



Genomic diversity and population structure analysis of four geographically distinct Holstein cattle populations using genome-wide DNA profiling

Ronak Salehi¹, Arash Javanmard¹, Mahdi Mokhber², and Sadegh Alijani¹

¹Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

²Department of Animal Science, Faculty of Agriculture, Urmia University, Urmia, Iran

Correspondence: Arash Javanmard (a.javanmard@tabrizu.ac.ir) and Mahdi Mokhber (m.mokhber@urmia.ac.ir)

Received: 19 March 2025 – Revised: 10 May 2025 – Accepted: 13 June 2025 – Published: 11 September 2025

Abstract. Holstein cattle are the world's most widely distributed dairy breed, present in more than 150 countries. Despite their large population and widespread presence, the genomic diversity of Holstein subpopulations across the globe is surprisingly low; therefore, genetic monitoring of these populations, especially in countries that import their genetic material, is crucial. This study aimed to compare the genomic diversity of Iranian Holstein cattle with populations from France, Poland, and Sweden, providing a comprehensive overview of the Holstein breed's genetic structure. In total, 399 Holstein cattle from four geographically distinct countries, namely, Iran ($n = 25$), France ($n = 153$), Poland ($n = 198$), and Sweden ($n = 23$), were selected for analysis. Quality control and data filtering were conducted using PLINK 1.9. After quality control, a pruning process was applied to the genomic data, resulting in 10 822 single-nucleotide polymorphism markers for genomic diversity analysis. Due to limitations in the software used for population structure analysis, missing data were imputed using Beagle 5.2. To analyze the genetic structure of the populations, discriminant analysis of principal components (DAPC) and population admixture analysis were performed using the adegenet package in R. Moreover, genomic diversity and distances were assessed using Weir and Cockerham's unbiased differentiation index (F_{ST}) and Nei's genetic distance. The results were visualized using R software. The DAPC analysis revealed minimal genetic differences among the studied populations. Based on the findings, the studied populations from Iran, France, Sweden, and Poland formed a single genetic group with no significant differentiation. These results were further confirmed through F_{ST} and Nei's genetic distance statistic and population admixture analysis. Overall, a high genetic similarity was found between Iranian Holstein cattle and Holstein cattle populations in other countries, particularly in Europe and North America. This similarity is likely due to the extensive importation of genetic material and the absence of an effective breeding and selection program in Iran. Future studies may investigate the precise influence of these factors on the genomic diversity and population structure of the Holstein breed. The results of this study can be used to explore the structure of the Holstein population. Likewise, they can be beneficial in national breeding schemes and planning for the importation of genetic material (especially semen).

1 Introduction

Throughout the history of animal husbandry, cattle – as large ruminants – have been widely dispersed and successfully adapted to a broad range of climatic conditions. Since their domestication, cattle have become an essential source of food and labor for humans (Papachristou et al., 2020). In 2022, the global cattle population reached approximately 1.55 billion heads, up from around 1.51 billion heads in 2021 (FAO, 2023). The Holstein Friesian is the most widespread breed across more than 150 countries, forming a large metapopulation, and is currently the leading dairy breed worldwide (Vostry et al., 2023).

The Holstein Friesian is the most widespread breed across more than 150 countries, forming a large metapopulation, and is currently the leading dairy breed worldwide.

The genomic diversity of Holstein cattle is relatively low despite its large population and wide distribution. For instance, the effective population size (N_e) of Holstein populations in countries such as Australia, Canada, Denmark, Spain, Ireland, and the USA ranges between 49 and 127 (Doekes et al., 2018). The low genomic diversity in this breed can be attributed to several factors. Among the main reasons are the intense selection and overuse of a small number of top-performing bulls, particularly those with higher milk production potential (Yue et al., 2015). Additionally, the high gene flow and strong connections between subpopulations have further diminished genomic diversity within the Holstein metapopulation (Ablondi et al., 2022; Doekes et al., 2018). As a result, managing genomic diversity and inbreeding has become a key focus in many national Holstein breeding programs (Rodríguez-Ramilo et al., 2015).

Understanding the genomic diversity of domestic species is essential for the development of effective management strategies and genetic conservation (Wultsch et al., 2016). Recent advancements in genome sequencing and the availability of high-density genomic data have facilitated the analysis of genomic diversity and population structure (Decker et al., 2014). Methods such as genetic distances, population clustering, and admixture analysis are commonly used to assess genomic diversity within populations (Al-Mamun et al., 2015).

In recent years, a growing body of research has focused on the genomic diversity of various domesticated species, including cattle, buffalo, goats, and sheep. The findings of different studies (e.g., Moradi Shahrabak et al., 2023; Rahim-madar et al., 2021; Colli et al., 2018; Doekes et al., 2018; Mastrangelo et al., 2018; Kukučková et al., 2017; Rodríguez-Ramilo et al., 2015; Gautier et al., 2010) have provided valuable insights into the genetic variation within these populations. Additionally, some studies have focused on the genomic diversity of goats and sheep (e.g., Giovannini et al., 2024; Kusza et al., 2024), contributing to further understanding of the genetic composition of these domesticated animals. Despite its global prominence, the Holstein breed faces

challenges in maintaining genomic diversity due to the extensive use of semen from elite bulls and artificial insemination practices. Therefore, the present study seeks to compare the genomic diversity of Iranian Holstein cattle with populations from Poland, France, and Sweden, providing a thorough overview of the genetic structure of the Holstein breed.

2 Material and methods

2.1 Animals and source of raw data

A total of 399 Holstein cattle from four geographically distinct countries, including Iran ($n = 25$), France ($n = 153$), Poland ($n = 198$), and Sweden ($n = 23$), were meticulously selected based on the study purpose. Genomic information used in this study was obtained through reliable online databases, including Dryad and WIDDE (Sempéré et al., 2015), or through electronic correspondence with relevant researchers (Table 1).

2.2 Quality control and filtering process

Data quality control and filtering were performed using PLINK 1.9, following the protocols outlined by Purcell et al. (2007) and Toro-Ospina et al. (2022). Loci located on sex chromosomes, loci with unknown chromosomal locations, single-nucleotide polymorphism (SNP) loci with call rate genotyping (CR_{SNP}) less than 95 % and minor allele frequency (MAF) less than 5 %, and SNP markers that were out of Hardy–Weinberg equilibrium (P value less than 10^{-6}) were excluded during this process. Similarly, individuals with more than 5 % missing genotypes were also excluded from the analysis. Subsequently, data aggregation and pruning were performed, resulting in 10 822 SNP markers for genomic diversity analysis. Due to the limitations of the software for population structure analysis, missing genotypes were imputed using Beagle software (Browning et al., 2021).

2.3 Genomic diversity statistical analysis

The genomic diversity and racial groups were studied using principal component analysis (PCA) and genetic group determination methods (Elhaik, 2022; Ma and Amos, 2012). To investigate the genetic structure of the population and its constituent subpopulations, discriminant analysis of principal components (DAPC) was conducted using the adegenet package in R software (Jombart and Ahmed, 2011). Moreover, Weir and Cockerham's differentiation index (1984) and genetic distances were calculated in the R environment (Toro-Ospina et al., 2022; Purcell et al., 2007). Furthermore, the adegenet package in the R environment was used to study the populations' genetic commonality and analyze population structure in greater detail (Jombart and Ahmed, 2011). Eventually, visual representations of the analysis results were created using the R environment (R Core Team, 2024).

Table 1. Genomic data of six used Holstein populations related to four countries: Iran, France, Poland, and Sweden.

No.	Mark	Country	Sample no.	Quick access address – genomic array	Relevant article
1	IRI	Iran	25	Illumina_BovineSNP50v2	Shakeri et al. (2021)
2	FRA_LD	France	30	https://doi.org/10.1371/journal.pone.0013038 & WIDDE	Gautier et al. (2010)
3	FRA_LD	France	63	https://doi.org/10.1371/journal.pone.0005350 & WIDDE Illumina_BovineSNP50v1	Matukumalli et al. (2009)
4	FRA_HD	France	60	http://widde.toulouse.inra.fr/widde/widde/main.do; jsessionid=EB8D8557C0E6C34E3040E06D29FDA729? module=cattle (last access: January 2023) & WIDDE illumina.com/applications/agriculture.ilmn Illumina BovineHD	Project: ILLUMINA_HD
5	POL	Poland	198	Illumina_BovineSNP50v1	Requested
6	SWE	Sweden	23	–	Salehi et al. (2023b)
Total			399		

3 Results

3.1 PCA and DAPC methods

Initially, the data were subjected to PCA (Price et al., 2006), followed by discriminant analysis to identify clusters according to the method described in previous research (Jombart and Ahmed, 2011). To determine the proximity of populations and their genetic relationships, information on identity by descent and identity by state was obtained from the genomic relationship matrix (Villanueva et al., 2021; Makgahlela et al., 2013). Figures 1 and 2 display the graphical representation of the separation of the studied Holstein cattle populations using SNP markers. The results of the DAPC indicated a lack of distinct clusters and groups in both cases (Figs. 1 and 2). The study utilized information from components 1–40 to deduce genetic grouping. Interestingly, the findings revealed that increasing the number of components resulted in greater differentiation between individuals and groups. Nonetheless, slight differentiation was observed among the studied groups when using the information from the first 40 components to draw graphs related to genetic grouping. This observation is attributed to the diversity being primarily related to subpopulations within the Holstein breed.

In the analysis, the results emphasized the significance of the first component (Fig. 2), suggesting potential population mixing or a shared origin among the studied animals. The observed diversity within and between populations was found in both visual representations. Based on the results (Fig. 2), the French population exhibited the highest genomic diver-

Table 2. Genetic distances of Holstein cattle populations in Iran, France, Poland, and Sweden.

Populations	FRA	IRI	POL	SWD
FRA	0.00	0.0038	0.0030	0.0144
IRI	0.0038	0.00	0.0066	0.0156
POL	0.0030	0.0066	0.00	0.0120
SWD	0.0144	0.0156	0.0120	0.00

sity, whereas the Iranian population exhibited the lowest genomic diversity.

3.2 Genetic distance across populations

To further analyze population diversity, genetic distances were calculated, and a phylogenetic tree was constructed for the studied populations. The outcomes of these calculations are presented in Table 2 and Fig. 3. The findings revealed significant variations between the populations. Furthermore, the analysis of population differentiation using the Weir and Cockerham differentiation index and the examination of genetic distances conform to the findings of the DAPC analysis.

3.3 Structure-based allele admixture analysis

The results (Fig. 4) confirmed the genetic mixing of the studied populations, serving as a supplementary analysis for genomic diversity studies. The findings in this section reaffirm the results of previous analyses. While this analysis is primarily utilized for studying the genetic mixing of different

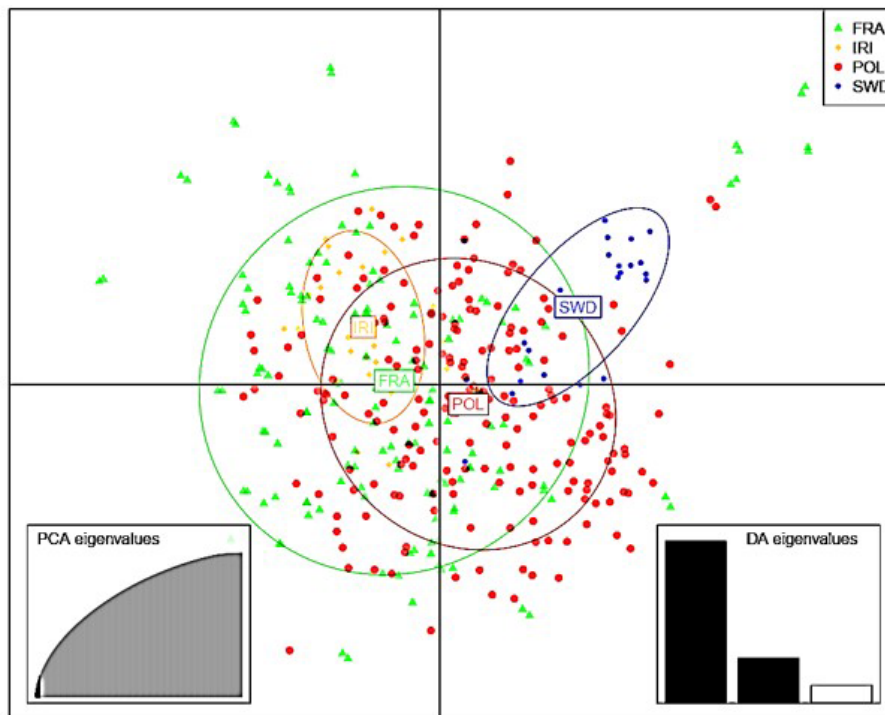


Figure 1. PCA results of discriminant analysis of principal components of Holstein cattle populations in Iran, France, Poland, and Sweden. PCA: principal component analysis; DA: discriminant analysis.

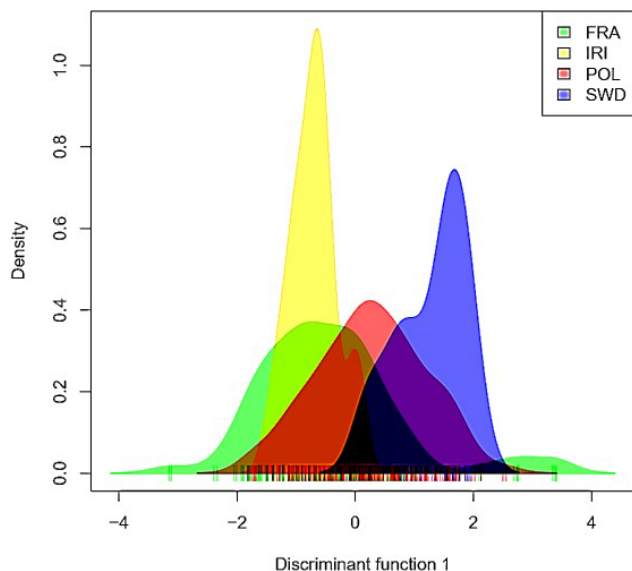


Figure 2. Advanced clustering outputs of discriminant analysis of principal components of Holstein cattle populations in Iran, France, Poland, and Sweden.

populations, it also demonstrates genomic similarities and can be applied to individuals within the same population. The genetic commonalities of each of the studied animals are separately presented, with a unique color assigned to each population to indicate the contribution of each individual's genome from the studied populations. In this analysis, a specific color, or “ K equals 4”, was designated for the number of studied populations. Assuming a K value lower than 4 would be possible to discuss similarities at the level of one or two genetic groups. However, in this instance, a value of 4 was chosen to discern the minimal differences through a more in-depth examination of the population.

The observed pattern indicates a significant genetic contribution from the French and Polish populations to Holstein cattle. Iranian Holsteins demonstrate a strong genetic affinity with the French population, followed by Polish Holsteins, with a close genetic affinity. Notably, the Swedish population exhibits the highest genetic divergence from the other three populations. It is worth mentioning that these genetic affinities reflect a shared ancestry, likely influenced by the historical movement of Holstein cattle from various countries. This genetic insight underscores the substantial impact of international exchanges in shaping the Holstein cow genome on a global scale.

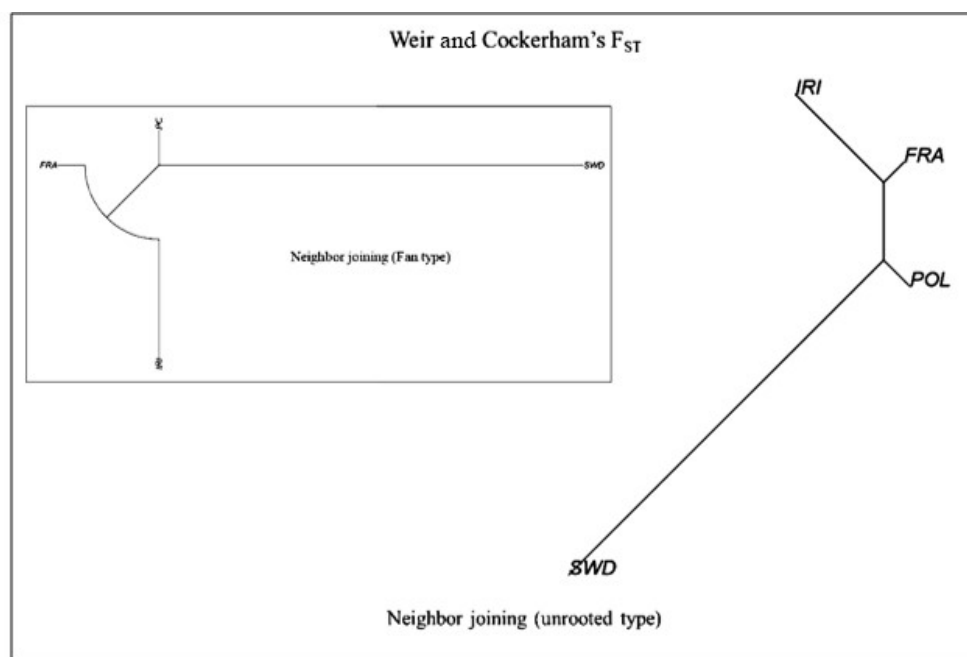


Figure 3. Phylogeny tree of Holstein cattle populations in Iran, France, Poland, and Sweden.

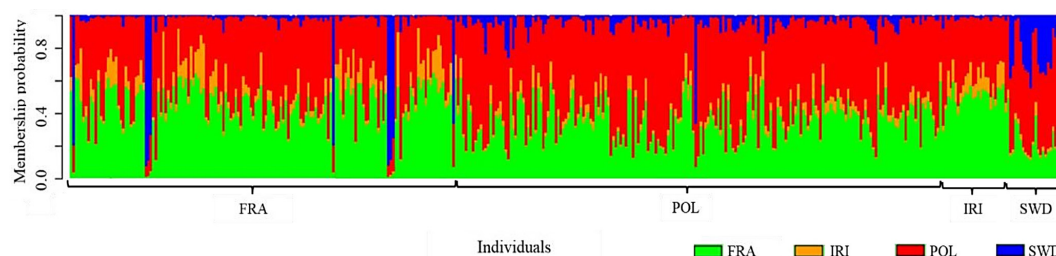


Figure 4. Graphical diagram of shared genomic diversity related to the overall appearance of the genome of each population of Holstein cattle with geographical differences.

4 Discussion

In this study, Weir and Cockerham's F_{ST} (1984) and Nei's genetic distance (1972) were selected as primary measures of genetic differentiation. These metrics are widely used for SNP-based diversity studies due to their robustness and ability to quantify population-level structure (Allendorf et al., 2013). While other metrics such as heterozygosity or allelic richness offer valuable insights, they can be more sensitive to sample size and may not fully reflect inter-population divergence (Kalinowski, 2004; Luikart et al., 2013). Our goal was to evaluate genetic differentiation across populations, for which F_{ST} and Nei's distance remain standard and well-validated tools. However, we acknowledge their limitations in capturing within-population diversity and recommend future studies to complement them with additional metrics as DAPC and population admixture analyses.

4.1 PCA and DAPC methods

Understanding the population's genetic structure is crucial for various purposes, such as identifying stratifications in livestock populations, conducting genome-wide association studies, designing breeding programs, and developing genetic resource conservation strategies. Our study used DAPC analysis, population differentiation index, genetic distance determination, and population mixing to assess the genetic structure. The DAPC method is a multivariate approach that helps identify and describe clusters of genetically related individuals. The reduced genomic diversity in the Holstein breed can be attributed to factors such as shared ancestry, rigorous selection practices, and a relatively small effective population size worldwide. The initial 10 components (Fig. 1) were utilized during the compilation of the genetic clustering analysis.

As illustrated in Fig. 1, the studied livestock populations exhibit a shared genetic origin with no clear signs of significant differentiation. The visual analysis of the generated graphs shows that intra-population diversity substantially contributes to overall diversity. Furthermore, the populations of France and Poland exhibit higher diversity than those of Iran and Sweden. The observed diversity can be attributed to the higher number and variety of samples from the respective populations.

The DAPC method has been widely employed in numerous studies (Kusza et al., 2024; Moradi Shahrehabak et al., 2023; Sharma et al., 2016) to assess genomic diversity. It is important to emphasize that the first two principal components provide a reliable measure of differentiation among populations and individuals within each population (Gaspar and Breen, 2019; Kijas et al., 2013; Jombart et al., 2010). However, leveraging additional component information enhances the accuracy and interpretability of the results. The decision to include the number of components in population differentiation analysis and genetic grouping is contingent upon the diversity of the populations under study.

Within the scope of the current study, the first two components accounted for 18 % of the variance, while the first 10 components collectively explained 25 % of the total variance. Despite justifying the same amount of variance, no significant population discrimination was recognizable. Conversely, in the investigation performed by Biabani et al. (2022) on genomic diversity between Holstein and native Iranian livestock, the studied populations were distinctly categorized, with the first two components elucidating 74.9 % of the variance. Moreover, Salehi et al. (2023a) observed justified variances of 12.87 and 5.44 for the first and second components, respectively.

In the study by Sharma et al. (2016), the justified variance of the initial 100 components equated to 52 % of the total variance.

4.2 Genetic distance across populations

Based on the genetic distance analysis results, the population of France demonstrated the lowest distance compared to the populations of Poland (0.0030) and Iran (0.0038), while the genetic distance between Poland and Iran was 0.0066. Additionally, the Swedish population exhibited genetic distances ranging from 0.0120 to 0.0156 compared to other populations, placing the French population between the Iranian and Polish populations. Notably, previous studies reported a high level of genetic compatibility between two subsets of data pertaining to the French population, namely, a high-density series (60 heads) and a low-density series (93 heads), which were considered inseparable in genetic terms (Salehi et al., 2023b). In this analysis, both datasets were collectively evaluated as a single population. Salehi et al. (2023b) further highlighted that the utilization of high-density arrays (777 962 SNPs) is unnecessary for the assessment of

genomic diversity, and genomic information obtained from lower-density arrays (the Illumina BovineSNP50 series, typically comprising ~54 000 markers) was adequate.

Genetic distance, a measure of the genetic differentiation between populations, can be quantified using the genetic differentiation index (F_{ST}). These values are categorized based on numerical intervals (Wright, 1978) into low (0–0.05), medium (0.05–0.15), high (0.15–0.25), and very high (more than 0.25) groups. They provide a classification system for reporting the genetic relationships between populations. According to Wright (1978), low F_{ST} values signify a close genetic relationship between populations. In the context of the studied Holstein populations, low F_{ST} values suggest minimal genetic differentiation, indicating a common origin and high gene flow between the populations.

In a comprehensive study on the genomic diversity of the Holstein population, Vostry et al. (2023) examined 2178 Holstein heads from 12 European and American Holstein subpopulations, along with the Simmental breed as an outgroup. Their phylogenetic analysis revealed interesting findings. They found that the Holstein populations of Belgium, England, Canada, the Netherlands, France, Germany, the Czech Republic, and the United States were grouped together, while the subpopulations of Switzerland and Ireland, as well as Croatia, were distinct from this main group. The genetic differentiation of the Simmental breed from the Holstein subpopulations was found to be in the numerical range of 0.049–0.060, while differentiation within the Holstein subpopulations showed two distinct intervals. The difference between Croatia, Ireland, and Switzerland populations compared to other Holsteins (Belgium, England, Canada, the Netherlands, France, Germany, Czech Republic, and the USA) was in the numerical range of 0.011–0.028. Moreover, the distance among the Holstein populations of Belgium, England, Canada, the Netherlands, France, Germany, Czech Republic, and the USA was in the numerical range of 0.001–0.008. The results and the research objectives of the present study are consistent with those of Vostry et al. (2023).

The above study compared the Holsteins of the Czech Republic with the Holstein populations of other countries and reported that the bulls used in artificial insemination in the Czech Republic differed from those in other countries, especially regarding genes affecting meat and carcasses. In addition, despite the common origin of the world's Holsteins, each country had adopted different selective paths, depending on its breeding goals, leading to relative differentiation of Holstein populations. This differentiation was particularly evident in the Swedish population, which was highly different from other studied populations, including Iran, France, and Poland. Additionally, the high similarity of the Iranian Holstein population with other genetic groups suggests high kinship and gene flow with the Iranian population (Vostry et al., 2023).

4.3 Structure-based allele admixture analysis

The observed pattern indicated a significant genetic contribution from the French and Polish populations to Holstein cattle. Iranian Holsteins demonstrated a strong genetic affinity with the French population, while Polish Holsteins represented a close genetic affinity. Notably, the Swedish population exhibited the greatest genetic divergence from the other three populations. Notably, these genetic affinities reflect a shared ancestry, likely influenced by the historical movement of Holstein cattle from various countries. This genetic insight underscores the substantial impact of international exchanges in shaping the Holstein cow genome on a global scale. In a separate investigation exploring the genomic diversity of Holsteins worldwide, data from China, Canada, the Netherlands, and Ireland were also included (Salehi et al., 2023b). However, due to the varying data formats for these populations, it was impossible to analyze the genetic differences alongside the data from the current study. The high differentiation of genetic groups could be attributed to the incompatible data formats with those of the present study. The division of cattle breeds globally relies on exploring the genetic differentiation between cattle breeds with *Bos taurus* and *Bos indicus* backgrounds (Ben Jemaa et al., 2015; Deckers et al., 2014), indicating that the origin or genetic background serves as a fundamental factor in the differentiation of cattle populations worldwide. Additionally, geographical distribution is another crucial factor in the racial differentiation of livestock populations. This is attributed to genetic differentiation among groups being influenced by their geographical distribution and the varying climatic conditions of their breeding regions (Berihulay et al., 2019).

The results of the present study regarding the genomic data of French Holstein were compared with those of the study by Vostry et al. (2023). This comparison could provide a better understanding of the genomic diversity of Iranian Holstein cows in relation to other European and American Holstein populations. To thoroughly discuss the genomic diversity of the Holstein population in Iran, it is important to investigate the origins of the genetic material. The low genetic distance between the French Holstein population and that of Iran in this study (0.0038) indicates a high likelihood of strong genetic relationships between Iranian Holsteins and other Holstein populations worldwide, including the USA, Ireland, Germany, the Netherlands, and Canada. This suggests that Iranian Holsteins maintain a very close genetic relationship with commercial Holstein populations worldwide. The genetic purity of Iranian Holsteins appears to be high, closely resembling the genetic profile of Holsteins from large commercial farms. Genetic resources for Iranian Holsteins are likely derived from external sources, particularly in cases where the genetic material is obtained from international sperm sources. Moreover, the low level of genetic differentiation in Iranian Holsteins indicates that targeted breeding efforts have not substantially shaped their ge-

netic structure. Geographic isolation and racial mixing may have contributed to the observed genetic differences between populations (Vostry et al., 2023; Berihulay et al., 2019).

The genomic diversity of Holstein cattle worldwide has decreased due to extensive breeding programs and selection, reducing the effective population size. Studies reported that the effective population size values for Holstein cattle in several countries ranged from 49 to 127 heads. This decrease in genomic diversity is not surprising, given the extensive exchange of genetic material and the high relatedness of this breed globally. Despite the high genetic connections, differences can be observed in the national diversity of Holstein cows due to selection programs and genetic drift in different regions of the world. For example, the genomic diversity of Danish Holstein has increased in recent years due to breeding programs. Additionally, some studies compared the genetic distances between Holstein subpopulations and other dairy and beef breeds worldwide, including indigenous cows of different regions. Karimi et al. (2016) found that the genetic differentiation of Iranian native cattle with Holstein varies from 0.129 in Kurdi to 0.332 in Sistani. Interestingly, the genetic differentiation of some native Iranian breeds with Holstein is less than the differentiation between native Iranian breeds. For example, Moradi Shahrehabak et al. (2023) reported the genetic differentiation of Sistani and Kurdish cows as 0.18 and Sistani and Sarabi as 0.166. Based on Moradi Shahrehabak et al. (2023) findings, the differences observed in the study groups of Holstein populations and native breeds of Iran were related to genetic background and geographical differentiation. Sistani cattle belong to the *Bos indicus* subspecies and have different living conditions from the *Bos taurus* breed, such as the Kurdish breed (Moradi Shahrehabak et al., 2023). Likewise, Makina et al. (2014) compared the genetic distance between Holstein and Angus with several native breeds from South Africa. The genetic distance between Holstein and Angus was 0.098, while the distance between Holstein and the native breeds of South Africa was reported to be up to 0.159.

Recent advancements in population genetics have raised important concerns about the sustainability of current Holstein breeding strategies in Iran. The widespread use of elite sires through artificial insemination has led to significant improvements in productivity. However, this strategy has also resulted in increased inbreeding and a reduction in effective population size, which in turn has caused a decline in genetic diversity – a crucial factor for long-term adaptability (Doekes et al., 2018; Vostry et al., 2023). To address these challenges, future strategies should focus on broadening the genetic base by incorporating less-related or underutilized lineages, including locally adapted bulls. Establishing a national genomic selection program and regularly monitoring diversity indices – such as inbreeding coefficients and runs of homozygosity – can help guide informed selection while preventing excessive genetic relatedness (Doekes et al., 2018). Additionally, the cryopreservation of genetically

diverse germplasm could serve as a valuable resource for future breeding flexibility and resilience (Vostry et al., 2023).

The findings from this study highlight the genetic homogenization of Iranian Holstein cattle due to the widespread use of imported semen from a narrow elite pool. While this approach has improved production traits, it has also raised concerns for long-term sustainability by reducing genetic diversity and effective population size. To mitigate these risks and preserve genetic resilience, Iran's breeding strategies should prioritize genomic monitoring, diversify sire origins, and incorporate local selection programs that are tailored to national needs (Ameri et al., 2023; Vostry et al., 2023).

5 Conclusions

Examining the genetic variation at the genomic level provides valuable insights into the differentiation between populations and individuals within a population. This information is crucial for breeders aiming to enhance livestock performance and for genetic conservation strategists. In this study, the Iranian Holstein cattle population's genomic diversity was comprehensively analyzed compared to that of the French, Polish, and Swedish Holstein populations. The findings revealed the genetic affinity of Iran's purebred Holstein cattle population with that of other countries that primarily export genetic material of this breed. The minimal difference between Iran's Holstein cattle and other global Holstein populations, particularly those in Europe and the United States, can be attributed to the substantial import of genetic material from these regions. Additionally, it reflects historical breeding practices and the lack of effective breeding programs in Iran. These results provide a meaningful foundation for understanding population structure and contribute to the refinement of breeding programs and the development of effective conservation plans. Furthermore, the findings can inform national decisions pertaining to the dairy cattle breeding industry, specifically in determining the sources of genetic material input (sperm) in the country. Future investigations are essential to confirm and expand upon the conclusions drawn from this study.

Data availability. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions. RS, AJ, and MM were responsible for the methodology and study design. MM and RS conducted the data analysis. AJ and MM prepared the original draft version of the manuscript. The manuscript was reviewed and edited by AJ, MM, and SA. The visualization (tables and figures) was prepared by RS and RS. AJ was responsible for supervision and fundraising.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Ethical statement. The study was carried out on a farm under standard rearing conditions and in accordance with Slovenian laws. The approval of an ethical review body was not required for this study.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. We appreciate the helpful feedback from the reviewers, which greatly enhanced the quality of this work.

Financial support. The research was funded by the University of Tabriz (grant no. 13/419596/1).

Review statement. This paper was edited by Henry Reyer and reviewed by Hossein Mohammadi and one anonymous referee.

References

- Ablondi, M., Sabbioni, A., Stocco, G., Cipolat-Gotet, C., Dadousis, C., van Kaam, J. T., and Summer, A.: Genomic 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>, 2022.
- Allendorf, F. W., Luikart, G., and Aitken, S. N.: *Conservation and the Genetics of Populations*, 2nd edn., Wiley-Blackwell, ISBN: 978-0-470-67145-0, 2013.
- Al-Mamun, H. A., Clark, S. A., Kwan, P., and Gondro, C.: Genome-wide linkage disequilibrium and genomic diversity in five populations of Australian domestic sheep, *Genet. Sel. Evol.*, 47, <https://doi.org/10.1186/s12711-015-0169-6>, 2015.
- Ameri, N. F., Moradian, H., Koshkoiyeh, A. E., Montazeri, M., Madabi, E. R., and Fozi, M. A.: Genetic diversity and positive signatures of selection in indigenous cattle breeds of Iran, *Genome*, 67, 31–42, 2023.
- Ben Jemaa, S., Boussaha, M., Ben Mehdi, M., Lee, J. H., and Lee, S. H.: Genome-wide insights into population structure and genetic history of Tunisian local cattle using the Illumina BovineSNP50 BeadChip, *BMC Genomics*, 16, 67, <https://doi.org/10.1186/s12864-015-1225-7>, 2015.
- Berihulay, H., Li, Y., Liu, X., Gebreselassie, G., Islam, R., Liu, W., Jiang, L., and Ma, Y.: Genetic diversity and population structure in multiple Chinese goat populations using a SNP panel, *Animal Genetics*, 50, 242–249, <https://doi.org/10.1111/age.12776>, 2019.
- Biabani, P., Yeganeh, H. M., Shahrabak, H. M., and Mokhber, M.: Detection of genetic differences between

- Holstein and Iranian north-west indigenous hybrid cattle using genomic data, *Res. Anim. Prod.*, 13, 175–186, <https://doi.org/10.52547/rap.13.37.175>, 2022.
- Browning, B. L., Tian, X., Zhou, Y., and Browning, S. R.: Fast two-stage phasing of large-scale sequence data, *Am. J. Hum. Genet.*, 108, 1880–1890, <https://doi.org/10.1016/j.ajhg.2021.08.005>, 2021.
- Colli, L., Milanesi, M., Vajana, E., Iamartino, D., Bomba, L., Puglisi, F., and Ajmone-Marsan, P.: New insights on water buffalo genomic diversity and post-domestication migration routes from medium density SNP chip data, *Front. Genet.*, 9, 53, <https://doi.org/10.3389/fgene.2018.00053>, 2018.
- Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., Regitano, L. C. A., Yildiz, M. A., Heaton, M. P., Liu, W. S., Lei, C. Z., Reecy, J. M., Saif-Ur-Rehman, M., Schnabel, R. D., and Taylor, J. F.: Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle, *PLoS Genet.*, 10, e1004254, <https://doi.org/10.1371/journal.pgen.1004254>, 2014.
- Doekes, H. P., Veerkamp, R. F., Bijma, P., Hiemstra, S. J., and Windig, J. J.: Trends in genome-wide and region-specific genomic diversity in the Dutch-Flemish Holstein-Friesian breeding program from 1986 to 2015, *Genet. Sel. Evol.*, 50, 1–16, <https://doi.org/10.1186/s12711-018-0385-y>, 2018.
- Elhaik, E.: Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated, *Scientific Reports*, 12, 14683, <https://doi.org/10.1038/s41598-022-14395-4>, 2022.
- FAO (Food and Agriculture Organization of the United Nations): FAOSTAT Statistical Database, <https://www.fao.org/faostat/en/> (last access: 15 October 2023), 2023.
- Gaspar, H. A. and Breen, G.: Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics, *BMC Bioinformatics*, 20, 116, <https://doi.org/10.1186/s12859-019-2680-1>, 2019.
- Gautier, M., Laloë, D., and Moazami-Goudarzi, K.: Insights into the Genetic History of French cattle from Dense SNP Data on 47 Worldwide Breeds, *PLoS ONE*, 5, e13038, <https://doi.org/10.1371/journal.pone.0013038>, 2010.
- Giovannini, S., Chessari, G., Riggio, S., Marletta, D., Sardina, M. T., Mastrangelo, S., and Sarti, F. M.: Insight into the current genomic diversity, conservation status and population structure of Tunisian Barbarine sheep breed, *Front. Genet.*, 15, 1379086, <https://doi.org/10.3389/fgene.2024.1379086>, 2024.
- 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 Genetics*, 11, 94, <https://doi.org/10.1186/1471-2156-11-94>, 2010.
- Kalinowski, S. T.: Counting alleles with rarefaction: private alleles and hierarchical sampling designs, *Conserv. Genet.*, 5, 539–543, 2004.
- Karimi, K., Strucken, E. M., Moghaddar, N., Ferdosi, M. H., Esmaeilzadeh, A., and Gondro, C.: Local and global patterns of admixture and population structure in Iranian native cattle, *BMC Genetics*, 17, 1–14, <https://doi.org/10.1186/s12863-016-0335-9>, 2016.
- Kijas, J. W., Ortiz, J. S., McCulloch, R., James, A., Brice, B., Swain, B., and International Goat Genome Consortium: Genomic diversity and investigation of polledness in divergent goat populations using 52,088 SNPs, *Anim. Genet.*, 44, 285–289, <https://doi.org/10.1111/age.12035>, 2013.
- Kukučková, V., Moravčíková, N., Ferenčaković, M., Simčič, M., Mészáros, G., Sölkner, J., Trakovická, A., Kadlecík, O., Curik, I., and Kasarda, R.: Genomic characterization of Pinzgau cattle: genetic conservation and breeding perspectives, *Conserv. Genet.*, 18, 893–910, <https://doi.org/10.1007/s10592-017-0935-9>, 2017.
- Kusza, S., Badaoui, B., and Wanjala, G.: Insights into the genomic homogeneity of Moroccan indigenous sheep breeds through the lens of runs of homozygosity, *Sci. Rep.*, 14, 16515, <https://doi.org/10.1038/s41598-024-67558-w>, 2024.
- Luikart, G., England, P. R., Tallmon, D., Jordan, S., and Taberlet, P.: The power and promise of population genomics: from genotyping to genome typing, *Nat. Rev. Genet.*, 4, 981–994, 2013.
- Ma, J., and Amos, C. I.: Principal components analysis of population admixture, *PLoS ONE*, 7, e40115, <https://doi.org/10.1371/journal.pone.0040115>, 2012.
- Makgahlela, M. L., Knürr, T., Aamand, G. P., Strandén, I., and Mäntysaari, E. A.: Single step evaluations using haplotype segments, *Interbull Bulletin*, 47, 217–221, 2013.
- Makina, S. O., Muchadeyi, F. C., van Marle-Köster, E., MacNeil, M. D., and Maiwashe, A.: Genomic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel, *Front. Genet.*, 5, 333, <https://doi.org/10.3389/fgene.2014.00333>, 2014.
- Mastrangelo, S., Sardina, M. T., Tolone, M., Di Gerlando, R., Suter, A. M., Fontanesi, L., and Portolano, B.: Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds, *Animal*, 12, 2480–2488, <https://doi.org/10.1017/S1751731118000629>, 2018.
- Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P. L., Sonstegard, T. S., and Van Tassell, C. P.: Development and Characterization of a High Density SNP Genotyping Assay for Cattle, *PLoS ONE*, 4, e3350, <https://doi.org/10.1371/journal.pone.0005350>, 2009.
- Moradi Shahrehabak, H., Biabani, P., Mehrbani Yeganeh, H., and Mokhber, M.: Investigating the genomic diversity of Iranian native and Holstein cattle breeds using genomic data, *Research Journal of Livestock Science*, 36, 87–98, <https://doi.org/10.22092/asj.2022.357456.2201>, 2023.
- Nei, M.: Genetic distance between populations, *Am. Nat.*, 106, 283–292, 1972.
- Papachristou, D., Koutsouli, P., Laliotis, G. P., Kunz, E., Upadhyay, M., Seichter, D., Russ, I., Götz, K.-U., and Medugorac, I.: Genomic diversity and population structure of the indigenous Greek and Cypriot cattle populations, *Genet. Sel. Evol.*, 52, 43, <https://doi.org/10.1186/s12711-020-00560-8>, 2020.
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D.: Principal components analysis corrects for stratification in genome-wide association studies, *Nat. Genet.*, 38, 904–909, <https://doi.org/10.1038/ng1847>, 2006.

- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., and Sham, P. C.: PLINK: a toolset for whole-genome association and population-based linkage analysis, *Am. J. Hum. Genet.*, 81, 559–575, <https://doi.org/10.1086/519795>, 2007.
- Rahimadad, S., Ghaffari, M., Mokhber, M., and Williams, J. L.: Linkage disequilibrium and effective population size of buffalo populations of Iran, Turkey, Pakistan, and Egypt using a medium density SNP array, *Front. Genet.*, 12, 608186, <https://doi.org/10.3389/fgene.2021.608186>, 2021.
- R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, <https://www.r-project.org/> (last access: 26 August 2024), 2024.
- Rodríguez-Ramilo, S. T., Fernández, J., Toro, M. A., Hernández, D., and Villanueva, B.: 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>, 2015.
- Salehi, R., Javanmard, A., Mokhber, M., and Alijani, S.: Analysis of genomic diversity in worldwide Holstein populations based on genomic array information, in: Proceedings of the 10th National and 2nd International Animal Science Congress of Iran, Karaj, Iran, 30–31 August 2023a.
- Salehi, R., Javanmard, A., Mokhber, M., and Alijani, S.: Genomic Selection Signatures in Two French and Swedish Holstein Cattle Breeds Provide Evidence for Several Potential Candidate Genes Linked to Economic Traits, *Iran. J. Appl. Anim. Sci.*, 13, 677–684, 2023b.
- Sempéré, G., Moazami-Goudarzi, K., Eggen, A., Laloë, D., Gautier, M., and Flori, L.: WIDDE: a Web-Interfaced next-generation database for genomic diversity exploration, with a first application in cattle, *BMC Genomics*, 16, 940, <https://doi.org/10.1186/s12864-015-2181-1>, 2015.
- Shakeri, R., Javanmard, A., Hasanpur, K., Abbasi, M., Khansefid, M., and Rahimi Varposhti, M.: Assessment of genetic diversity within Holstein population using bovine SNP chip data, *Res. Anim. Prod.*, 12, 140–149, 2021.
- Sharma, A., Lee, S. H., Lim, D., Chai, H. H., Choi, B. H., and Cho, Y.: A genome-wide assessment of genomic diversity and population structure of Korean native cattle breeds, *BMC Genet.*, 17, 1–9, <https://doi.org/10.1186/s12863-017-0472-z>, 2016.
- Toro-Ospina, A. M., Herrera Rios, A. C., Pimenta Schettini, G., Vallejo Aristizabal, V. H., Bizarria dos Santos, W., Zapata, C. A., and Ortiz Morea, E. G.: Identification of runs of homozygosity islands and genomic estimated inbreeding values in Caqueteño Creole cattle (Colombia), *Genes*, 13, 1232, <https://doi.org/10.3390/genes13071232>, 2022.
- Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E., Toro, M. A., and Pong-Wong, R.: The value of genomic relationship matrices to estimate levels of inbreeding, *Genet. Sel. Evol.*, 53, 42, <https://doi.org/10.1186/s12711-021-00635-0>, 2021.
- Vostry, L., Vostra-Vydrova, H., Moravcikova, N., Kasarda, R., Cubric-Curik, V., Brzakova, M., Sölkner, J., Shihabi, M., Hidalgo Moreno, J. A., and Spehar, M.: Genomic diversity and population structure of the Czech Holstein cattle, *Livest. Sci.*, 273, 105261, <https://doi.org/10.1016/j.livsci.2023.105261>, 2023.
- Weir, B. S. and Cockerham, C. C.: Estimating F-statistics for the analysis of population structure, *Evolution*, 38, 1358–1370, <https://doi.org/10.2307/2408641>, 1984.
- Wright, S.: Evolution and the genetics of populations, in: Variability Within and Among Natural Populations, Vol. 4, University of Chicago Press, Chicago, IL, USA, ISBN: 978-0-226-91052-9, 1978.
- Wultsch, C., Caragiulo, A., Dias-Freedman, I., Quigley, H., Rabbinowitz, S., and Amato, G.: Genetic diversity and population structure of Mesoamerican jaguars (*Panthera onca*): implications for conservation and management, *PLoS ONE*, 11, e0162377, <https://doi.org/10.1371/journal.pone.0162377>, 2016.
- Yue, X. P., Dechow, C., and Liu, W. S.: 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>, 2015.