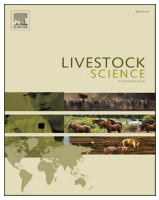
Livestock Science 273 (2023) 105261

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci



Genomic diversity and population structure of the Czech Holstein cattle

Lubos Vostry [a,*], Hana Vostra-Vydrova [a], Nina Moravcikova [b], Radovan Kasarda [b], Vlatka Cubric-Curik [c], Michaela Brzakova [d], Johan Solkner [e], Mario Shihabi [c], Jorge Angel Hidalgo Moreno [f], Maria Spehar [g], Ino Curik [c,*]

[a] Czech University of Life Science Prague, Kamycka 129, 16500 Prague, Czech Republic
[b] Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovak Republic
[c] University of Zagreb, Faculty of Agriculture, Svetošimunska cesta 25, 10000 Zagreb, Croatia
[d] Institute of Animal Science, Přátelství 815, 104 00 Prague, Czech Republic
[e] University of Natural Resources & Life Sciences Vienna, Gregor-Mendel-Strasse 33, 1180 Vienna, Austria
[f] Department of Animal and Dairy Science, University of Georgia, 425 River Road, Athens, GA, USA
[g] Centre for Livestock Breeding, Svetošimunska cesta 25, 10000 Zagreb

HIGHLIGHTS

- The genomic diversity of the Czech and other Holstein subpopulations was investigated.
- Inbreeding in the AI bulls was very high and far exceeded the inbreeding in the cows.
- Using of highly inbred AI bulls will lead to a high level of inbreeding in the Czech subpopulation.
- The differences in the Holstein subpopulations are consequence of different breeding programs.

ARTICLE INFO

Keywords:
Effective populations size
Genomic relatedness
Runs of homozygosity
SNPs
Holstein Friesian breed

ABSTRACT

Holstein-Friesian (HF) is a cosmopolitan breed distributed in more than 150 countries. It represents a large metapopulation with intensive gene flow, realised mainly through artificial insemination and the intensive use of the popular "star" bulls. The breed is known for its superiority in milk yield, production efficiency and black and white coat color. In contrast to the enormous size of the census population, which exceeds nine million animals in the U.S. alone, the genetic diversity of this highly commercialised breed is surprisingly low, necessitating genetic monitoring, especially of subpopulations in individual countries. Our main objective in this study was to analyze genomic diversity (estimated by genomic inbreeding and effective population size) and population structure (relationship to other subpopulations) of the subpopulation from the Czech Republic and, based on high-throughput SNP array genotypes. We analysed 2178 animal samples (32,865 autosomal SNP) from 12 sub-populations and the Simental cattle breed (98 animals), which represents an outlier population. Czech bulls showed high genomic inbreeding ($F_{ROH>2Mb}$=0.133), well above the inbreeding level of Czech cows ($F_{ROH>2Mb}$=0.091), with particularly high recent inbreeding (ROH>8Mb). Unexpectedly, the estimated effective population size (Ne_{LD}) was relatively high, ranging from 202 (GONE) to 283 (NeEstimator v2), depending on the estimation algorithm. Our phylogenetic analyses showed that the Czech HF belonged to the "core metapopulation HF ", together with Belgian, British, Canadian, Dutch, French, German, and USA subpopulations, which was separated from the Swiss, Irish, and Croatian subpopulations. We also showed that Czech AI bulls differed slightly from cows, especially in genes affecting meat and carcass. Our results have defined the population structure of the Czech HF and indicate the potential problems of increased inbreeding due to selection of AI bulls.

* Corresponding author.
E-mail addresses: vostry@af.czu.cz (L. Vostry), icurik@agr.hr (I. Curik).

https://doi.org/10.1016/j.livsci.2023.105261
Received 10 December 2022; Received in revised form 13 April 2023; Accepted 24 April 2023
Available online 21 May 2023

1. Introduction

Management of genetic diversity in intensively selected cattle populations is an essential component to ensure successful long-term increase in selection response to production and adaptation traits that provide optimal genetic improvement (Boichard et al., 2016; Goddard, 2009; Hill, 2000). Maintaining sufficient genetic variability is also necessary to meet production requirements in different environments, to enable sustained genetic improvement and to adapt to changing breeding objectives (de Jong and Bijma, 2002; Strandén et al., 2019). Genetic diversity also represents a key factor to avoid negative consequences due to inbreeding (Doekes et al., 2018; Ferenčaković et al., 2017; Howard et al., 2017) and/or increased prevalence of genetic defects (Cole, 2015; VanRaden et al., 2011). However, the genetic diversity of dairy breeds has been greatly reduced, jeopardizing successful long-term genetic improvement. For example, despite large numbers of animals (large census), the genetic diversity of commercial breeds is often relatively low because only a few sires are used to reproduce with large numbers of offspring (Hodges, 2006).

Holstein-Friesian (HF) is a trans-border breed spread in more than 150 countries and represents a large metapopulation. It is known for its superiority in milk yield, production efficiency and black and white piebald coat color, and is now the most widespread dairy breed in the world. In the Czech Republic, the Holstein subpopulation (CZE) has long been bred for high milk yield. (www.holstein.cz). Similar to other national subpopulations, the CZE has been developed over the past fifty years from domestic breeding material and imported semen and embryos from international bulls and cows with high milk yields. More specifically, more than 1 million AI semen and several thousand embryos have been imported, along with 200 live bulls and 25,000 heifers imported mainly from France, Germany, the Netherlands, and Denmark.

The genomic diversity of subpopulations of HF is relatively low worldwide, and for example, the effective population size (N_e) for Holstein populations in Australia, Canada, Denmark, Spain, Ireland, and the United States of America ranges from 49 to 127 (Doekes et al., 2018). The main reason for reduced genomic diversity is the overuse of a small number of superior bulls. This is most evident in the variability of the Y chromosome, where it was shown that 220,872 Holstein bulls, all from the Interbull database, descended from only five founders, three of which came from North America (Yue et al., 2015). This example also shows that intensive use of a small number of popular bulls known for their superior milk production results in high gene flow and connectivity between subpopulations, but this further reduces the genetic diversity of the overall metapopulation. Management of genomic diversity and inbreeding has therefore become an important issue in many national HF breeding programs (Ablondi et al., 2022; Doekes et al., 2018; Forutan et al., 2018; Rodríguez-Ramilo et al., 2015). The CZE subpopulation is no exception, as current breeding objectives focus on high milk yield and intensive use of a small number of mainly U.S. elite bulls is substantial, while selection (especially genomic) is known to increase inbreeding.

The main objective of this study was to estimate the genomic diversity status (genetic diversity, genomic inbreeding, and effective population size) of the CZE and compare it with other HF subpopulations. Genomic relatedness, population structure, and admixture of HF subpopulations, especially those affecting breeding of CZE, were estimated. We made additional efforts to analyze the influence of imported bulls on current CZE (cows).

2. Material and method

2.1. Analysed animals and quality control

The genome-wide data of 2178 animals of Holstein cattle from twelve countries; Belgium - 43 bulls (BEL_B), Canada - 98 bulls (CAN_B), Croatia - 84 cows (HRV_C), Czech Republic - 301 bulls (CZE_B), 298 Czech Republic - cows (CZE_C), France - 141 bulls (FRA_B), Netherlands - 290 bulls (NLD_B), Ireland - 129 bulls (IRL_B), Great Britain - 42 bulls (GBR_B), Germany - 343 bulls (DEU_B), Switzerland - 177 bulls (CHE_B), USA - 232 bulls (USA_B), were analysed. Animals were genotyped using Illumina BovineSNP50 BeadChip v1, v2 and v3 ($n = 2094$) and GGP100K Neogen Ltd ($n = 84$). In addition, genome-wide data from 98 animals, originally Czech Simental cattle (SIM_C), genotyped with the Illumina BovineSNP50 BeadChip v3 were used as an outlier population. Only SNPs co-occurring in these three panels were selected for analysis. The merged data set included 32 865 successfully reassigned SNPs occurring in all panels used. Quality control of genotype data was performed using PLINK 1.9 software (Chang et al., 2015). Data were trimmed using the following parameters: only autosomal SNPs with known chromosomal positions to exclude bias between males and females, individual call rate > 0.9, and SNP call rate > 0.9. This process ultimately resulted in the use of 29,502 SNPs.

2.2. Genetic diversity and runs of homozygosity inbreeding

Genomic diversity, represented as observed (H_O) and expected (H_E) heterozygosity (gene diversity), and Wright's inbreeding coefficient (F_{IS}) were calculated in PLINK 1.9. Genome-wide F_{IS} values indicate deviation from Hardy-Weinberg heterozygosity equilibrium, with negative values usually the result of avoidance of mating between close relatives, while positive values indicate inbreeding or mating between closely related individuals (Wright, 1965). Genome-wide diversity was quantified using a segment-based approach. Runs of homozygosity (ROH) segments were determined using detectRUNS (Biscarini et al., 2019). ROH segments were determined according to the following criteria (Ferenčaković et al., 2013): the minimum number of SNPs included in the ROH was set at 15; zero missing calls per window were allowed for ROHs $>$ 2 Mb, one for ROHs from 4 to 8 Mb, and two for ROHs $>$ 8 Mb categories, while no heterozygous SNPs were allowed. The maximum distance between the two SNPs was set at 1 Mb. The ROH-based inbreeding coefficient (F_{ROH}) was defined as the proportion of the genome in ROH relative to the total autosomal genome covered by SNPs on the chip. The F_{ROH} was estimated as the proportion of the autosome in ROH covering 29 chromosomes (F_{ROH}=$L_{ROH}/L_{AUTOSOME}$). With respect to ROH length, three inbreeding coefficients were estimated, indicating recent ($F_{ROH>8Mb}$), intermediate ($F_{ROH4–8Mb}$), and distant ($F_{ROH2–4Mb}$) inbreeding.

2.3. Effective population size

We calculated the contemporary population effective size for each subpopulation using two different approaches. To estimate only the contemporary effective population size, we used the approach described and implemented in NeEstimator v.2 software (Do et al., 2014), which uses the Jack-Knife method to estimate 95% confidence intervals and removes SNPs with frequencies below 5%. We also applied a recently developed approach implemented in the software GONE that provides both historical and contemporary estimates (Saura et al., 2021). The estimation approach developed in GONE is based on the functional relationship between gamete and/or linkage disequilibrium patterns and effective population size (Hill, 1981), but is calculated using a complex modeling approach. A genetic algorithm is implemented in the software GONE to derive the historical set of effective population size series that best minimises the sum of squared differences between the observed d^2 values (averaged squared correlations between two loci allele frequencies weighted by their variance) of the bins and those predicted at corresponding different demographic trajectories (Saura et al., 2021). In this study, we have referred to our estimates as Ne_{LD} to indicate that the estimates are based on gametic and/or linkage disequilibrium. Ne_{LD} is the size of an idealised population, often considered a Wright-Fisher population, that exhibits the same degree of genetic drift-change in gametic or linkage disequilibrium-as the

population under consideration. For more information on NeLD and potential biases, see Ryman et al. (2019), Waples (2021) and Waples et al. (2014).

2.4. Genetic relationship and population structure

The phylogenetic relationship between the 12 subpopulations of HF and the outlying Simental cattle population was represented by the Neighbor-Net network inferred from Nei's pairwise genetic distance (Nei, 1972). Nei's genetic distances were calculated using the StAMPP package (Pembleton et al., 2013). The phylogenetic network (Neighbor-Net) was created and drawn using SitsTree5 software (Huson and Bryant, 2006). Discriminant analysis of principal components (DAPC), implemented in the R package Adegenet (Jombart and Ahmed, 2011), was used to determine genetic structure and differentiation among Holstein subpopulations. The optimal number of principal components (PCs) reflecting the highest proportion of variance in the database was tested based on the a-score (Jombart and Ahmed, 2011). In addition, DAPC was used to assign individuals and obtain the affiliation probability, which represents the total genetic background of an individual. The population structure was further evaluated using a Bayesian clustering approach implemented in the software STRUCTURE (Pritchard et al., 2000). Before analysing genetic differentiation within and between populations, data sets were pruned based on the threshold for linkage disequilibrium between SNPs (0.05) with a window size of 50 and a step size of five SNPs. After this cleaning of the data, only 2223 SNPs remained for subsequent analyses. The analysis was performed with an admixture and correlated allele frequency model, using 106 iterations with a burn-in period of 105. Runs were repeated 20 times for each assumed K (1–13). The most likely K value in the data set was determined according to (Evanno et al., 2005) using STRUCTURE HARVESTER (Earl and vonHoldt, 2012). Visualisations of population structure were performed using the web-based tool CLUMPAK (Kopelman et al., 2015).

2.5. The genetic influence of imported bulls on the Czech Holstein subpopulation

The Czech Holstein population is open, with significant import of genetic material from abroad. The term Czech bulls refers to bulls born in the Czech Republic. These are bulls from imported embryos (mainly from the USA) and sons of heifers born from these embryos. The sires of the Czech bulls are the best bulls of the world population of Holstein cattle (mainly from the USA). The Czech Holstein cow entered the herd in the nineties of the last century by crossing with foreign bulls (from different countries, mainly from Europe). Most cows in the Czech Republic are inseminated by foreign bulls (from different countries). Insemination doses of Czech bulls make up only a small part of the insemination doses used in the Czech Republic. As mentioned above, most of the insemination doses used for production cows come from abroad (Europe, USA). Therefore, the Czech Holstein cows (CZE_C) have a different genetic background than the studied Czech Holstein bulls (CZE_B). Therefore, to measure the influence of imported bulls, we calculated the genetic differentiation between female and male Czech Holsteins (CZE_C versus CZE_B).

Thus, genome-wide F_{ST} (coefficient of population differentiation) estimates were calculated for each SNP (Weir and Cockerham, 1984) using the SNP & Variation Suite (SVS) v8.7.0 software package (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). Subsequently, F_{ST} estimates were normalised by mean frequency, while transformed values were represented as -log(P). As outliers, 0.1% SNPs with the highest values were selected (30 SNPs), which corresponded to F_{ST} value of 0.074 and a -log(P) value of 9.058. In addition, candidate genes and QTLs were annotated within 0.4 Mb wide genomic regions around each outlier (0.2 Mb from each side). While genes were annotated using information from the SNP & Variation Suite (SVS) v8.7.0 software package (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com), QTLs were annotated using the GALLO R package (Fonseca et al., 2020) to query the Animal QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/index, accessed 04/29/2022) for previously identified QTLs in the regions of interest. Trait enrichment analysis was performed for the annotated QTLs, and enriched classes and traits per chromosome with an FDR-adjusted P value of less than 0.05 were considered significant and represented as -log(P).

3. Results and discussion

3.1. Genetic diversity and runs of homozygosity inbreeding

Our analyses showed that inbreeding levels were higher in all subpopulations of HF compared to SIM_C (Table 1 and Fig. 1). For example, the estimated mean $F_{ROH>2Mb}$ in SIM_C (0.045) was more than twice as small as in the twelve HF subpopulations (see below for more details on population structure), where the $F_{ROH>2Mb}$ ranged from 0.091 to 0.133. At the same time, the estimated $F_{ROH>2Mb}$ was lower in tree subpopulations (CHE_B, HRV_C, and IRL_B) outside the "core" HF, ranging from 0.077 to 0.088 (Table 1). The similar trend was observed for other indicators of genetic diversity (H_O, H_E, and F_{IS}), although the observed differences for H_O and H_E were not significant. Interestingly, with the exception of HRV_C, CHE_B, IRL_B, but also CZE_C, and GBR_B subpopulations, all the remaining subpopulation had significantly positive F_{IS} values. The F_{IS} negative values usually indicating of avoidance of mating between close relatives, while positive values indicate inbreeding or mating between closely related individuals.

In contrast, F_{IS} values were significantly negative in CHE_B, HRV_C, and IRL_B, indicating avoidance of inbreeding, while a negative F_{IS} value was also observed in the SIM_C breed. Quite large differences, higher F_{IS} and $F_{ROH>2Mb}$ values, were observed in CZE_B than in CZE_C, indicating that the reduction in diversity and extreme increase in inbreeding is due to imported semen. Of particular concern is the very high recent inbreeding ($F_{ROH>8Mb}$) observed in CZE_B, which exceeded the levels of

Table 1
Genetic diversity indicators (H_O, H_E and F_{IS}) and runs of homozygosity based inbreeding level in 12 HF subpopulations and Czech Simental cows.

	Animals	H_O	H_E	F_{IS}	$F_{ROH>2Mb}$
BEL_B	43	0.365 ±0.024	0.358 ±0.022	0.012±0.006	0.112 ±0.006
CAN_B	98	0.363 ±0.015	0.357 ±0.015	0.017±0.004	0.114 ±0.004
CZE_C	298	0.372 ±0.008	0.371 ±0.008	−0.007 ±0.002	0.091 ±0.002
CZE_B	301	0.356 ±0.008	0.358 ±0.008	0.036±0.003	0.133 ±0.003
DEU_B	343	0.363 ±0.008	0.363 ±0.007	0.017±0.002	0.113 ±0.002
FRA_B	141	0.366 ±0.012	0.360 ±0.012	0.010±0.003	0.108 ±0.003
GBR_B	42	0.368 ±0.025	0.359 ±0.022	0.005±0.005	0.102 ±0.005
HRV_C	84	0.373 ±0.016	0.368 ±0.015	−0.010 ±0.003	0.088 ±0.003
CHE_B	177	0.373 ±0.011	0.368 ±0.010	−0.010 ±0.002	0.084 ±0.002
IRL_B	129	0.374 ±0.012	0.375 ±0.011	−0.013 ±0.003	0.077 ±0.003
NLD_B	290	0.365 ±0.008	0.365 ±0.008	0.012±0.002	0.108 ±0.002
USA_B	232	0.364 ±0.010	0.358 ±0.009	0.016±0.003	0.113 ±0.002
SIM_C	90	0.376 ±0.015	0.373 ±0.014	−0.016 ±0.004	0.045 ±0.004

H_O is observed heterozygosity, H_E is expected heterozygosity, F_{IS} is Wright's inbreeding coefficient and $F_{ROH>2\ MB}$ is ROH based genomic inbreeding coefficient.

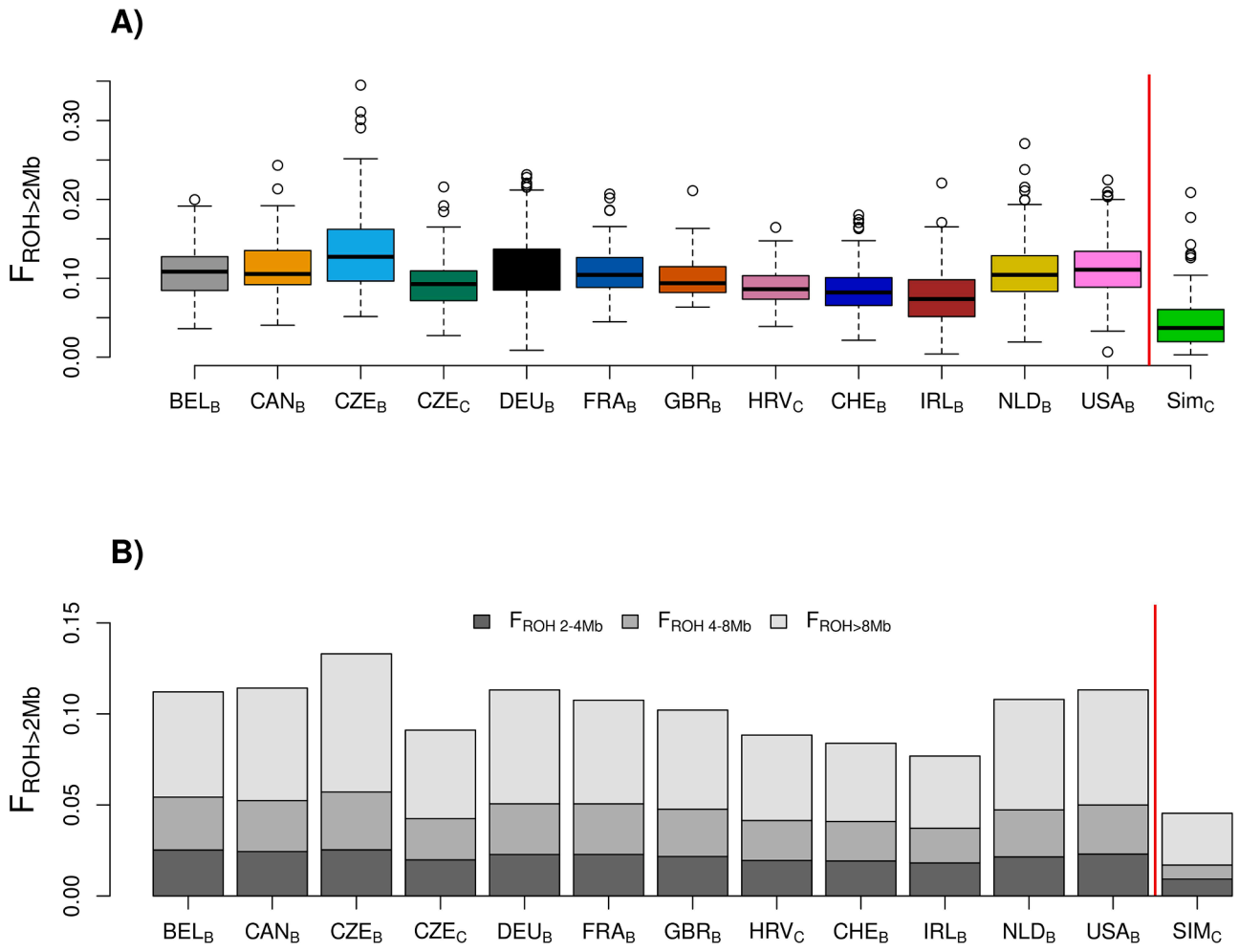


Fig. 1. Distribution of individual genomic inbreeding in 12 HF subpopulations [Belgium - BEL_B, Canada - CAN_B, Croatia - HRV_C, Czech Republic - CZE_B (bulls), Czech Republic - CZE_C (cows), France - FRA_B, Netherlands - NLD_B, Ireland - IRL_B, Great Britain - GBR_B, Germany - DEU_B, Switzerland - CHE_B, USA - USA_B)] and Czech Simmental breed - SIM_C (cows). (a) Boxplot representation of genomic inbreeding ($F_{ROH>2Mb}$); (b) stacked bar representation of the partitioning of genomic inbreeding to remote ($F_{ROH2-4Mb}$), intermediate ($F_{ROH4-8Mb}$) and recent ($F_{ROH>8Mb}$) origins of autozygosity.

all other (sub)populations analysed (Fig. 1B).

In general, very high recent inbreeding was observed in all HF subpopulations ($F_{ROH>8Mb}$), reflecting the strong intensity of selection in bulls. This explains why genetic diversity was low in all bull populations, while inbreeding was high. However, the observed differences in inbreeding levels, e.g., HF bulls sampled subpopulations versus SIM_C, were too large to affect the observed trends. For example, a very high level of inbreeding ($F_{ROH>2Mb}$) was observed in the Italian HF population (mainly cows), where the observed level of inbreeding ranged from 0.09 for animals born between 2002 and 2005 to 0.16 for animals born between 2016 and 2020 (Ablondi et al., 2022). However, the estimated $F_{ROH>2Mb}$ in Ablondi et al. (2022) is based on a denser SNP array (≈70 K) and is not fully comparable, although the difference is not expected to be significant. The estimated level of inbreeding of 0.12 ($F_{ROH>2Mb}$) reported by Lozada-Soto et al. (2022) in the USA HF is consistent with our estimates. A similarly high degree of inbreeding ($F_{ROH>2Mb}$ equal to 0.111) was also estimated by Szmatoła et al. (2019) in the Polish HF (bulls and cows).

3.2. *Effective population size*

Contemporary estimates of linkage (gametic) disequilibrium effective population size obtained using algorithms from NeEstimator v2 (Do et al., 2014) and GONE (Santiago et al., 2020) software are shown in Table 2. There were often no significant differences between the estimates obtained with two different approaches because the 95% confidence intervals (95% CI) largely overlapped, although some exceptions were also observed (CZE_C, CZE pooled, HRV_C, and IRL_B).

Unfortunately, in some subpopulations (HRV_C, BEL_B, and GBR_B) the observed 95% CIs were quite wide and provided less informative estimates (Table 2).

In addition to the estimates presented separately for the CZE_C and CZE_B samples, we also estimated Ne_{LD0} for the pooled CZE sample (bulls and cows). Compared with the other HF subpopulations, estimated Ne_{LD0} was among the highest in the pooled CZE sample (Table 2). However, we should be aware that the estimates based on the bull samples are much lower, as expected and commented by Ablondi et al. (2022). Although our subpopulation samples consisted mainly of bulls, much higher estimates than expected (Table 2) were observed, mostly above 100, which is higher than the critical value of 50 recommended by (FAO, 1998). Estimates of effective population size (Ne_{LD}) obtained from genome-wide data in other studies ranged from about 80 in the North American (Canada and USA) HF population (Sargolzaei et al., 2008) to about 150 in the Australian HF cattle (Hayes et al., 2003). Our estimates for the NLD_B were much higher than those of (Doekes et al., 2018), which ranged from 69 to 102 for the period from 1995 to 2015, although these estimates related to inbreeding effective population size were reported without confidence intervals. For example, Ablondi et al. (2022) observed an inbreeding effective population size of 55 in the Italian population HF, estimated from the pedigree, whereas a much higher

Table 2
Contemporary and historical linkage (gametic) disequilibrium effective population size estimated by two different approaches implemented in the NeEstimator v2 (Do et al., 2014) and GONE (Santiago et al., 2020) software in HF subpopulations.

HF subpopulation (Origin)	Ne_{LD0} - NeEstimator v2	Ne_{LD0} - GONE	Ne_{LD10} - GONE
BEL_B (Belgium)	**135** (89–255)	**120** (90–161)	**95** (73–126)
CAN_B (Canada)	**107** (85–140)	**97** (83–114)	**102** (87–120)
CHE_B (Switzerland)	**102** (88–118)	**110** (99–123)	**118** (106–132)
CZE_C (Czech Republic)	**359** (329–395)	**181** (167–197)	**117** (109–127)
CZE_B (Czech Republic)	**147** (132–164)	**154** (143–167)	**86** (81–93)
CZE pooled (Czech Republic)	**261** (240–283)	**202** (193–213)	**109** (105–114)
DEU_B (Germany)	**196** (181–212)	**188** (175–203)	**94** (89–101)
FRA_B (France)	**146** (123–176)	**155** (136–178)	**117** (103–134)
GBR_B (Great Britain)	**145** (107–219)	**108** (81–145)	**124** (91–168)
HRV_C (Croatia)	**130** (105–166)	**206** (167–255)	**55** (48–65)
IRL_B (Ireland)	**61** (51–72)	**105** (92–121)	**200** (172–234)
NLD_B (Netherlands)	**207** (190–226)	**179** (165–195)	**109** (102–118)
USA_B (USA)	**158** (144–175)	**140** (128–154)	**99** (91–108)

Ne_{LD0} is the estimated linkage (gametic) effective population size in the contemporary population, whereas Ne_{LD10} refers to historical estimates of the same populations 10 generations back.

estimate (120) was based only on SNPs placed on the same chromosomes (Ne_{LD}). Surprisingly, historical estimates (Ne_{LD10}) were mostly lower than contemporary estimates (Ne_{LD0}), suggesting that Ne_{LD} has increased over the past 10 generations, whereas a significant decrease was observed only in the IRL_B. The increase in inbreeding effective population size from 65 (animals born between 1960 and 1979) to 101 (animals born between 2000 and 2013) was also observed in the Spanish HF (Rodríguez-Ramilo et al., 2015).

We would like to emphasize that the observed Ne_{LD} estimates should be used with caution for several reasons that may have influenced our estimates. For example, estimating effective population size is very complex when populations are subdivided, especially when there is a high unbalanced migration rate between subpopulations, which is the case in this study where HF can be considered as a large metapopulation (Ryman et al., 2019). Thus, the presence of admixture or migration from other HF subpopulations or even breeds (Simmental cattle), as observed in this study (see below), may lead to biased estimates of Ne_{LD}. At the same time, estimates of historical Ne_{LD} are even more unreliable when there is a possible bias due to the presence of admixed individuals. The subpopulations analysed had overlapping generations, which is another cause of the potential bias in estimating effective population size in this study (Waples et al., 2014). Finally, estimates of effective population size can differ drastically between inbreeding and linkage (gametic) disequilibrium in populations undergoing migration (Ryman et al., 2019), which is to be expected in HF subpopulations, with additional complications arising from the presence of selection and non-random mating.

3.3. Genetic relationship and population structure

The population structure of the HF metapopulation was analysed using different approaches. For example, Fig. 2 shows the relationships between 12 HF subpopulations and SIM_C resulting from DAPC and Neighbor-Net analysis. The first four discriminant functions of DAPC explained 97% of the total variation, whereas each discriminant function contributed to the separation of SIM_C, HRV_C, IRL_B, and CHE_B from the "core" HF metapopulation (Fig. 1A, B, C). The NeighborNet network derived from the pairwise genetic distances of Nei is consistent with the DAPC analysis and shows clear differentiation of the HRV_C, IRL_B, and CHE_C subpopulations toward SIM_C apart from the other HF subpopulations (Fig. 2D). The results support the hypothesis that the formation of the HF subpopulations in some EU countries is due to the crossing of imported HF semen with local Simmental cows. Our hypothetically subdivided Czech HF subpopulations of bulls and cows (CZE_B and CZE_C), although slightly separated from each other, were placed in the "core" HF metapopulation with subpopulations BEL_B, CAN_B, DEU_B, FRA_B, GBR_B, NLD_B and USA_B. This result is logical because semen from popular and highly productive North American bulls was intensively imported into the Czech Republic and other European countries.

This also explains the neighbor position of CZE_B near USA_B, whereas CZE_C is on the other side of the Neighbor-Net network toward HRV_C (Fig. 2D). The visual illustration of the relationship between the subpopulations of HF is further quantified by the pairwise genome-wide F_{ST} estimates (Table 3). The three slightly separated subpopulations (HRV_C, IRL_B, and CHE_B) had high estimated mean F_{ST} (MF_{ST}) that ranged from 0.019 to 0.027, whereas the MF_{ST} for the main HF "gene pool" ranged from 0.010 to 0.014.

For unsupervised identification of population structure and estimation of admixture level, we used the algorithm implemented in the software STRUCTURE because it can reveal "hidden structure" and quantify admixture without determining a priori membership in individual clusters. We ran STRUCTURE from $K = 1$ to $K = 13$, assuming that the highest number of potential clusters is 13, i.e., one breed plus 12 subpopulations, each representing a single cluster. The Ln Pr(G|K) value increased slightly and steadily, with only a significant drop at $K = 6$ caused by lower Ln Pr(G|K) values in some runs (Fig. 3A). At the same time, the highest rate of change of Ln Pr(G|K) between successive K values was observed at $K = 5$ (Fig. 3B). Thus, following the recommendations of Pritchard et al. (2000), Falush et al. (2007), and Evanno et al. (2005), it is very likely that $K = 5$ applies to the analysed data set. While SIM_C clearly stood out as a different "yellow" cluster (in this case representing "SIM_C cluster") from the HF metapopulations, the other four clusters were distributed across all HF subpopulations, confirming the extensive gene flow known to occur through semen importation (Fig. 3C).

This result suggests that any differences in the subpopulations of HF are due to the composition of the estimated five clusters, probably because of their different breed development histories, selection intensities, and breeding goals. Interestingly, the main cluster defining SIM_C with the mean individual membership equal to 0.777 (presented by yellow color in the Fig. 3C) was present in all HF subpopulations, ranging from 0.089 to 0141 in the "core" HF metapopulation. Much higher proportion of the "SIM_C cluster" (yellow) was estimated in the HRV_C (0.356), CHE_B (0.296) and IRL_B (0.214) subpopulations further explaining its slight differentiation from the "core" HF metapopulation presented in Fig. 2. In CZE_B, we found some degree of clustering with a considerable presence of "red" clusters, which was also noticeable in DEU_B, NLD_B, and USA_B (Fig. 3C).

3.4. The genetic influence of imported bulls on the Czech Holstein subpopulation

Our phylogenetic network (Neighbor-Net) and unsupervised STRUCTURE analyses revealed slight genetic differences between samples from the cows (CZE_C) and samples from the imported AI bulls (Figs. 2 and 3). To predict future genetic changes expected from high use of AI semen in the Czech Republic HF, we identified genomic regions where genetic differences (estimated F_{ST} values) between CZE_C and CZE_B were most pronounced, and their genomic positions are shown in Fig. 4.

The genes located in these regions were identified and are presented

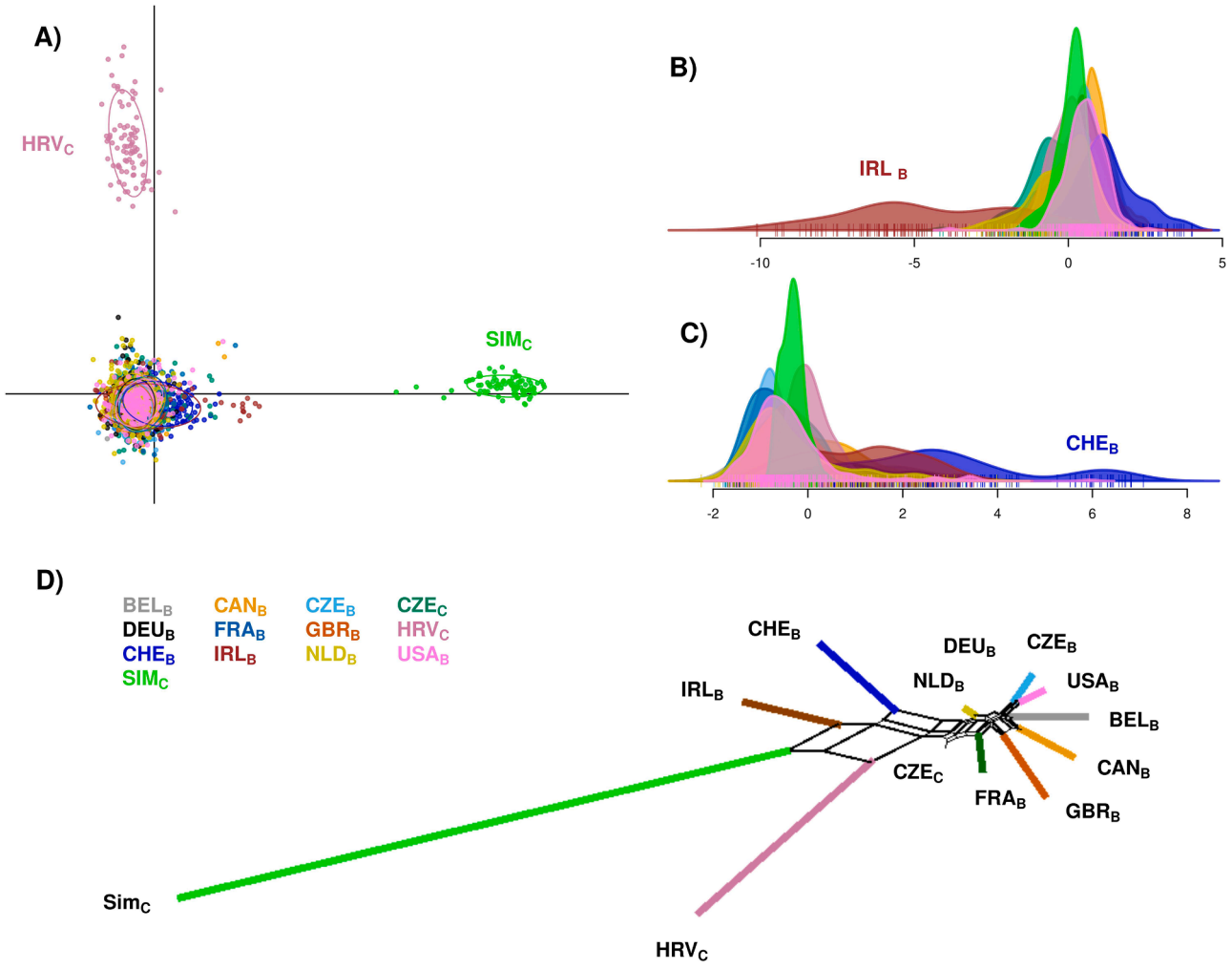


Fig. 2. visualization of the genetic relationship of 12 HF subpopulations [Belgium - BEL_B, Canada - CAN_B, Croatia - HRV_C, Czech Republic - CZE_B (bulls), Czech Republic - CZE_C (cow), France - FRA_B, Netherlands - NLD_B, Ireland - IRL_B, Great Britain - GBR_B, Germany - DEU_B, Switzerland - CHE_B, USA - USA_B)] and Czech Simmental breed – SIM_C (cows). A) Variation of the first two discriminant function of the DAPC. B) Variation of the third discriminant function of the DAPC. C) Variation of the fourth discriminant function of the DAPC. D) Neighbor-Net inferred from pairwise Nei's genetic distances.

Table 3
Population differentiation among 12 HF subpopulations [Belgium - BEL_B, Canada -CAN_B, Croatia - HRV_C, Czech Republic - CZE_B (bulls), Czech Republic - CZE_C (cow), France - FRA_B, Netherlands - NLD_B, Ireland - IRL_B, Great Britain - GBR_B, Germany - DEU_B, Switzerland - CHE_B, USA - USA_B)] and Czech Simmental breed – SIM_C (cows) based on genome-wide F_{ST} estimates.

	BEL_B	CAN_B	CHE_B	CZE_B	CZE_C	DEU_B	FRA_B	GBR_B	HRV_C	IRL_B	NLD_B	USA_B	MF_{ST}
BEL_B													0.011
CAN_B	0.003												0.014
CHE_B	0.013	0.013											0.019
CZE_B	0.004	0.008	0.019										0.014
CZE_C	0.005	0.008	0.013	0.007									0.011
DEU_B	0.001	0.004	0.012	0.002	0.004								0.010
FRA_B	0.005	0.008	0.017	0.005	0.003	0.003							0.012
GBR_B	0.003	0.003	0.013	0.005	0.004	0.002	0.003						0.011
HRV_B	0.025	0.028	0.028	0.027	0.021	0.023	0.023	0.023					0.027
IRL_B	0.017	0.022	0.019	0.020	0.011	0.016	0.015	0.016	0.028				0.020
NLD_B	0.002	0.007	0.014	0.004	0.003	0.002	0.003	0.003	0.022	0.013			0.011
USA_B	0.003	0.006	0.018	0.002	0.007	0.002	0.005	0.004	0.027	0.021	0.004		0.013
SIM_C	0.056	0.059	0.049	0.060	0.051	0.055	0.056	0.057	0.057	0.047	0.054	0.059	0.055

in the Supplement along with their physical location, estimated F_{ST} values, and significance (Supplemental Table 1). The identified genes are located on chromosomes 2 (PLEKHA3, FKBP7, DFNB59, PRKRA, OSBPL6, NFE2L2, HNRNPA3, MALRD1, PLXDC2, SPAG16, RHBDD1, COL4A3), 3 (NOTCH2, REG4, HMGCS2, PHGDH, DDX18, CCDC93, HDAC4), 4 (SFRP4, EPDR1, STARD3NL, TRGC6, BBS9, RP9, NT5C3A, FKBP9, KBTBD2, AVL9, CREB5, JAZF1), 5 (MYRFL, RAB3IP, BEST3, LRRC10, CCT2, FRS2), 6 (LDB2), 7 (FNIP1, RAPGEF6, CGC42SE2), 8 (MSRA, KIF13B), 9 (OLIG3, UTRN, SASH1, UST), 11 (HMCN2, ASS1, FUBP3, PRDM12, EXOSC2, ABL1), 13 (MACROD2, TAF3, ATP5C1, KIN,

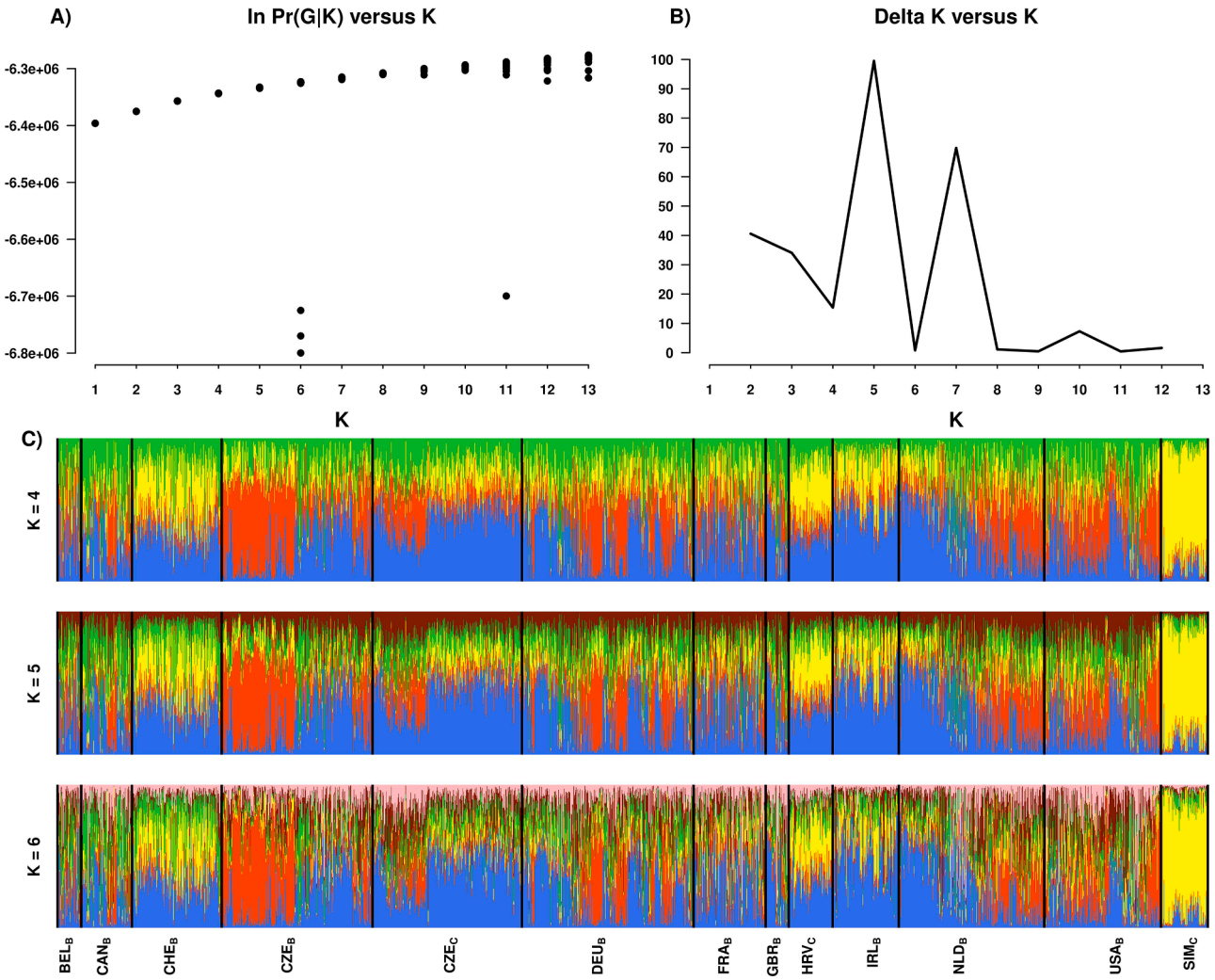


Fig. 3. Results of unsupervised population structure and admixture analysis using the algorithm STRUCTURE for 2178 individuals HF from 12 subpopulations HF (Belgium - BEL_B, Canada – CAN_B, Croatia - HRV_C, Czech Republic bulls - CZE_B, Czech Republic cows - CZE_C, France - FRA_B, Netherlands - NLD_B, Ireland - IRL_B, Great Britain - GBR_B, Germany - DEU_B, Switzerland - CHE_B, USA - USA_B) and for 98 Czech Simmental cows (SIM_C). A) Plot of Ln Pr(G|K) values as a function of the number of clusters (K). B) Plot of K values for each K based on the second order rate of change of the likelihood function as a function of K. C) Graphical representation of the selection of results from STRUCTURE at $K = 4$, $K = 5$, and $K = 6$, where each individual is represented by a vertical line divided into K coloured segments whose length is proportional to the estimated membership of the inferred cluster.

ITIH2. ITIH5, PTPN1, FAM65C, PARD6B, BCAS4, ADNP, DPM1), 19 (PRPSAP1, QRICH2, RNF157, FOXJ1, EXOC7, ZACN, GALR2, SRP68, EVPL, CDK3, TEN1, ACOX1, FBF1, MRPL38, TRIM65, TRIM47) and 29 (PAG3, PAG6, PAG11, FGF19, FGF4, FGF3, ANO1, MRPL21, IGHMBP2, MRGPRF, TPCN2). Based on QTL annotation and their enrichment analyses (Fonseca et al., 2020), it was recognised that the majority of QTLs annotated in these regions mainly affect meat and carcass traits (Supplementary Figure 1). This result informs breeders, as the main decision makers, that the intensive use of current AI bulls influences meat and carcass traits in addition to increasing milk production, which is defined by the breeding objective of the Czech HF.

4. Conclusion

In this study, we have shown that the effective population size (NE_{LD}) in the Czech HF population was relatively high, ranging from 202 (GONE) to 283 (NeEstimator v2). Compared to other HF subpopulations, this should not be a problem for breeding. However, the inbreeding observed in the AI bull sample was very high ($F_{ROH>2Mb}$=0.133) and far exceeded the inbreeding observed in the cow sample ($F_{ROH>2Mb}$=0.091). Of particular concern was the very high recent inbreeding ($F_{ROH>8Mb}$=0.08) of the CZE_B, indicating intentional mating of close relatives, as further evidenced by the high and significant F_{IS} (0.036 ±0.003). It is very likely that the future use of highly inbred AI bulls will lead to a high level of inbreeding in the entire Czech HF subpopulation. Our phylogenetic analyses showed that the Czech HF belonged to the "core HF metapopulation", together with Belgian, British, Canadian, Dutch, French, German, and USA subpopulations, which was slightly separated from the Swiss, Irish, and Croatian subpopulations. Finally, to predict future genetic changes expected from the high use of AI bull semen in the Czech Republic HF, we identified genomic regions where genetic differences (estimated F_{ST} values) between CZE_C and CZE_B were most pronounced. Our further enrichment analyses of SNPs with the 0.1% highest F_{ST} values revealed that genes located in the most differentiating genomic regions primarily influence meat and carcass traits. This information is of importance to breeders, as they might expect continued use of current AI bulls to result in changes in meat and carcass performance in addition to increased milk production.

CRediT authorship contribution statement

Lubos Vostry: Writing – original draft, Methodology, Visualization,

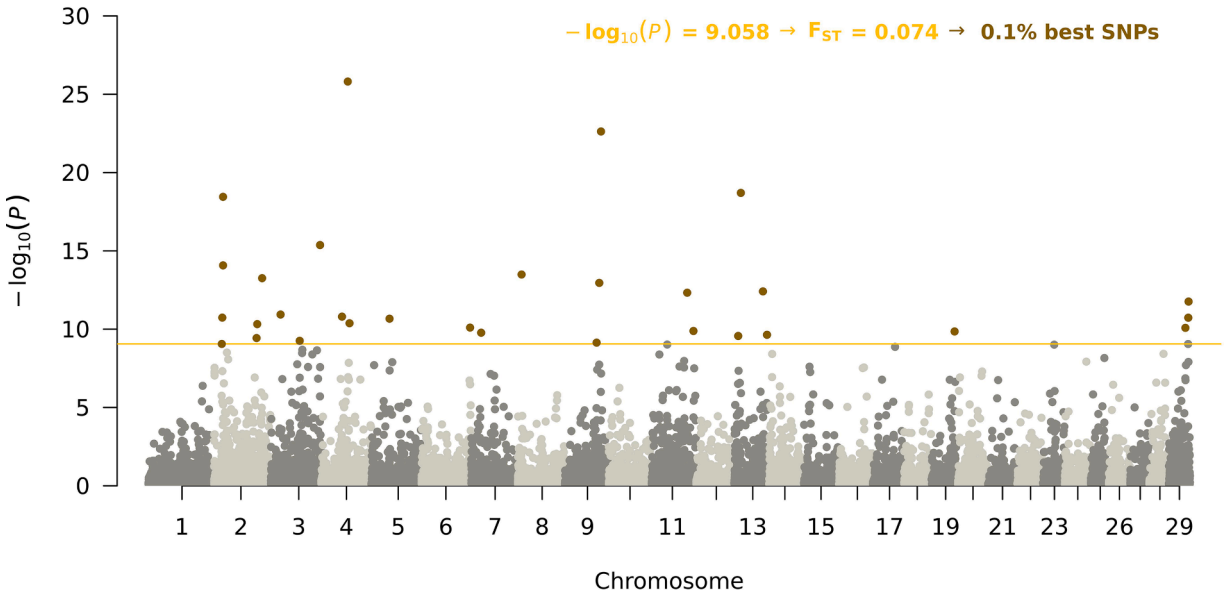


Fig. 4. Manhattan plot showing genomic distribution of F_{ST} values distinguishing Czech HF cows (CZE_C) and bulls (CZE_B). The horizontal yellow line represents the significance threshold ($F_{ST} = 0.074$) for the SNPs with the highest F_{ST} differentiation (0.1%).

Formal analysis, Data curation, Writing – review & editing. **Hana Vostra-Vydrova:** Writing – original draft, Methodology, Visualization, Formal analysis, Data curation, Writing – review & editing. **Nina Moravcikova:** Formal analysis, Methodology. **Radovan Kasarda:** Methodology. **Vlatka Cubric-Curik:** Writing – original draft, Methodology. **Michaela Brzakova:** Formal analysis. **Johan Solkner:** Methodology. **Mario Shihabi:** Formal analysis, Visualization. **Jorge Angel Hidalgo Moreno:** Data curation. **Maria Spehar:** Data curation. **Ino Curik:** Writing – original draft, Methodology, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Acknowledgement

L.V., H.V.-V. and M.B. acknowledge support from the Ministry of Education Youth and Sports of the Czech Republic and Ministry of Agriculture of the Czech Republic (Projects LTAUSA19117, QK1810253 and MZE-RO0723), N.M. and R.K. acknowledge support from the Slovak Research and Development Agency (grant numbers APVV-20–0161 and APVV-17–0060), V.C.-C., M.Sh. and I.C. acknowledge support from the Croatian Science Foundation (Project ANAGRAMS-IP-2018–01–8708).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.livsci.2023.105261.

References

Ablondi, M., Sabbioni, A., Stocco, G., Cipolat-Gotet, C., Dadousis, C., Kaam, J.-T.V., Finocchiaro, R., Summer, A., 2022. Genetic diversity in the Italian Holstein dairy cattle based on pedigree and SNP data prior and after genomic selection. Front. Vet. Sci. 8, 773985 https://doi.org/10.3389/fvets.2021.773985.

Biscarini, F., Cozzi, P., Gaspa, G., Marras. G., 2019. Detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.5. Accessed June 20, 2019. https: //cran.r-project.org/web/packages/detectRUNS/.

Boichard, D., Ducrocq, V., Croiseau, P., Fritz, S., 2016. Genomic selection in domestic animals: principles, applications and perspectives. C. R. Biol. 339, 274–277. https://doi.org/10.1016/j.crvi.2016.04.007.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaSci 4, 7. https://doi.org/10.1186/s13742-015-0047-8.

Cole, J.B., 2015. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. Genet. Sel. Evol. 47, 94. https://doi.org/10.1186/s12711-015-0174-9.

de Jong, G., Bijma, P., 2002. Selection and phenotypic plasticity in evolutionary biology and animal breeding. Livest. Prod. Sci. 78, 195–214. https://doi.org/10.1016/S0301-6226(02)00096-9.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Resour. 14, 209–214. https://doi.org/10.1111/1755-0998.12157.

Doekes, H.P., Veerkamp, R.F., Bijma, P., Hiemstra, S.J., Windig, J.J., 2018. Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. Genet. Sel. Evol. 50, 15. https://doi.org/10.1186/s12711-018-0385-y.

Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361. https://doi.org/10.1007/s12686-011-9548-7.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Falush, D., Stephens, M., Pritchard, J.K., 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles: TECHNICAL ARTICLE. Mol. Ecol. Notes 7, 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x.

FAO, 1998. Secondary Guidelines For Development of National Farm Animal Genetic Resources Management plans: Animal Recording For Medium Input Production Environment. FAO, Rome.

Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genet. Sel. Evol. 45, 42. https://doi.org/10.1186/1297-9686-45-42.

Ferenčaković, M., Sölkner, J., Kapš, M., Curik, I., 2017. Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. J. Dairy Sci. 100, 4721–4730. https://doi.org/10.3168/jds.2016-12164.

Fonseca, P.A.S., Suárez-Vega, A., Marras, G., Cánovas, Á., 2020. GALLO: an R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. Gigascience 9, 149. https://doi.org/10.1093/gigascience/giaa149 giaa.

Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F.S., Sargolzaei, M., 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics 19, 98. https://doi.org/10.1186/s12864-018-4453-z.

Goddard, M., 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245–257. https://doi.org/10.1007/s10709-008-9308-0.

Hayes, B.J., Visscher, P.M., McPartlan, H.C., Goddard, M.E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13, 635–643. https://doi.org/10.1101/gr.387103.

Hill, W.G., 2000. Maintenance of quantitative genetic variation in animal breeding programmes. Livest. Prod. Sci. 63, 99–109. https://doi.org/10.1016/S0301-6226(99)00115-3.

Hill, W.G., 1981. Estimation of effective population size from data on linked genes. Adv. Appl. Probab. 13 https://doi.org/10.2307/1426455, 4–4.
Hodges, J., 2006. Conservation of genes and culture: historical and contemporary issues. Poult. Sci. 85, 200–209. https://doi.org/10.1093/ps/85.2.200.
Howard, J.T., Pryce, J.E., Baes, C., Maltecca, C., 2017. Invited review: inbreeding in the genomics era: inbreeding, inbreeding depression, and management of genomic variability. J. Dairy Sci. 100, 6009–6024. https://doi.org/10.3168/jds.2017-12787.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030.
Jombart, T., Ahmed, I., 2011. 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.
Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across. Mol. Ecol. Resour. 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387.
Lozada-Soto, E.A., Tiezzi, F., Jiang, J., Cole, J.B., VanRaden, P.M., Maltecca, C., 2022. Genomic characterization of autozygosity and recent inbreeding trends in all major breeds of US dairy cattle. J. Dairy Sci. 105, 8956–8971. https://doi.org/10.3168/jds.2022-22116.
Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292. https://doi.org/10.1086/282771.
Pembleton, L.W., Cogan, N.O.I., Forster, J.W., 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. 13, 946–952. https://doi.org/10.1111/1755-0998.12129.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Rodríguez-Ramilo, S.T., Fernández, J., Toro, M.A., Hernández, D., Villanueva, B., 2015. Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in the Spanish Holstein Population. PLoS ONE 10, e0124157. https://doi.org/10.1371/journal.pone.0124157.
Ryman, N., Laikre, L., Hössjer, O., 2019. Do estimates of contemporary effective population size tell us what we want to know? Mol. Ecol. 28, 1904–1918. https://doi.org/10.1111/mec.15027.
Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J., Caballero, A., 2020. Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. Mol. Biol. Evol. 37, 3642–3653. https://doi.org/10.1093/molbev/msaa169.
Sargolzaei, M., Schenkel, F.S., Jansen, G.B., Schaeffer, L.R., 2008. Extent of linkage disequilibrium in Holstein Cattle in North America. J. Dairy Sci. 91, 2106–2117. https://doi.org/10.3168/jds.2007-0553.
Saura, M., Caballero, A., Santiago, E., Fernández, A., Morales-González, E., Fernández, J., Cabaleiro, S., Millán, A., Martínez, P., Palaiokostas, C., Kocour, M., Aslam, M.L., Houston, R.D., Prchal, M., Bargelloni, L., Tzokas, K., Haffray, P., Bruant, J.-S., Villanueva, B., 2021. Estimates of recent and historical effective population size in turbot, seabream, seabass and carp selective breeding programmes. Genet. Sel. Evol. 53, 85. https://doi.org/10.1186/s12711-021-00680-9.
Strandén, I., Kantanen, J., Russo, I.-R.M., Orozco-terWengel, P., Bruford, M.W., Consortium, the Climgen, 2019. Genomic selection strategies for breeding adaptation and production in dairy cattle under climate change. Heredity (Edinb) 123, 307–317. https://doi.org/10.1038/s41437-019-0207-1.
Szmatoła, T., Gurgul, A., Jasielczuk, I., Ząbek, T., Ropka-Molik, K., Litwińczuk, Z., Bugno-Poniewierska, M., 2019. A Comprehensive analysis of runs of homozygosity of eleven cattle breeds representing different production types. Animals 9, 1024. https://doi.org/10.3390/ani9121024.
VanRaden, P.M., Olson, K.M., Null, D.J., Hutchison, J.L., 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J. Dairy Sci. 94, 6153–6161. https://doi.org/10.3168/jds.2011-4624.
Waples, R.S., 2021. Relative precision of the sibship and LD methods for estimating effective population size with genomics-scale datasets. J. Heredity 112, 535–539. https://doi.org/10.1093/jhered/esab042.
Waples, R.S., Antao, T., Luikart, G., 2014. Effects of overlapping generations on linkage disequilibrium estimates of effective population size. Genetics 197, 769–780. https://doi.org/10.1534/genetics.114.164822.
Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution (N Y) 38, 1358. https://doi.org/10.2307/2408641.
Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution (N Y) 19, 395–420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x.
Yue, X.-P., Dechow, C., Liu, W.-S., 2015. A limited number of Y chromosome lineages is present in North American Holsteins. J. Dairy Sci. 98, 2738–2745. https://doi.org/10.3168/jds.2014-8601.