Arch. Anim. Breed., 64, 91–102, 2021
https://doi.org/10.5194/aab-64-91-2021

Open Access
Archives Animal Breeding

Genetic diversity and population structure in divergent German cattle selection lines on the basis of milk protein polymorphisms

Lisa G. Hohmann, Christina Weimann, Carsten Scheper, Georg Erhardt, and Sven König

Institute of Animal Breeding and Genetics, University of Giessen, 35390 Giessen, Germany

Correspondence: Sven König (sven.koenig@agrar.uni-giessen.de)

Received: 14 October 2020 – Revised: 1 January 2021 – Accepted: 5 February 2021 – Published: 11 March 2021

Abstract. The aim of this study was to analyze the genetic structure of the casein cluster in eight selection lines of the Holstein Friesian (HF), German Simmental (GS) and German Black Pied cattle ("Deutsches Schwarzbuntes Niederungsrind", DSN) breeds. A total of 2962 milk samples were typed at α_{s1}-casein (α_{s1}-CN), β-casein (β-CN), α_{s2}-casein (α_{s2}-CN) and κ-casein (κ-CN) loci using isoelectric focusing. The number of alleles per locus ranged from one (α_{s2}-CN) to five (β-CN), and the average expected heterozygosity and polymorphic information content of all loci were 0.33 and 0.27, respectively. The unrooted dendrogram revealed that the selection lines of the endangered DSN breed were clearly separated from the HF and GS breeds due to their predominance of the β-CN A1 allele and the comprehensive haplotype BA1A (in the abbreviation of α_{s1}-β-κ-CN). Temporal changes in allele distributions indicated decreasing genetic diversity at the casein loci, explaining the moderate level of genetic differentiation among selection lines (7.1 %). The variability of the casein should be exploited in future using breeding programs to select genetic lines for specific protein production in bovine milk but also to preserve biodiversity.

1 Introduction

Since domestication 8000–10 000 years ago, natural as well as man-made factors including geography, environment, culture and directional artificial selection contributed to cattle trait modifications phenotypically and genetically (Loftus et al., 1994). From a time perspective, in contrast to natural selection, artificial selection has the ability to change the genome rapidly. The consequence is a targeted displacement in allele frequencies, implying deviations from Hardy–Weinberg equilibrium (HWE) (Lachance, 2009). Two types of selection appear on the genomic level. Positive (Darwinian) selection promotes the spread of beneficial alleles, so that frequencies for these alleles increase and the selected alleles might be fixed over generations (Maynard Smith and Haigh, 1974; Kreitman, 2000). Negative or purifying selection hinders the spread of unfavorable alleles, causing decreasing allele frequencies up to the complete loss from the population (Kreitman, 2000). Selection not only affects the favored or unfavored mutations directly. In addition, selection causes a "hitchhiking" effect on the frequency of neutral alleles at linked loci (Maynard Smith and Haigh, 1974). The cattle genome therefore represents an opportunity for the identification of genetic variation that contributes to phenotypic diversity and for inferring genome responses to strong artificial selection. The different methods to detect selection signatures are based either on the distribution of allele frequencies, on the properties of haplotypes segregating within populations or on genetic differentiation between populations (reviewed by Hohenlohe et al., 2010).

Along with divergent selection criteria, the long-lasting intensive specific improvement of economically important traits contributed to the formation of diverse genetic lines within breeds. For example, artificial selection in the dual-purpose Simmental breed implied the establishment of divergent strains which are specialized for either milk or meat production (Campbell and Marshall, 2016). As the future requires promotion of more efficient sustainable livestock systems and utilization of greater proportions of non-human competitive products for animal feed, attention is given on adaptation to grazing systems (Delaby et al., 2018). Pasture-based systems reflect harsh environments, emphasizing the importance of animal traits associated with grazing behavior and robustness. Functional traits required under grazing include feed efficiency, health, fertility and longevity (Washburn and Mullen, 2014). In predestinated locations in Ireland or New Zealand, the development of grazing systems is accompanied by animal breeding and selection strategies on adaptation to local conditions (e.g., Lopez-Villalobos et al., 2000). The New Zealand total merit index favors robust, lightweight, long-living and efficient milk producing pasture converters (Jaeger, 2018). However, German Holstein (HF_G) cows have been selected for modern and large-scale indoor systems during decades, raising questions of possible genotype–environment interactions with impact on adaptation capabilities to harsh environments (König et al., 2005). In consequence, so-called "pasture breeding projects" were initiated in Germany (Brügemann et al., 2015; May et al., 2017), aiming at genetic line comparisons in grassland systems. Specific pasture-based selection lines within the HF_G breed were created by mating, e.g., HF_G cows with Holstein Friesian sires from New Zealand (HF_NZ). The close genetic relationships between selection lines with the same founder animals suggest genetic comparisons on the basis of milk protein compositions, in order to study effects of selection in dairy lines during the past decades.

Genes influencing milk yield and protein content are the casein genes *CSN1S1*, *CSN1S2*, *CNS2* and *CSN3*, encoding the proteins α_{s1}-casein (α_{s1}-CN), β-casein (β-CN), α_{s2}-casein (α_{s2}-CN) and κ-casein (κ-CN), respectively (Ng-Kwai-Hang et al., 1984). Several single-nucleotide polymorphisms (SNPs) of the casein genes change their protein sequences, implying different casein variants. A recent review of milk protein nomenclature (Gallinat et al., 2013) indicated 10 variants for α_{s1}-CN (A, B, C, D, E, F, G, H, I, J), 15 for β-CN (A1, A2, A3, B, C, D, E, F, G, H^1, H^2, I, J, K, L), five for α_{s2}-CN (A, B, C, D, E) and 14 for κ-CN (A, A1, B, B2, C, D, E, F^1, F^2, G^1, G^2, H, I, J) in *Bos genus*. The tight genetic linkage among the casein genes within a 250 kb cluster on chromosome 6 (BTA6) implies limited recombination and suggests the creation of casein haplotypes (Ferretti et al., 1990; Lien et al., 1993). Casein polymorphisms were used for the characterization of domesticated breeds and for tracing the evolutionary history (Caroli et al., 2009). Beja-Pereira et al. (2002) and Jann et al. (2004) provided evidence for a geographically associated distribution of casein haplotypes, and they identified a decline of genetic diversity for taurine breeds in Europe from the south to the north and from the east to the west. Mahé et al. (1999) discriminated between *Bos taurus* and *Bos indicus* origins at the milk protein level. Furthermore, casein genes harbor a number of variants with beneficial effects on milk production, milk composition and technological properties (reviewed by Caroli et al., 2009). Additionally, numerous studies (e.g., Ehrmann et al., 1997; Çardak et al., 2003) focused on the effects of polymorph milk proteins on the individual milk protein content. For example, Ng-Kwai-Hang et al. (1984) identified causal relationships between the homozygote genotypes BB of the respective casein α_{s1}-CN and κ-CN with the protein and casein content of milk. Protein yield and protein percentage are included into the overall production index (RZM) for German dairy cattle since decades and have been used as a major selection criterion (König et al., 2007). In consequence, monitoring casein genetic variants is a useful tool to inferring signatures of selection.

To our knowledge, there are no studies addressing genetic diversity in individual selection lines – especially in pasture-based selection lines – based on alleles and haplotypes of the whole casein cluster. We hypothesize that divergent directions of positive selection (e.g., towards pasture ability, dairy or meat production) have altered allele and haplotype frequencies of the casein. Therefore, the aims of the present study were to (i) compare allele and haplotype frequencies across selection lines, (ii) study temporal changes in allele frequencies since the past 25 years and (iii) analyze genetic diversity between individual selection lines and evaluate effects of selection on casein frequencies.

2 Materials and methods

2.1 Animals

Milk samples from 2962 cows from first to third lactation of the Holstein Friesian (HF), German Black Pied cattle ("Deutsches Schwarzbuntes Niederungsrind", DSN) and German Simmental (GS) breeds were collected in 2018. The samples were obtained from 50 small and medium-sized farms spread over Germany. Herd sizes ranged from 24 to 228 milking cows, with an average of 76 cows per farm.

The breeds were subdivided into eight selection lines based on divergent breeding strategies (Table 1). With regard to the HF breed, a total of four selection lines was considered. Three HF lines were established in the framework of the "German pasture genetics project" (Brügemann et al., 2015; May et al., 2017) considering a specific mating design in participating grazing herds. The first line in the grazing herds (HF_NZ) based on inseminations of HF_G cows with HF sires from New Zealand. The second line ($\mathrm{HF\_G_p}$) was established considering mating between HF_G cows from the grazing herds with HF_G "pasture" sires. The selected HF_G pasture sires are suited to grazing conditions and represented favorable breeding values for traits that were important in New Zealand (i.e., small body size, high fat percentage, high non-return rate, short interval from calving to first insemination) (May et al., 2017). The third HF line ($\mathrm{HF\_G_m}$) from the grazing herds included female offspring from mating of HF_G cows with HF_G sires representing outstanding breeding values for milk yield. The fourth HF line ($\mathrm{HF\_G_{ref}}$) considered HF_G cows from intensive indoor systems, i.e., herds with a strong selection focus on milk yield. Continuous se-

Table 1. Description of selection lines of the Holstein Friesian (HF), German Black Pied (DSN) and German Simmental (GS) breeds and lactation production records for milk yield, fat percentage and protein percentage for the HF lines.

Selection line	Abbreviation	Description	Production records	Number of cows
German Holstein (milk)	HF_G_m	HF_G–HF_G sires with high breeding values for milk yield in pasture-based systems	9894 kg 3.99 % 3.31 %	64
German Holstein (pasture)	HF_G_p	HF_G–HF_G sires selected for pasture conditions	8702 kg 4.21 % 3.45 %	50
Holstein Friesian (New Zealand)	HF_NZ	German Holstein cow (HF_G)–New Zealand sires	8003 kg 4.40 % 3.60 %	25
German Holstein (reference)	HF_G_{ref}	HF_G kept indoors	10 229 kg 4.04 % 3.36 %	1069
German Black Pied (east)	DSN_{east}	DSN from the new federal states of Germany		1158
German Black Pied (west)	DSN_{west}	DSN from the old federal states of Germany		293
German Simmental (milk)	GS_m	GS; dual-purpose breed in milk production systems		124
German Simmental (beef)	GS_b	GS; beef cattle breed with high value for meat production		179

lection strategies within lines contributed to production trait differences, especially for lactation milk yield and fat percentage as indicated in Table 1.

The local dual-purpose DSN cattle population is the founder breed of the modern HF population and has a long breeding history in the grassland region of East Frisia, Lower Saxony, Germany (Mügge et al., 1999). The DSN breeding goal considers both output traits milk and meat. DSN is defined as robust cattle under harsh environmental conditions and showed superiority over HF in terms of physiological traits (Al-Kanaan, 2016). Due to divergent breeding strategies under different housing conditions after World War II (separation into East and West Germany), two selection lines for DSN were considered (DSN_{east} and DSN_{west}, respectively). For Simmental cattle, the most famous dual-purpose breed for milk and beef production in Germany, two selection lines were included: GS cows of the dual-purpose breed in milk production systems (GS_m) and GS suckler cows as used in beef production systems (GS_b).

2.2 Milk protein typing

Skimmed milk samples from 2962 cows were analyzed for milk protein polymorphisms of α_{s1}-CN, α_{s2}-CN, β-CN and κ-CN by isoelectric focusing in 0.3 mm thin polyacrylamide gels according to Seibert et al. (1985) and Erhardt (1989). This method describes the simultaneous separation of the known α_{s1}-CN, β-CN, α_{s2}-CN and κ-CN variants due to their isoelectric point and considers genetic variants which cannot be detected via commercial SNP chip applications.

2.3 Statistical analyses

Allele frequencies were calculated by direct counting, and HWE was tested by applying a χ^2 test using the packages adegenet version 2.1.1 (Jombart, 2008; Jombart and Ahmed, 2011) and pegas (Paradis, 2010), as implemented in the software package R, version 2.14.2 (R Core Team, 2019). The polymorphic information content (PIC) was computed for each locus within and across populations using the R package polysat (Clark and Jasieniuk, 2011). The observed (H_o) and expected (H_e) heterozygosity were calculated using the R package adegenet. Wright's F-statistic parameters (F_IS, F_IT, F_ST; Wright, 1965) describing the expected level of heterozygosity at various levels of population structure were calculated for each locus across all selection lines using the R package hierfstat (Goudet and Jombart, 2015). The most widely used fixation index (F_ST) serves as a measure of population differentiation due to genetic structure. The overall inbreeding coefficient (F_IT) measures the reduction in heterozygosity of an individual relative to the total population, whereas Wright's inbreeding coefficient (F_IS) measures the reduction in heterozygosity of an individual due to nonrandom mating within its subpopulation (Wright, 1965). The R package hierfstat was also applied for the calculation of F_IS per population and loci. Negative F_IS values indicate heterozygote excesses and positive F_IS values imply a deficiency of heterozygotes, indicating a considerable level of inbreeding. Haplotypes were inferred using the software package PHASE version 2.1 (Stephens et al., 2001), in order to evaluate the haplotype variability within and among populations.

The standard genetic distance (D_s) according to Nei (1972) was calculated from haplotype frequencies using the R package adegenet. The unrooted dendrogram was constructed using the unweighted pair-group method with arithmetic mean (UPGMA) (Sneath and Sokal, 1973) to reconstruct phylogenetic relationships. The robustness of the phylogenies was evaluated by bootstrap values, considering 10 000 replications of resampling loci.

Discriminant analysis of principal components (DAPC) as implemented in the R packages ade4 (Bougeard and Dray, 2018) and adegenet was used to illustrate the admixture within the populations. In contrast to other common multivariate approaches (e.g., principal component analysis or factorial correspondence analysis), DAPC maximizes the separation between groups while minimizing variation within a group, providing a clear discrimination of pre-defined genetic groups (Jombart et al., 2010; Alves et al., 2015).

3 Results

3.1 Allele frequencies and test for Hardy–Weinberg equilibrium

A total of 11 alleles were detected in eight selection lines at four casein loci. The number of alleles per locus ranged from five (β-CN) over three (κ-CN) to two (α_s1-CN) alleles. For α_s2-CN, only the allele A was identified, so that the monomorphic locus α_s2-CN was excluded from further analyses. Allele frequencies of the remaining casein loci α_s1-CN, β-CN and κ-CN in the eight studied selection lines are presented in Table 2. For α_{s1}-CN, all selection lines showed an average high frequency for the common B allele (97 %) and a minor allele frequency (MAF) of 3 % for the C allele. Only the selection lines HF_G$_\mathrm{p}$ (7 %), GS$_\mathrm{b}$ (6 %) and GS$_\mathrm{m}$ (6 %) showed a MAF larger than 3 % for the C allele. For β-CN, the variant A2 (53 %) was the predominant allele, but the two DSN subpopulations had a higher proportion of the A1 allele (in average 67.3 %). The highest frequency of A2 was found in HF_NZ (68 %). The A3 allele revealed highest frequencies in HF_G$_\mathrm{p}$ (7 %) and HF_NZ (2 %) but was zero in both GS subpopulations. The C allele of β-CN only occurred in GS$_\mathrm{b}$ (2 %) and GS$_\mathrm{m}$ (0.4 %) and thus represents a breed-specific allele for GS. With regard to κ-CN, the allele A had the largest frequency in most of the selection lines, but exceptions with a higher or equal frequency of the κ-CN B allele were HF_G$_\mathrm{p}$ and HF_NZ with 51 % and 50 %, respectively (Table 2). The highest frequency of the κ-CN E allele was found in HF_G$_\mathrm{ref}$ (10 %). The rare allele C, which was detected by Erhardt (1993) in GS with a frequency of 0.02 %, was not identified in the sampled animals. In the χ^2 test for HWE, three and five selection lines showed significant deviation ($P < 0.05$) for the β-CN and κ-CN loci, respectively, with corresponding degrees of freedom (d.f.) ranging from 1 to 6 (Table 2). The calculated χ^2 value for α_s1-CN was 0.57 on average (d.f. = 1), indicating HWE in all populations ($P > 0.5$, Table 2).

3.2 Genetic variation of casein loci

The breed- and casein-wise estimates of H_o, H_e and F_IS as well as the PIC are presented in Table 3. Across selection lines, the locus β-CN had the highest H_e (0.53) and the highest PIC (0.42), while the locus α_s1-CN displayed the lowest H_e and PIC values (both 0.03). The κ-CN locus was characterized by intermediate H_e (0.42) and PIC (0.37) values. Mean H_o values for each locus were 0.03 for α_s1-CN, 0.46 for β-CN and 0.41 for κ-CN. Within selection lines, the level of genetic variation was highest for HF_G$_\mathrm{p}$ (average H_e overall loci = 0.41) and lowest for DSN$_\mathrm{east}$ (average H_e overall loci = 0.26). The F_IS values ranged from −0.07 (HF_G$_\mathrm{p}$) to 0.14 (GS$_\mathrm{b}$) at α_s1-CN, from −0.23 (HF_NZ) to 0.27 (DSN$_\mathrm{east}$) at β-CN and from −0.74 (HF_NZ) to 0.14

Table 2. Allele frequencies at the casein loci and tests for Hardy–Weinberg equilibrium with corresponding χ^2 values from the χ^2 test statistics in the studied selection lines. (Significant deviation from Hardy–Weinberg equilibrium at $P<0.05$ is indicated with *; n.s. indicates a non-significant deviation; d.f. indicates the degrees of freedom used in the χ^2 test statistics).

	Selection lines[1]								Erhardt (1993)[2]		
	DSN_{west}	DSN_{east}	HF G_{ref}	HF G_m	HF G_p	HF NZ	GS_b	GS_m	Mean	HF_G	GS
α_{s1}-CN											
B	0.987	0.993	0.987	0.988	0.932	0.978	0.945	0.944	0.969	0.989	0.887
C	0.013	0.007	0.013	0.012	0.068	0.022	0.055	0.056	0.031	0.011	0.113
χ^2	0.03	0.03	0.23	0.00	0.28	0.01	3.5	0.44			
P	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.			
d.f.	1	1	1	1	1	1	1	1			
β-CN											
A1	0.678	0.667	0.320	0.374	0.428	0.300	0.288	0.310	0.422	0.487	0.252
A2	0.291	0.311	0.640	0.624	0.481	0.679	0.611	0.646	0.534		0.657
A3	0.000	0.001	0.004	0.003	0.070	0.021	0.000	0.000	0.014	0.011	0.000
B	0.031	0.021	0.036	0.002	0.021	0.000	0.081	0.040	0.029	0.020	0.091
C	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.004	0.003	0.000	0.010
χ^2	8.16	111.18	15.02	1.60	7.59	2.13	4.52	4.01			
P	*	*	*	n.s.	n.s.	n.s.	n.s.	n.s.			
d.f.	3	6	6	6	6	3	6	6			
κ-CN											
A	0.687	0.822	0.657	0.547	0.450	0.500	0.721	0.768	0.644	0.811	0.760
B	0.222	0.136	0.241	0.421	0.510	0.500	0.258	0.232	0.315	0.134	0.220
E	0.091	0.042	0.102	0.032	0.040	0.000	0.021	0.000	0.041	0.055	0.000
C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020
χ^2	34.44	14.29	7.30	12.48	12.68	9.01	1.11	1.67			
P	*	*	n.s.	*	*	*	n.s.	n.s.			
d.f.	3	3	3	3	3	1	3	1			

[1] Abbreviations for selection lines are given in Table 1. [2] For comparisons of present results with historic data, allele frequencies of the HF_G and GS breeds from the study by Erhardt (1993) were considered.

(DSN_{west}) at κ-CN. The negative F_{IS} values of some breeds indicated an excess of heterozygotes.

The fixation coefficients of subpopulations within the total population, measured as F_{ST} value for the three loci α_{s1}-CN, β-CN and κ-CN, varied from 0.016 (α_{s1}-CN) to 0.080 (β-CN), with a mean of 0.071. It means that 7.1 % of the total genetic variation in the selection lines corresponds to genetic differences among populations, while the remaining 92.9 % explained differences among individuals within population. Additionally, results of F statistics revealed on average an excess of heterozygotes of 11.3 % for each of the analyzed subpopulations (F_{IS}) and 3.4 % in the whole population (F_{IT}). In comparison to the negative F_{IS} values of κ-CN (-0.233) and β-CN (-0.017) among the eight selection lines, the casein locus α_{s1}-CN showed a deficit of heterozygotes due to its positive F_{IS} value (0.001).

3.3 Casein haplotype distributions

Table 4 represents the results of the haplotype analysis of the α_{s1}–β–κ-CN cluster (in order according to their location on BTA6). A total of 13 haplotypes was identified. More than 80 % of all individuals carried one of the BA1A, BA2A or BA2B haplotypes (abbreviation of the specific combination of α_{s1}–β–κ-CN alleles), with mean frequencies of 35 %, 32 % or 20 %, respectively. DSN_{east}, DSN_{west} and HF_G_p revealed the highest frequencies for BA1A, while the most frequent haplotype for the remaining selection lines was BA2A (Table 4).

3.4 Genetic distances and population structure

The matrix of Nei's D_s among the studied selection lines is presented in Table 5. We identified a very close relationship between GS_b and GS_m (0.004). A close relationship was also found between HF_G_m and HF_NZ (0.011), followed by HF_G_m and GS_m (0.013). The selection line HF_G_{ref} was the most divergent from the two DSN subpopulations: 0.318

Table 3. The breed- and casein-wise estimates of observed (H_o) and expected heterozygosity (H_e), Wright's inbreeding coefficient (F_{IS}; Wright, 1965) and polymorphic information content (PIC).

	α_{s1}-CN				β-CN				κ-CN			
Selection lines	H_o	H_e	F_{IS}	PIC	H_o	H_e	F_{IS}	PIC	H_o	H_e	F_{IS}	PIC
DSN_{west}	0.02	0.02	−0.01	0.020	0.53	0.45	−0.17	0.374	0.41	0.47	0.14	0.413
DSN_{east}	0.01	0.01	−0.01	0.010	0.34	0.46	0.27	0.368	0.30	0.30	0.02	0.278
HF_G_m	0.02	0.02	0.00	0.020	0.45	0.48	0.07	0.373	0.76	0.51	−0.44	0.413
HF_G_p	0.14	0.13	−0.07	0.122	0.60	0.58	−0.03	0.490	0.78	0.52	−0.48	0.429
HF_NZ	0.04	0.04	0.00	0.038	0.56	0.45	−0.23	0.364	0.88	0.50	−0.74	0.375
HF_G_{ref}	0.03	0.03	−0.01	0.020	0.53	0.49	−0.09	0.401	0.50	0.50	−0.01	0.437
GS_b	0.10	0.11	0.14	0.106	0.48	0.53	0.10	0.468	0.54	0.41	−0.05	0.291
GS_m	0.11	0.11	−0.06	0.106	0.52	0.49	−0.08	0.398	0.40	0.35	−0.11	0.370
Total	0.03	0.03	0.001	0.03	0.46	0.53	−0.02	0.42	0.41	0.42	−0.23	0.37

Table 4. Haplotype frequencies in the studied selection lines.

Haplotype[1]	DSN_{west}	DSN_{east}	HF_G_{ref}	HF_G_m	HF_NZ	HF_G_p	GS_b	GS_m	All[2]
BA1A	0.50	0.57	0.21	0.34	0.28	0.33	0.27	0.31	0.35
BA1B	0.11	0.07	0.03	0.02	0.02	0.08	0.01	0.004	0.04
BA2A	0.18	0.26	0.44	0.35	0.38	0.19	0.34	0.38	0.32
BA2B	0.09	0.05	0.18	0.25	0.30	0.29	0.20	0.21	0.20
BA1E	0.08	0.04	0.09	0.01	0.00	0.02	0.01	0.00	0.03
BBA	0.01	0.001	0.001	0.00	0.00	0.00	0.06	0.03	0.01
CA2A	0.01	0.001	0.01	0.00	0.00	0.00	0.06	0.06	0.02
BBB	0.01	0.02	0.04	0.01	0.00	0.02	0.01	0.01	0.02
BA2E	0.01	0.00	0.01	0.02	0.00	0.00	0.01	0.00	0.01
CA3A	0.00	0.001	0.004	0.01	0.02	0.07	0.00	0.00	0.01
BCB	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.004	0.003
CA2B	0.00	0.003	0.002	0.00	0.00	0.00	0.003	0.00	0.001
BBE	0.002	0.00	0.00	0.00	0.00	0.00	0.003	0.00	0.001

[1] In the abbreviation of α_{s1}–β–κ-CN. [2] Average haplotype frequencies across all selection lines.

in relation to DSN_{west} and 0.290 in relation to DSN_{east}. The pasture-based selection lines HF_G_m, HF_G_p and HF_NZ showed a close relationship, as documented by low values for D_s ranging from 0.011 to 0.068. The HF_G_{ref} subpopulation distanced itself from the pasture based subpopulations with D_s values up to 0.193. In the UPGMA dendrogram (Fig. 1), two main clusters and two subclusters were identified. In the first main cluster, HF_NZ, HF_G_m, GS_b, GS_m and HF_G_{ref} were placed together, while HF_G_p was allocated to the other subcluster. Both DSN subpopulations (DSN_{east}, DSN_{west}) were placed in the second main cluster. Finally, the DAPC was used to show the genetic admixture between the selection lines. The respective results are shown in Fig. 2. For the best discrimination of haplotypes into predefined clusters, DAPC was run using 12 principal components and seven discriminant functions. The first two linear discriminants, which are illustrated in the scatterplot, contributed to 56 % and 27 % of the total variation, respectively. The first linear discriminant separated DSN and HF populations, whereas the second linear discriminant distinguished between the Simmental subpopulations from all other selection lines.

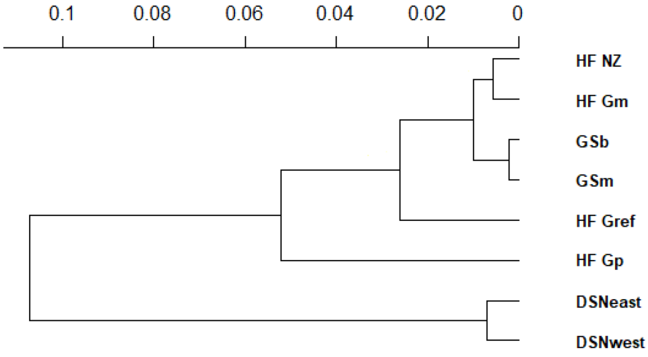

Figure 1. Unrooted dendrogram with bootstrap support using Nei's standard genetic distances (D_S) (Nei, 1972). The x axis represents the genetic distances between the eight studied selection lines.

Table 5. Matrix of Nei's standard genetic distance (D_s) (Nei, 1972) obtained from the haplotype frequencies.

	DSN_{west}	HF_G_m	HF_NZ	HF_G_p	GS_b	GS_m	DSN_{east}	HF_G_{ref}
DSN_{west}	0.000	0.167	0.268	0.155	0.210	0.203	0.014	0.318
HF_G_m		0.000	0.011	0.056	0.019	0.013	0.152	0.061
HF_NZ			0.000	0.068	0.026	0.022	0.253	0.053
HF_G_p				0.000	0.102	0.104	0.188	0.193
GS_b					0.000	0.004	0.190	0.050
GS_m						0.000	0.174	0.044
DSN_{east}							0.000	0.290
HF_G_{ref}								0.000

4 Discussion

4.1 Temporal changes of milk protein polymorphisms

Temporal changes in allele frequencies of milk protein polymorphisms in the HF and GS common cattle breeds are evident when comparing results from the present study with allele frequencies for the same breeds 25 years ago (Erhardt, 1993; Table 2). The 25-year period reflects six generations of mating and selection schemes, with an additional possible impact of random genetic drift. At the β-CN locus, frequencies for the A2 allele were larger in the present than in the historical data, in particular for HF cows. In this regard, Erhardt (1993) detected an A2 allele frequency of 49 % in HF_G, but in the present study, the average allele frequency across the HF populations was 60 %. Chessa et al. (2019) observed a similar trend in temporal changes for A2 in Italian Holstein dairy cattle. Frequencies for A2 were 38.9 %, 49.0 %, 53.1 % and 55.7 % for cows born in 1990, 2000, 2010 and after 2010, respectively. Freyer et al. (1999) and Bech and Kristiansen (1990) reported a favorable impact of the A2 allele on milk and protein yield. Hence, ongoing selection of bulls and cows according to genetic merits for milk or protein yield indirectly increased the A2 variant for β-CN. The relatively high A2 frequencies in the "milk lines" HF_G_{ref} (64 %) and GS_m (65 %) support such hypothesis. Another explanation addresses the relation of A1 milk consumption with the release of the opioid peptide β-casomorphin-7, which may play a role in the development of some human diseases (i.e., ischemic heart disease, type 1 diabetes) (Tailford et al., 2003; Kamiński et al., 2007; Cieślińska et al., 2012; Sheng et al., 2019). As the production of milk with special nutrition properties (i.e., hypoallergenic milk) benefits from the A2 variant of bovine β-CN (De Noni, 2008), farmers are encouraged to select favorable alleles for milk production in niche markets. Additionally, A2 variant information for HF_G sires recently is given in sire catalogues, public journals and discussion forums (Gödert et al., 2017). Nevertheless, the pasture-based genetic line HF_NZ revealed the highest frequency with 68 % for the A2 allele. On the one hand, this may be the result of crossing with the breed Jersey, which generally displays a high frequency of 67 % for β-CN A2 (Erhardt, 1993). In New Zealand, the crosses between HF and Jersey generated the so-called "kiwi cross", a new synthetic breed with favorable values for milk composition traits (Rowarth, 2013; Buckley et al., 2014; Mogollón-García et al., 2020). Another explanation might be the intensive selection for the A2 variant as initiated by the "a2 Milk Company" founded in New Zealand (The a2 Milk Company, 2020). The "a2 Milk Company" initiated a milk marketing program, considering only cows carrying the homozygous β-CN genotype A2A2. Up to now, there have been no progresses regarding active marketing strategies for bovine milk with defined milk protein variants (e.g., A2 milk) in Germany (Gödert et al., 2017).

In contrast to the increasing frequencies of the A2 allele, the β-CN A1 allele declined with progressing time in all selection lines, apart from HF_G_p. The genetic line HF_G_p reflects the A1 and A2 allele frequencies as identified by Erhardt (1993) in HF_G cows. Such a result indicates genomic characteristic similarities of low yielding HF_G cows from low input systems with the broad HF population 25 years ago.

With regard to κ-CN, frequencies of the B allele increased in all populations with progressing time. The B allele frequency in HF_G was 13 % in 1993 (Erhardt, 1993) but increased to 39 % (average from all HF selection lines). The increasing frequency may be due to the positive effect of the κ-CN B allele on milk protein percentage and therefore its favorable cheese-making properties (Hallén et al., 2008; Heck et al., 2009; Mohammadi et al., 2013). Such association stimulated interest in using casein polymorphism in marker-assisted selection schemes to improve milk performance traits in farm animals (Kumar et al., 2006). Simultaneously, the rare allele C was suppressed until its complete loss from the subpopulations. Departures from HWE in both loci (β-CN and κ-CN) reflect temporal changes, i.e., increasing frequencies of the favorable alleles A2 (β-C) and B (κ-CN), due to the impact of selection (Lachance, 2009).

For α_{s1}-CN, frequencies in HF did not differ between present and historic data, because high frequencies of the α_{s1}-CN B allele, close to fixation, were already reported by Erhardt (1993). In GS, the B allele frequency increased from 89 % (Erhardt, 1993) towards fixation (94 %; Table 2).

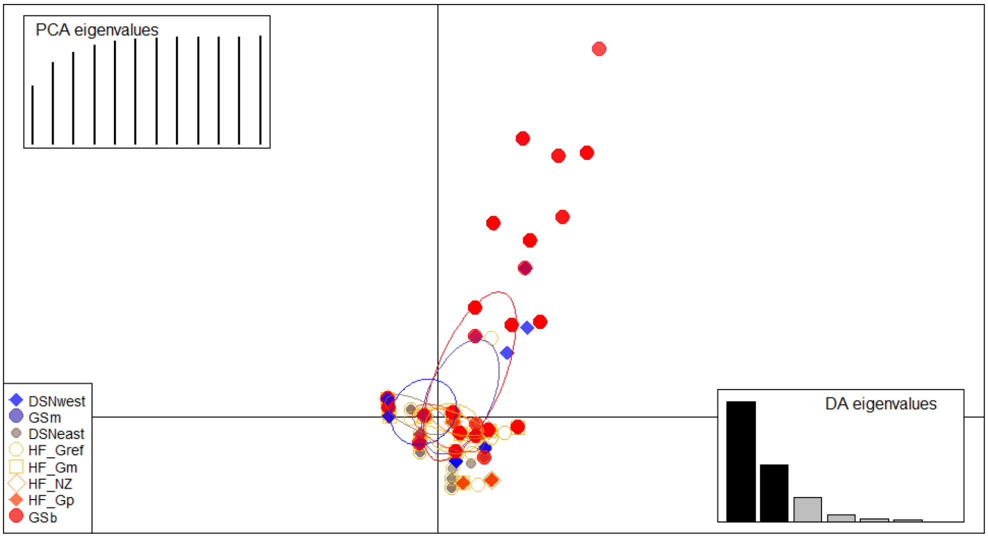

Figure 2. Scatterplot of the discriminant analysis of principal components (DAPC) based on casein haplotype frequencies. Eight selection lines are plotted according to the eigenvectors corresponding to the first (56 %) and second (27 %) linear discriminants. Each circle represents a cluster and each dot represents an individual.

With regard to the α_{s2}-CN locus, only the allele A was identified in all selection lines. The D allele is rather common in French breeds (e.g., Montbéliarde) (Grosclaude et al., 1979) but was also described for HF and GS with low frequencies of 0.2 % and 2 %, respectively (Erhardt, 1993; Meier et al., 2019). Such a loss of rare alleles (e.g., α_{s2}-CN D allele, κ-CN C allele) indicates genetic drift, a mechanism of evolution in which allele frequencies change over generations by chance (Hartl and Clark, 2007), with an impact on decreasing genetic diversity.

4.2 Genetic diversity parameters

Among selection lines, HF_G_p displayed the highest gene diversity over all loci (average H_e overall loci = 0.41). Alternative selection of HF in grazing systems with a focus on a broad pattern of functional traits including especially female fertility and somatic cells might explain their variability at protein loci. Additionally, observed (but rather limited) genetic exchange with DSN contributed to genetic diversity. The H_e for each locus in the reference line HF_G_{ref} is in agreement with the commercial Portuguese HF population (Beja-Pereira et al., 2002). The lowest values for H_e across all loci were observed for DSN_{east}, which might be due to the larger inbreeding increase in the DSN east subpopulation compared to the subset for DSN cows from former West Germany (Jaeger et al., 2018a). As the deficiency of heterozygotes is an indication of inbreeding, the positive F_{IS} values for DSN_{east} at both loci (β-CN and κ-CN) (0.27 and 0.02, respectively) underline this assumption. An explanation for the mating of closely related animals in the past DSN_{east} is the restricted gene flow from foreign countries in the former German Democratic Republic. In contrast, in DSN_{west}, sires from the Netherlands have been used in the period from 1970 to 1980 (Jaeger et al., 2018a). Nevertheless, also for DSN_{west}, the diversity measurement ($H_e = 0.31$) suggests a general small effective population size for DSN, reflecting a small real population with only 2800 registered cows in Germany (Rinderproduktion Berlin-Brandenburg GmbH, 2016). A decreasing population size is a major cause for losses in genetic diversity (Kantanen et al., 1999). In such context, Jaeger et al. (2018b) calculated an increase of inbreeding per year in DSN of 0.1 %, implying a rather small effective population size of 85 animals.

4.3 Relationships between selection lines

In the UPGMA dendrogram, the genetically closely related subpopulations of the DSN breed (DSN_{east}, DSN_{west}; $D_S = 0.014$) built their own cluster, clearly differentiated from the remaining selection lines. The majority of selection lines (i.e., selection lines within HF and GS) revealed the highest frequency for the haplotype BA2A, which was also detected for Italian Friesian cattle (Boettcher et al., 2004). In contrast, in both DSN subpopulations, BA1A was the most frequent casein haplotype with a frequency up to 57 %. The predominant impact of such chromosomal segment and especially of the A1 allele (67.3 % on average for both DSN subpopulations) is in agreement with results of Ng-Kwai-Hang et al. (1984) and Meier et al. (2019). Generally, in addition to DSN, breeds originating from northern Europe, including European Red cattle (Bech and Kristiansen, 1990), Black and White Lowland breeds (McLean et al., 1984) and Danish Red cattle (Meier et al., 2019), showed highest frequencies for the A1 allele and the corresponding haplotypes. These results indicate that the β-CN A1 allele is a major characteristic for breeds with a certain geographic location in Nordic countries. A further explanation for genomic similarities in Nordic breeds including DSN addresses identical breeding objectives towards a dual-purpose phenotype. A shared characteristic of Nordic breeds and DSN is the similarity in fat and protein percentages (Meier et al., 2019).

The average genetic distance of 0.30 between the founder DSN breed ($\mathrm{DSN_{east}}$, $\mathrm{DSN_{west}}$) and the modern $\mathrm{HF\_G_{ref}}$ population indicates the breeding particularities in both lines in the past decades. The robust DSN cattle were subject of extensive breeding mostly in pasture-based production systems, predominantly considering mating with natural service sires. In contrast, $\mathrm{HF\_G_{ref}}$ cows were intensively selected for milk yield, indirectly favoring casein haplotypes as already reported in goats (Grosclaude and Martin, 1997). The results for Nei's D_S indicate a close relationship between both DSN subpopulations and $\mathrm{HF\_G_p}$. The predominance of the A1 allele in $\mathrm{HF\_G_p}$ makes them more similar to $\mathrm{DSN_{east}}$ and $\mathrm{DSN_{west}}$ than to the current $\mathrm{HF\_G_{ref}}$ population. This might be due to the higher genetic percentage of DSN in their ancestors, as the genetic pasture line was selected for robust animals (Jaeger et al., 2018a, b). The high frequencies for the α_{s1}-CN C allele in $\mathrm{HF\_G_p}$ (7 %) and both DSN subpopulations (6 %) support such hypothesis.

Genetic distances were observed between HF selection lines, as they clustered separately. With regard to allele frequencies at the κ-CN locus, some specific patterns in selection lines were noticed. First, the selection lines $\mathrm{HF\_G_p}$, HF_NZ and $\mathrm{HF\_G_m}$ from the grazing herds displayed the highest allele frequencies for the favorable κ-CN B allele (51 %, 50 % and 42 %, respectively). The importance of specific breeding goal traits differed in divergent feeding systems (Washburn and Mullen, 2014; Delaby et al., 2018). In pasture-based systems, the focus of selection has emphasized fertility, fitness and robustness. The prevalence of the B allele in pasture-based selection lines may be the result of indirect selection for these traits. In this regard, Hiendleder et al. (2003) detected quantitative trait locus (QTL) linked with the milk protein genes on BTA6 for udder quality and limb conformation (e.g., quality of feed and leg), being traits reflecting the pasture ability.

In both GS subpopulations ($\mathrm{GS_b}$, $\mathrm{GS_m}$), we identified a private allele in the β-CN C variant and therefore the breed-specific haplotype BCB. This is in agreement with results by Çardak (2005), who found the C allele with a frequency of 2.3 % in Simmental cows but not in HF. The occurrence of β-CN C explains the lowest D_A between $\mathrm{GS_b}$ and $\mathrm{GS_s}$ as well as their differentiation from the other selection lines (Fig. 1). The breed-specific C allele may be linked to a favorable mutation on BTA6 for carcass and body weight, promoting the breeding value for beef production in a dual-purpose breed like GS. In this regard, QTL for growth traits (i.e., body length, carcass weight) have been detected within the *NCAPG* gene located on BTA6 in local beef cattle breeds (e.g., Chinese Qingchuan and Japanese Black and Brown beef cattle), indicating overlapping mechanisms of bone and muscle growth with lipid deposition (Setoguchi et al., 2009; Liu et al., 2015). Furthermore, the gene *SPP1* on BTA6 was associated with body weight in Polish Holstein Friesian cattle (Pareek et al., 2008). In a functional genomic approach, Sheehy et al. (2009) suggested *SPP1* as an important regulator of bovine milk protein gene expressions, explaining the possible link between the casein and *SPP1*.

4.4 Genetic differentiation among selection lines

In the present study, the average F_ST among selection lines was 7.1 %, reflecting a moderate level of population differentiation (Hartl and Clark, 2007). Hence, 7 % of the total genetic variation corresponds to selection line particularities, and the remaining 93 % is due to individual differences. The illustration for DAPC (Fig. 2) indicates that the selection lines do not clearly distinguish divergent clusters. The DAPC visualizes a high admixture between the subpopulations. We only identified a separation between the DSN and HF subpopulations along the linear discriminant 1 and a separation between the Simmental subpopulations with the remaining selection lines. The slight genetic variation among subpopulations might be a result of the decreased variability at the casein loci, which is indicated by the average PIC of 0.27 over all loci. Genotyping of the casein genes (e.g., β-CN and κ-CN) is of increasing relevance for practical breeding and selection, also from a genetic diversity monitoring perspective.

5 Conclusions

The results of the present study indicate that different selection strategies (e.g., pasture ability, meat or dairy production) indirectly contributed to the variability of the casein polymorphisms linked to milk production traits. The selection lines of the endangered DSN breed showed the lowest gene diversity and clearly separated from the HF and GS breeds due to their predominance of the β-CN A1 allele. The pasture-based selection lines of the HF breed carried the favorable κ-CN B allele with highest frequency, which is related to a higher protein content in milk. Temporal changes in allele distributions reflect that casein loci or selected mutations in close proximity to the casein underlie selective breeding. Fixation of alleles and results for evaluated indicators of heterozygosity (e.g., H_e, F_ST, F_IS) showed diversity loss at the casein loci. The present study revealed differences in allele frequencies at casein loci across selection lines, indicating breeding potential for specific milk markets. Furthermore, genetic milk protein variants can be used to monitor genetic diversity.

Data availability. The data that support the findings of this study are available from the authors upon reasonable request.

Author contributions. All authors made substantial contributions to the study. SK, GE and CW designed the experiment and supervised the research. SK supported LGH in writing and data validation. LGH performed the laboratory analysis, supported by CW. LGH and CS were responsible for data preparation and statistical analyses. All authors read and approved the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors are grateful to the Justus Liebig University of Gießen for providing the JLU graduate scholarship to Lisa G. Hohmann.

Financial support. This research has been supported by the German Federal Ministry of Food and Agriculture (BMEL), granted by the Federal Office for Agriculture and Food (BLE; grant no. FKZ 2818BM090).

This open-access publication was funded
by Justus Liebig University Giessen.

Review statement. This paper was edited by Antke-Elsabe Freifrau von Tiele-Winckler and reviewed by Stanislaw Kaminski and one anonymous referee.

References

Al-Kanaan, A.: Heat stress response for physiological traits in dairy and dual purpose cattle populations on phenotypic and genetic scales, PhD thesis, Faculty of Organic Agriculture, University of Kassel, Kassel, Germany, 142 pp., 2016.

Alves, J. M., Carneiro, M., Afonso, S., Lopes, S., Garreau, H., Boucher, S., Allain, D., Queney, G., Esteves, P. J., Bolet, G., and Ferrand, N.: Levels and patterns of genetic diversity and population structure in domestic rabbits, Plos One, 10, e0144687, https://doi.org/10.1371/journal.pone.0144687, 2015.

Bech, A. M. and Kristiansen, K. R.: Milk protein polymorphism in Danish dairy cattle and the influence of genetic variants on milk yield, J. Dairy Res., 57, 53–62, https://doi.org/10.1017/S0022029900026601, 1990.

Beja-Pereira, A., Erhardt, G., Matos, C., Gama, L., and Ferrand, N.: Evidence of a geographical cline of casein haplotypes in Portuguese cattle breeds, Anim. Genet., 33, 295–300, https://doi.org/10.1046/j.1365-2052.2002.00871.x, 2002.

Boettcher, P. J., Caroli, A., Stella, A., Chessa, S., Budelli, E., Canavesi, F., Ghiroldi, S., and Pagnacco, G.: Effects of Casein Haplotypes on Milk Production Traits in Italian Holstein and Brown Swiss Cattle, J. Dairy Sci., 87, 4311–4317, https://doi.org/10.3168/jds.S0022-0302(04)73576-6, 2004.

Bougeard, S. and Dray, S.: Supervised Multiblock Analysis in R with the ade4 Package, J. Stat. Softw., 86, 1–17, https://doi.org/10.18637/jss.v086.i01, 2018.

Brügemann, K., Jaeger, M., Ruebesam, K., König von Borstel, U., and König, S.: Differences in novel traits between genetic groups of dairy cows in pasture-based production systems, in: Proceedings of the 66th Annual Meeting of the European Federation of Animal Science, Warsaw, Poland, 31 August–4 September 2015, 576 pp., 2015.

Buckley, F., López-Villalobos, N., and Heins, B. J.: Crossbreeding: Implications for dairy cow fertility and survival, Animal, 8, 122–133, https://doi.org/10.1017/S1751731114000901, 2014.

Campbell, J. R. and Marshall, R. T.: Dairy Production and Processing: The Science of Milk and Milk Products, Waveland Press, Long Grove, USA, 2016.

Çardak, A. D.: Effects of genetic variants in milk protein on yield and composition of milk from Holstein-Friesian and Simmentaler cows, S. Afr. J. Anim. Sci., 35, 41–47, https://doi.org/10520/EJC94433, 2005.

Çardak, A. D., Bartenschlager, H., and Geldermann, H.: Effects of polymorph milk proteins on the individual milk protein content of Holstein-Friesian and Simmental cows, Milchwissenschaft, 58, 235–238, 2003.

Caroli, A. M., Chessa, S., and Erhardt, G. J.: Invited review: Milk protein polymorphisms in cattle: Effect of animal breeding and human nutrition, J. Dairy Sci., 92, 5335–5352, https://doi.org/10.3168/jds.2009-2461, 2009.

Chessa, S., Gattolin, S., Finocchiaro, R., Van Kaam, J.-T., Marusi, M., and Civati, G.: Beta-casein A2 variant: is the frequency changing in Holstein cattle?, Ital. J. Anim. Sci., 18, 125–126, 2019.

Cieślińska, A., Kostyra, E., Kostyra, H., Olenski, K., Fiedorowicz, E., and Kaminski, S.: Milk from cows of different beta-casein genotypes as a source of beta-casomorphin-7, Int. J. Food Sci. Nutr., 63, 426–430, https://doi.org/10.3109/09637486.2011.634785, 2012.

Clark, L. and Jasieniuk, M.: polysat: an R package for polyploid microsatellite analysis, Mol. Ecol. Resour., 11, 562–566, https://doi.org/10.1111/j.1755-0998.2011.02985.x, 2011.

Delaby, L., Buckley, F., McHugh, N., and Blanc, F.: Robust animals for grass based production systems, in: Proceedings of the 23/27th European Grassland Federation General Meeting on “Sustainable Meat and Milk Production from Grasslands”, Cork, Ireland, 17–21 June 2018, 389–400, 2018.

De Noni, I.: Release of β-casomorphins 5 and 7 during simulated gastro-intestinal digestion of bovine β-casein variants and milk-based infant formulas, Food Chem., 110, 897–903, https://doi.org/10.1016/j.foodchem.2008.02.077, 2008.

Ehrmann, S., Bartenschlager, H., and Geldermann, H.: Quantification of gene effects on single milk proteins in selected groups of dairy cows, J. Anim. Breed. Genet., 114, 121–132, https://doi.org/10.1111/j.1439-0388.1997.tb00499.x, 1997.

Erhardt, G.: κ-Kaseine in der Rindermilch – Nachweis eines weiteren Allels (κ-Cn^E) in verschiedenen Rassen, J. Anim. Breed. Genet., 106, 225–231, https://doi.org/10.1111/j.1439-0388.1989.tb00233.x, 1989.

Erhardt, G.: Allele frequencies of milk proteins in German cattle breeds and demonstration of α_{s2}-casein variants by isoelectric focusing, Arch. Anim. Breed., 36, 145–152, 1993.

Ferretti, L., Leone, P., and Sgaramella, V.: Long range restriction analysis of the bovine casein genes, Nucleic Acids Res., 18, 6829–6833, https://doi.org/10.1093/nar/18.23.6829, 1990.

Freyer, G., Liu, Z., Erhardt, G., and Panicke, L.: Casein polymorphism and relation between milk production traits, J. Anim. Breed. Genet., 116, 87–97, https://doi.org/10.1046/j.1439-0388.1999.00181.x, 1999.

Gallinat, J. L., Qanbari, S., Drögemüller, C., Pimentel, E. C. G., Thaller, G., and Tetens, J.: DNA-based identification of novel bovine casein gene variants, J. Dairy Sci., 96, 699–709, https://doi.org/10.3168/jds.2012-5908, 2013.

Gödert, M., Brandt, H., and Erhardt, G.: Beta-Casein A2 in Rindermilch – Hintergründe, züchterische und milchwirtschaftliche Strategien und Begrenzungen im Hinblick auf eine mögliche neue Nachfragesituation, Züchtungskunde, 89, 451–474, 2017.

Goudet, J. and Jombart, T.: Hierfstat: Estimation and tests of hierarchical F-statistics, R package version 0.04-22, available at: https://CRAN.R-project.org/package=hierfstat (last access: 10 September 2020), 2015.

Grosclaude, F. and Martin, P.: Casein polymorphisms in the goat, Milk Protein Polymorphism, 9702, 241–253, 1997.

Grosclaude, F., Joudrier, P., and Mahe, M. F.: A genetic and biochemical analysis of a polymorphism of bovine α_{S2}-casein, J. Dairy Res., 46, 211–213, https://doi.org/10.1017/S0022029900017052, 1979.

Hallén, E., Wedholm, A., Andrén, A., and Lundén, A.: Effect of β-casein, κ-casein and β-lactoglobulin genotypes on concentration of milk protein variants, J. Anim. Breed. Genet., 125, 119–129, https://doi.org/10.1111/j.1439-0388.2007.00706.x, 2008.

Hartl, D. L. and Clark, A. G.: Principles of Population Genetics, 4th ed, Sinauer Associates Inc., Sunderland, USA, 2007.

Heck, J. M. L., Schennink, A., van Valenberg, H. J. F., Bovenhuis, H., Visker, M. H. P. W., van Arendonk, J. A. M., and van Hooijdonk, A. C. M.: Effects of milk protein variants on the protein composition of bovine milk, J. Dairy Sci., 92, 1192–1202, https://doi.org/10.3168/jds.2008-1208, 2009.

Hiendleder, S., Thomsen, H., Reinsch, N., Bennewitz, J., Leyhe-Horn, B., Looft, C., Xu, N., Medjugorac, I., Russ, I., Kühn, C., Brockmann, G. A., Blümel, J., Brenig, B., Reinhardt, F., Reents, R., Averdunk, G., Schwerin, M., Weimann, C., Förster, M., Kalm, E., and Erhardt, G.: Mapping of QTL for body conformation and behavior in cattle, J. Hered., 94, 496–506, https://doi.org/10.1093/jhered/esg090, 2003.

Hohenlohe, P., Phillips, P., and Cresko, W.: Using population genomics to detect selection in natural populations: key concepts and methodological considerations, Int. J. Plant. Sci., 171, 1059–1071, https://doi.org/10.1086/656306, 2010.

Jaeger, M.: Phenotypic, Quantitative Genetic and Genomic Characterization of the German Black and White dual-purpose Cattle Breed, PhD thesis, Faculty of Agricultural Sciences, Nutritional Sciences, and Environmental Management, University of Giessen, Giessen, Germany, 203 pp., 2018.

Jaeger, M., Scheper, C., König, S., and Brügemann, K.: Studien zur Inzucht und Verwandtschaft beim ‘Deutschen Schwarzbunten Niederungsrind’ (DSN) auf Basis eigens berechneter Rasseanteile, Züchtungskunde, 90, 262–279, 2018a.

Jaeger, M., Brügemann, K., Naderi, S., Brandt, H., and König, S.: Variance heterogeneity and genotype by environment interactions in native Black and White dual-purpose cattle for different herd allocation schemes, Animal, 13, 2146–2155, https://doi.org/10.1017/S1751731119000144, 2018b.

Jann, O. C., Ibeagha-Awemu, E. M., Özbeyaz, C., Zaragoza, P., Williams, J. L., Ajmone-Marsan, P., Lenstra, J. A., Moazami-Goudarzi, K., and Erhardt, G.: Geographic distribution of haplotype diversity at the bovine casein locus, Genet. Sel. Evol., 36, 243–257, https://doi.org/10.1186/1297-9686-36-2-243, 2004.

Jombart, T.: adegenet: a R package for the multivariate analysis of genetic markers, Bioinformatics, 24, 1403–1405, doi.org/10.1093/bioinformatics/btn129, 2008.

Jombart, T. and Ahmed, I.: adegenet 1.3-1: new tools for the analysis of genome-wide SNP data, Bioinformatics, 27, 3070–3071, https://doi.org/10.1093/bioinformatics/btr521, 2011.

Jombart, T., Devillard, S., and Balloux, F.: Discriminant analysis of principal components: a new method for the analysis of genetically structured populations, BMC Genet., 11, 94, https://doi.org/10.1186/1471-2156-11-94, 2010.

Kamiński, S., Cieślińska, A., and Kostyra, E.: Polymorphism of bovine beta-casein and its potential effect on human health, J. Appl. Genet., 48, 189–198, https://doi.org/10.1007/BF03195213, 2007.

Kantanen, J., Olsaker, I., Adalsteinsson, S., Sandberg, K., Eythorsdottir, E., Pirhonen, K., and Holm, L.-E.: Temporal changes in genetic variation of North European cattle breeds, Anim. Genet., 30, 16–27, https://doi.org/10.1046/j.1365-2052.1999.00379.x, 1999.

König, S., Dietl, G., Raeder, I., and Swalve, H. H.: Genetic relationships for dairy performance between large-scale and small-scale farm conditions, J. Dairy Sci., 88, 4087–4096, https://doi.org/10.3168/jds.S0022-0302(05)73093-9, 2005.

König, S., Lessner, S., and Simianer, H.: Application of controlling instruments for improvements in cow sire selection, J. Dairy Sci., 90, 1967–1980, https://doi.org/10.3168/jds.2006-473, 2007.

Kreitman, M.: Methods to Detect Selection in Populations with Applications to the Human, Annu. Rev. Genom. Hum. G., 1, 539–559, https://doi.org/10.1146/annurev.genom.1.1.539, 2000.

Kumar, D., Gupta, N., Ahlawat, S., Satyanarayana, R., Sunder, S., and Gupta, S.: Single strand confirmation polymorphism (SSCP) detection in exon I of the lactalbumin gene of Indian Jamunapri milk goats (*Capra hircus*), Genetic Mol. Biol., 29, 271–274, https://doi.org/10.1590/S1415-47572006000200016, 2006.

Lachance, J.: Detecting selection-induced departures from Hardy-Weinberg proportions, Genet. Sel. Evol., 41, 15, https://doi.org/10.1186/1297-9686-41-15, 2009.

Lien, S., Kamiński, S., Aleström, P., and Rogne, S.: A Simple and Powerful Method for Linkage Analysis by Amplification of DNA from Single Sperm Cells, Genomics, 16, 41–44, https://doi.org/10.1006/geno.1993.1137, 1993.

Liu, Y., Duan, X., Chen, S., He, H., and Liu, X.: *NCAPG* is differentially expressed during longissimus muscle development and is associated with growth traits in Chinese Qinchuan beef cattle, Genet. Mol. Biol., 38, 450–456, https://doi.org/10.1590/S1415-475738420140287, 2015.

Loftus, R. T, MacHugh, D. E., Bradley, D. G., Sharp, P. M., and Cunningham, P.: Evidence for two independent domestications of cattle, P. Natl. Acad. Sci. USA, 91, 2757–2761, https://doi.org/10.1073/pnas.91.7.2757, 1994.

Lopez-Villalobos, N., Garrick, D. J., Holmes, C. W., Blair, H. T., and Spelman, R. J.: Effects of selection and crossbreeding strategies on industry profit in the New Zealand Dairy industry, J. Dairy Sci., 83, 164–172, https://doi.org/10.3168/jds.S0022-0302(00)74867-3, 2000.

Mahé, M.-F., Miranda, G., Queval, R., Bado, A., Zafindrajaona, P. S., and Grosclaude, F.: Genetic polymorphism of milk proteins in African Bos taurus and Bos indicus populations. Characterization of variants α_{s1}-Cn^H and β-Cn^J, Genet. Sel. Evol., 31, 239–253, doi.org/10.1186/1297-9686-31-3-239, 1999.

May, K., Brügemann, K., Yin, T., Scheper, C., Strube, C., and König, S.: Genetic line comparisons and genetic parameters for endoparasite infections and test-day milk production traits, J. Dairy Sci., 100, 7330–7344, https://doi.org/10.3168/jds.2017-12901, 2017.

Maynard Smith, J. and Haigh, J.: The hitch-hiking effect of a favourable gene, Genet. Res., 23, 23–35, https://doi.org/10.1017/S0016672300014634, 1974.

McLean, D. M., Graham, E. R. B., Ponzoni, R. W., and McKenzie, H. A.: Effects of milk protein genetic variants on milk yield and composition, J. Dairy Res., 51, 531, https://doi.org/10.1017/S0022029900032854, 1984.

Meier, S., Korkuć, P., Arends, D., and Brockmann, G. A.: DNA Sequence Variants and Protein Haplotypes of Casein Genes in German Black Pied Cattle (DSN), Front. Genet., 10, 1129, https://doi.org/10.3389/fgene.2019.01129, 2019.

Mohammadi, Y., Aslaminejad, A. A., Nassiry, M. R., and Koshkoieh, A. E.: Allelic polymorphism of κ-casein, β-Lactoglobulin and leptin genes and their association with milk production traits in Iranian Holstein cattle, JCMR, 5, 75–80, https://doi.org/10.22067/JCMR.V5I2.21086, 2013.

Mogollón-García, H. D., Nieto-Sierra, D. F., and Castro-Rincón, E.: Productive performance of Holstein and the crossbreeding Kiwi Cross x Holstein cattle, Artículo científico, 31, 329–339, https://doi.org/10.15517/am.v31i2.38541, 2020.

Mügge, B., Lutz, W. E., Südbeck, H., and Zelfel, S.: Deutsche Holsteins: Die Geschichte einer Zucht, Eugen Ulmer, Stuttgart, Germany, 1999.

Nei, M.: Genetic Distance between Populations, Am. Nat., 106, 283–292, https://doi.org/10.1086/282771, 1972.

Ng-Kwai-Hang, K. F., Hayes, J. F., Moxley, J. E., and Monardes, H. G.: Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle, J. Dairy Sci., 67, 835–840, https://doi.org/10.3168/jds.S0022-0302(84)81374-0, 1984.

Paradis, E.: pegas: an R package for population genetics with an integrated-modular approach, Bioinformatics, 26, 419–420, https://doi.org/10.1093/bioinformatics/btp696, 2010.

Pareek, C. S., Czarnik, U., a Pierzchal, M., and Zwierzchowski, L.: An association between the C>T single nucleotide polymorphism within intron IV of osteopontin encoding gene (*SPP1*) and bodyweight of growing Polish Holstein-Friesian cattle, Anim. Sci. Pap. Rep., 26, 251–257, 2008.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: https://www.R-project.org/ (last access: 10 September 2020), 2019.

Rinderproduktion Berlin-Brandenburg GmbH: Deutsches Schwarzbuntes Niederungsrind – lebendes Kulturerbe, available at: https://www.rinderzucht-bb.de/zucht/dsn-genreserve/ (last access: 14 August 2020), 2016.

Rowarth, J. S.: Dairy cows: economic production and environmental protection, in: Ecosystem Services in New Zealand: Conditions and Trends, edited by: Dymond, J. R., Manaaki Whenua Press, Lincoln, New Zealand, 85–93, 2013.

Seibert, B., Erhardt, G., and Senft, B.: Procedure for simultaneous phenotyping of genetic variants in cow's milk by isoelectric focusing, Anim. Blood Groups Biochem. Genet., 16, 183–191, https://doi.org/10.1111/j.1365-2052.1985.tb01469.x, 1985.

Setoguchi, K., Furuta, M., Hirano, T., Nagao, T., Watanabe, T., Sugimoto, Y., and Takasuga, A.: Cross-breed comparisons identied a critical 591-kb region for bovine carcass weight QTL (*CW-2*) on chromosome 6 and the Ile-442-Met substitution in *NCAPG* as a positional candidate, BMC Genet., 4, 10–43, https://doi.org/10.1186/1471-2156-10-43, 2009.

Sheehy, P. A., Riley, L. G., Raadsma, H. W., Williamson, P., and Wynn, P. C.: A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6, Anim. Genet., 40, 492–498, https://doi.org/10.1111/j.1365-2052.2009.01862.x, 2009.

Sheng, X., Li, Z., Ni, J., and Yelland, G.: Effects of Conventional Milk Versus Milk Containing Only A2 β-Casein on Digestion in Chinese Children: A Randomized Study, J. Pediatr. Gastroenterol. Nutr., 69, 375–382, https://doi.org/10.1097/MPG.0000000000002437, 2019.

Sneath, P. H. A. and Sokal, R. R.: Numerical taxonomy. The principles and practice of numerical classification, W. H. Freeman and Company, San Francisco, USA, 1973.

Stephens, M., Smith, N., and Donnelly, P.: A new statistical method for haplotype reconstruction from population data, Am. J. Hum. Genet., 68, 978–989, https://doi.org/10.1086/319501, 2001.

Tailford, K. A., Berry, C. L., Thomas, A. C., and Campbell, J. H.: A casein variant in cow's milk is atherogenic, Atherosclerosis, 170, 13–19, https://doi.org/10.1016/S0021-9150(03)00131-X, 2003.

The a2 Milk Company: Fonterra Customer Services, Auckland, New Zealand, available at: https://a2milk.nz/, last access: 12 August 2020.

Washburn, S. P. and Mullen, K. A. E.: Invited review: Genetic considerations for various pasture-based dairy systems, J. Dairy S., 97, 5923–5938, https://doi.org/10.3168/jds.2014-7925, 2014.

Wright, S.: The interpretation of population structure by F-statistics with special regard to systems of mating, Evolution, 19, 395–420, https://doi.org/10.2307/2406450, 1965.