls pre-1980 (cryopreserved semen samples from SDM-1965 cattle), OB2: Old bulls post-1980 (cryopreserved from n=14 Jutland and n=4 SDM-1965 cattle), SDM: SDM-1965 cattle, WNF: Western Norwegian Fjord cattle, WNR: Western Norwegian Red-pollled cattle. Migration arrows are coloured according to their weight, and indicate admixture between Old bulls post-1980 and Kortegaard, as well as between Old bulls post-1980 and SDM-1965.

**Figure 6.** Plots displaying runs of homozygosity (ROH) per autosomal chromosome and outliers from pairwise comparisons of native and commercial breeds central to this study. Outliers and ROH could both be indicative of selection, and we mapped our findings to examine the degree of overlap. ROH shared by at least six individuals per breed are shown in vertical coloured lines. Outlier loci where focal genes were found within 3000 bp flanking regions are marked as dark grey horizontal lines. The plots show pairwise comparisons of (a) Jutland cattle vs. Holstein cattle, and (b) Jutland cattle vs. Western-Norwegian Red-pollled cattle.

This article is protected by copyright. All rights reserved.
Cross-validation error

K

This article is protected by copyright. All rights reserved.