Evaluation of single nucleotide polymorphisms identified through the use of SNP assay in Romanian and Chinese Holstein and Simmental cattle breeds

Daniela Elena Ilie1,2✉, Yahui Gao3, Ioana Nicolae2, Dongxiao Sun3✉, Junya Li4, Bo Han3 and Dinu Gavojdian2

Simmental and Holstein cattle, being among the most widely distributed breeds worldwide, have been subjected to continuous selection for distinct purposes. In the current study, we evaluated the levels of SNPs identified through the use of SNP assay in Romanian Holstein and Romanian Simmental cattle, which were then compared to the data from Chinese Holstein and Chinese Simmental cattle. In total, 282 animals were genotyped: Romanian Holstein (n=30), Romanian Simmental (n=22), Chinese Holstein (n=96) and Chinese Simmental cattle (n=136), using 39,724 common SNPs to analyze minor allele frequency, genetic variability and level of SNPs. Among studied breeds, the average percentage of polymorphic markers was 90.84%, with the highest value in Chinese Simmental (91.37%) and lowest in Romanian Simmental cattle (90.31%). The average H0 ranged from 0.426 in Romanian Holstein to 0.416 in Romanian Simmental, and from 0.425 in Chinese Holstein to 0.422 in Chinese Simmental. The distribution of SNPs was homogenous across the breeds, except the Romanian Simmental which displayed the lowest percentage of polymorphic markers (24.66 and 32.48%) from higher MAF category (0.3 to <0.4 and 0.4 to <0.5) and the highest percentage (3.82 and 12.00%) for SNPs from low and intermediate MAF categories (0.05 to <0.1 and 0.1 to <0.2). In the current study, the SNP assay was successfully used to analyze the level of SNP sites of Romanian cattle breeds, however, a higher number of samples and production data are needed for future applications of the results in genomic selection, genome-wide association studies and genetic diversity analysis.

Key words: Chinese cattle; Romanian cattle; Holstein; Simmental; single nucleotide polymorphisms.

Received: 29 November, 2020; revised: 24 April, 2020; accepted: 15 July, 2020; available on-line: 17 September, 2020

INTRODUCTION

It is estimated that about 3017 breeds of cattle exist globally (Bos taurus, Bos indicus, Bibos hanteng and Bibos frontalis). These unique breeds were selected for hundreds of years in a wide range of environments following domestication. Over the last century, more than 1,900 of these breeds were declared extinct or classified at risk of extinction (FAO-DAD-IS, 2019). Currently, FAO estimated the worldwide cattle number as 1.47 billion heads (FAOSTAT, 2016), with 84.5 million cattle being found in China and over 90 million in the European Union (FAO, 2018). Thus, in both areas, the cattle sector is of significant importance due to its economic and social roles.

Romania is among the top ten countries in the EU regarding the cattle population with 1,977,200 heads (FAO, 2019) and two predominant breeds: Romanian Simmental/Fleckvieh (representing approximately 29% of the breed structure, national name Bălțăță Românească) and Romanian Black and White/Holstein-Friesian (representing approximately 22% of the breed structure, national name Bălțăță cu Negru Românească). The Simmental breed is widespread in all European countries and due to its special adaptability is the most widely distributed of all breeds of cattle worldwide. In addition, the Holstein-Friesian cattle have been continuously selected so that the breed is now the dairy breed with the highest milk production in the world (Shin, 2017). Both breeds have been subjected to artificial selection for distinct purposes (Chen et al., 2016).

As stated by the Domestic Animal Diversity Information System (FAO-DAD-IS, 2019), China has over 70 cattle breeds, including 53 indigenous cattle breeds (Zhang et al., 2018; Wang et al., 2018) that are classified mainly into two categories: Bos taurus and Bos indicus (Xia et al., 2019). Geographically, all Chinese cattle breeds are divided into three categories, the northern group classified as humpless breeds and distributed in Northern China (Bos taurus), the southern group classified as humped breeds and spread in Southern China (Bos indicus) and the central category found in the middle and lower areas of the Yellow River and the Huaihe River (influenced by both Bos taurus and Bos indicus) (Wang et al., 2018; Xia et al., 2019). However, a small number of cattle breeds, such as Holstein, Simmental, Hereford and Aberdeen Angus are predominant worldwide (Wang et al., 2018).

Application of SNP arrays in animal genetic selection has gained remarkable attention in recent years. Thereby,
the accessibility of numerous SNPs distributed across the genome proved to be useful for genetic analysis and selection in farm animals. In bovine, genotyping using SNP array has become a common practice in developed countries, for both dairy and beef cattle breeding programs applying genomic selection. The assay was used in genetic disease mapping (Charlier et al., 2008; VanRaden et al., 2011; Murugiano et al., 2014), genomic selection for economic traits (Neves et al., 2014; Garcia-Ruiz et al., 2016; Taylor et al., 2016; Brown et al., 2016; Boisson et al., 2017) and genome-wide association studies (Bolorma et al., 2011; Olsen et al., 2011; Wu et al., 2013; Lee et al., 2013; Streit et al., 2013; Tiezzi et al., 2015). Since the bovine genome sequencing, several SNP arrays from Illumina, Affymetrix and Neogen/GeneSeek were developed and are currently available for cattle (Khatkar et al., 2010; Kasarda et al., 2014), such as lower-density SNP panels (3K, 7K, 15K, 25K), medium (50K) up to high-density SNP panel (150K, 250K, 650K, 800K). This availability of high numbers of SNPs resulted in new research opportunities and thereby the bovine SNP arrays were used widely in a variety of studies to improve selection in cattle breeding programs. In Chinese cattle breeds, several studies on genome-wide SNP marker development and utilization were reported. However, to our knowledge, no previous study has been reported on the evaluation of single nucleotide polymorphisms identified by the SNP assay in Romanian cattle breeds.

The aim of the current study was to evaluate the single nucleotide polymorphisms identified by the SNP assay in Romanian Holstein and Romanian Simmental cattle and to compare the results to the data from the Chinese Holstein and Chinese Simmental cattle in order to analyze the level of SNPs for future applications in genomic selection, genome-wide association studies and genetic diversity analysis in Romanian cattle. Given the extensive use of high genetic merit bulls imported from North America and Germany to both China and Romania, we hypothesized that strong breed relationships and overlapping genetic makeup must exist, thus, future genetic exchanges would be feasible without disrupting the selection programs.

MATERIALS AND METHODS

Ethics statement. The experimental design, sampling collection protocols and procedures used in this study were reviewed and approved by each institution. All research activities were performed in accordance with the European Union’s Directive for animal experimentation (Directive 2010/63/UE).

Table 1. Breed, origin, sample size (n), common SNPs and SNPs remained after quality control and genetic diversity indexes ($H_e$ and $H_t$) in the four studied cattle breeds

| Breed | Origin | Sample size | Common SNPs | SNPs remained after QC | $P_{pol}$ (MAF) | $H_e$ ($E$) | $H_t$ ($O$) |
|-------|--------|-------------|-------------|------------------------|-----------------|-----------|-----------|
| HOL-CHN | China | 96 | 39724 | 36481 (91.11) | 0.425 | 0.418 |
| HOL-RO | Romania | 30 | 39724 | 36306 (90.58) | 0.426 | 0.405 |
| SIM-CHN | China | 136 | 39724 | 46568 (91.37) | 0.422 | 0.423 |
| SIM-RO | Romania | 22 | 39724 | 36214 (90.31) | 0.416 | 0.398 |

$P_{pol}$: polymorphic markers; $H_e$ and $H_t$: expected and observed heterozygosity

Animals and sample collection. A total of 282 individuals (Table 1) from four cattle breeds were used in the study, as follows: Romanian Holstein (HOL-RO, n=30), Romanian Simmental (SIM-RO, n=22), Chinese Holstein (HOL-CHN, n=96) and Chinese Simmental cattle (SIM-CHN, n=36). The Romanian Holstein and Romanian Simmental cattle samples were collected from China Agricultural University and the Chinese Academy of Agricultural Sciences, Beijing Institute of Animal Sciences.

Samples were randomly collected from cows, avoiding highly related animals, according to the pedigree information available. After collection, all samples were transferred to the laboratory and were kept at 4°C until further processing. Genomic DNA was extracted from both blood and hair follicles.

Bovine SNPs genotyping. For Romanian cattle DNA samples, the Geneseeek Genomic Profiler (GGP) Bovine 50K (GeneSeek Genomic Profiler, Neogen Corp., Lincoln, USA) which includes 47843 SNPs was used for genotyping. The GGP Bovine 50K covers a large percentage of SNPs overlapping with other commercially available arrays, including the original Illumina Bovine SNP50k. The GGP Bovine 50K contains more than 44000 SNPs which overlap with the Illumina Bovine HD array. The BeadChip technology includes SNPs specifically chosen for high minor allele frequency (MAF) values, with an average between 0.2826 and 0.3598 across all loci in different cattle breeds and uniform genome coverage for the majority of the Bos taurus and several of the Bos indicus breeds. The reference bovine genome build for the GGP Bovine 50K was UMD3.1 (bosTau6) (Zimin et al., 2009).

For the Chinese cattle breeds, all 232 animals were genotyped using Illumina BovineHD Genotyping BeadChip which includes 777962 SNPs evenly distributed across the genome. Over 99% of the SNPs of BovineHD BeadChip were mapped to the UMD3.1 bovine genome assembly (Zimin et al., 2009) which includes coverage of autosomal, mitochondrial, and sex-linked (X/Y) SNPs.

The average minor allele frequency across all loci on the 770K BovineHD BeadChip is 0.25. Marker positions and chromosomes in the map for each panel refer to the UMD3.1 assembly (Zimin et al., 2009). After excluding...
non-shared markers, a total of 43984 SNPs, common to all breeds, were retained for subsequent analyses.

**Quality control and data analysis.** Data quality control (QC) of SNPs and DNA samples was performed separately for each panel (50K and 770K) using PLINK software (Purcell et al., 2007). The quality control filtering of SNPs to exclude low-quality markers was applied based on variables to remove SNPs with insufficient genotyping quality. Minor allele frequency (MAF) was computed based on the frequency of the least common allele for every SNP in the given population. We removed SNPs with call rates <95% or with MAF <0.05, samples with more than 10% missing genotypes and SNPs with genotypes not in Hardy–Weinberg equilibrium (P>10^-6). SNPs with unknown positions and those located on the x sex chromosome were not considered in the current study. Only SNPs located on autosomes were further analyzed. Additional quality control parameter investigated was the GenTrain score, which is a measurement of SNP calling quality that ranges from 0 to 1, with higher value meaning better quality. Over the loci, the average GenTrain score was 0.740.

Due to the fact that in the present study two different genotype data sets consisting solely of the SNPs that were shared between the panels. The genotypes from both 50K and 770K panels were merged using PLINK software (Purcell et al., 2007) and we retained only the common SNPs. Finally, a data set of 39,724 SNPs across the entire bovine genome was common for both BeadChips and was subjected to further analysis.

The MAF for each autosome and the overall mean MAF in each of the four studied cattle breeds were calculated by using PLINK (Purcell et al., 2007). The distribution of MAF was grouped into five classes as follows: (0.0 to <0.1), (0.1 to <0.2), (0.2 to <0.3), (0.3 to <0.4) and (0.4 to 0.5). PLINK (Purcell et al., 2007) was used to calculate the observed heterozygosity (H_e) and expected heterozygosity (H_o) for each cattle breed. The H_o within breed was calculated based on SNPs which passed the quality control and compared to the H_e under Hardy-Weinberg equilibrium. Principal component analysis (PCA) was used to describe the genetic structure of the cattle breeds using GCTA software and the neighbor-joining tree was built using MEGA 10. The R (R Core Team) was used for processing SNP loci and generating the chromosome ideograms.

**RESULTS AND DISCUSSION**

Evaluation of single nucleotide polymorphisms covered 39,724 SNPs (Table 1) across the entire bovine genome after eliminating the non-shared SNPs (between GGP Bovine 50K and 770K BovineHD BeadChip) and after quality control which excluded SNPs with call rates less than 0.95 or SNPs significantly deviated from the Hardy-Weinberg equilibrium (P>10^-6). The average call rate for individual samples was greater than 95% in all studied cattle breeds which is in accordance with previously published results published in studies that used a call rate of ≥0.90 (Cooper et al., 2013; McClure et al., 2018).

Genetic variability parameters for the studied breeds were presented in Table 1. In Romanian cattle breeds, after analyzing a dataset consisting of 39,724 SNPs, the average H_o was 0.426 in the Romanian Holstein and 0.416 in the Romanian Simmental, with a maximum value of 0.890 (BovineHD0700014141, chr. 7) in the Romanian Simmental and 0.9 (BovineHD1900001632, chr19; BovineHD3000019302, chr. 30) for the Romanian Holstein. Similarly, in Chinese cattle, when using the same dataset of SNPs, the overall mean H_o was 0.425 in Chinese Holstein and 0.422 in the Chinese Simmental, with a maximum value of 1 (ARS-BFGL-NGS-19773, chr. 14) in Chinese Holstein and 0.6618 (BTB-00701972, chr.18; Hapmap41591-BTA-57970, chr. 25) in Chinese Simmental. Romanian Holstein was relatively more diverse (H_o 0.426) compared to the rest of the studied breeds. Its H_o values were higher than those reported previously for twenty Chinese indigenous cattle breeds, which ranged from 0.275 to 0.374 (Zhang et al., 2018).

The mean value for expected heterozygosity across the breeds was 0.411 with Romanian Simmental cattle exhibiting the lowest H_e value (0.398) and Chinese Simmental cattle the highest (0.423). The obtained H_e values were similar to those observed in five indigenous Ethiopian cattle populations (0.370-0.410) by Edea et al. (2012) and higher than the values reported previously by Canas-Alvarez and others (Canas-Alvarez et al., 2015) in Spanish cattle breeds (0.307). However, analysis of the genetic diversity using the average H_o and H_e showed a higher level of genetic diversity (H_o=0.422, H_e=0.411) in all investigated cattle breeds compared to the previously published results, where the average values of H_o and H_e across polymorphic loci in Slovak Pinzgauer cattle population were 0.375 and 0.362, respectively (Kasar-da et al., 2014).

The level of SNPs observed in the studied cattle breeds was summarized in Table 1. On average, 90.84% of the SNP markers were polymorphic (MAF≥0.05) and their percentage detected in each breed varied from 91.37% in Chinese Simmental to 90.31% in Romanian Simmental. In order to observe the degree of polymorphism across the breeds, we computed the average proportion of SNPs for different ranges of MAF (Fig. 1). A relatively higher proportion of SNPs (MAF=0.05) were observed in the Chinese breeds (91.11% and 91.37% in Chinese Holstein and Chinese Simmental, respectively) compared to the Romanian breeds (90.58% and 90.31% in Romanian Holstein and Romanian Simmental, respectively). The lowest number of rare SNPs (MAF 0.0 to <0.1) after filtering was found in Chinese Holstein and Romanian Simmental, respectively. The number of rare SNPs (MAF 0.0 to <0.1) after filtering was found in Chinese Holstein, and Chinese Simmental (90.58% and 90.31% in Romanian Holstein and Romanian Simmental, respectively). The lowest number of rare SNPs (MAF 0.0 to <0.1) after filtering was found in Chinese Holstein and Romanian Simmental (3.68%), while Romanian Simmental showed the highest proportion (7.57%). The proportion of polymorphic markers with high MAF values (0.4–0.5) ranged from 38.15% in Chinese Holstein to 32.48% in Romanian Simmental.

The distribution of SNPs on autosomal chromosomes was homogenous across the breeds, except the Romanian Simmental which displayed the lowest percentage...
of SNPs (24.66 and 32.48%) from high-MAF categories (0.3 to <0.4 and 0.4 to <0.5) and the highest (3.82 and 12.00%) for SNPs from rare and intermediate MAF categories (0.05 to <0.1 and 0.1 to <0.2). Among the filtered SNPs, a number of 729, 909, 338 and 1487 SNPs were fixed or close to fixed and showed MAF values lower than 0.05 in HOL-RO, HOL-CHN, SIM-RO and SIM-CHN, respectively. The fixed SNPs accounted for a small number of all cases in all studied breeds with the highest proportion in the Romanian Simmental (0.94%). Several monomorphic markers were overlapping among breeds. As the MAF values range was comparable among the studied breeds, our results might suggest similar variability of the breeds, which was also confirmed by the similar expected heterozygosity across the breeds. The degree of polymorphism observed in the present study was similar to that reported by Li and others (Li et al., 2019) for Chinese cattle breeds (89.9%) and by Cañas-Alvarez and others (Cañas-Alvarez et al., 2015) for Spanish breeds (86–98%), however, it was higher than previously reported for Korean cattle, where the proportion of SNPs with MAF>0.05 ranged between 72.7% and 76.16% (Kim et al., 2018). On the other hand, according to Zhang and others (Zhang et al., 2018), the percentage of SNPs observed in twenty Chinese indigenous cattle breeds was quite varied: between 72.9% and 97.2%. In addition, other authors found a varying degree of polymorphism among cattle populations, ranging from 62 to 95.21% (Edea et al., 2013; Ramey et al., 2013; Pertoldi et al., 2014; Porto-Neto et al. 2014). However, in a larger study that analyzed several breeds, covering 14 dairy and beef breeds (including Simmental and Holstein), three breeds of predominantly Bos taurus indicus background, two breeds that were Bos taurus × Bos taurus indicus composites and two African breeds, the authors observed that ~95% of investigated markers were polymorphic (Matukumalli et al., 2009).

The chromosome-wise information on SNPs in the four studied cattle breeds was presented in Table 2 and Fig. 2. In this study, the data set of 39,724 SNPs

### Table 2. Chromosome-wise numbers of polymorphic filtered markers on autosomes in the four studied cattle breeds

| Chr | Length (Mbp) | HOL-CHN | HOL-RO | SIM-CHN | SIM-RO |
|-----|--------------|---------|--------|---------|--------|
| 1   | 158.337      | 2252    | 2237   | 2256    | 2237   |
| 2   | 137.060      | 1937    | 1922   | 1931    | 1921   |
| 3   | 121.430      | 1693    | 1682   | 1693    | 1681   |
| 4   | 120.829      | 1757    | 1748   | 1768    | 1756   |
| 5   | 119.458      | 1781    | 1766   | 1785    | 1768   |
| 6   | 112.638      | 1577    | 1571   | 1583    | 1567   |
| 7   | 113.384      | 1592    | 1584   | 1599    | 1583   |
| 8   | 105.708      | 1595    | 1588   | 1597    | 1579   |
| 9   | 104.305      | 1526    | 1511   | 1528    | 1519   |
| 10  | 107.310      | 1526    | 1514   | 1529    | 1522   |
| 11  | 91.163       | 1222    | 1216   | 1225    | 1211   |
| 12  | 84.240       | 1307    | 1303   | 1310    | 1296   |
| 13  | 84.648       | 1276    | 1265   | 1276    | 1256   |
| 14  | 85.296       | 1222    | 1211   | 1220    | 1217   |
| 15  | 81.724       | 1120    | 1114   | 1128    | 1117   |
| 16  | 75.158       | 1090    | 1086   | 1094    | 1090   |
| 17  | 66.004       | 1016    | 1013   | 1016    | 1005   |
| 18  | 64.057       | 1011    | 1008   | 1013    | 1005   |
| 19  | 72.042       | 1134    | 1127   | 1137    | 1122   |
| 20  | 71.599       | 1009    | 1006   | 1009    | 1000   |
| 21  | 61.435       | 903     | 901    | 903     | 879    |
| 22  | 52.530       | 810     | 808    | 810     | 806    |
| 23  | 62.714       | 940     | 935    | 942     | 930    |
| 24  | 42.904       | 630     | 627    | 630     | 626    |
| 25  | 51.681       | 758     | 754    | 762     | 752    |
| 26  | 45.407       | 670     | 671    | 671     | 668    |
| 27  | 51.505       | 699     | 698    | 699     | 693    |
| 28  | 51.505       | 699     | 698    | 699     | 693    |
| 29  | 51.505       | 699     | 698    | 699     | 693    |
| 30  | 51.505       | 699     | 698    | 699     | 693    |
| 31  | 51.505       | 699     | 698    | 699     | 693    |
| Total | 2511.997   | 36481   | 36306  | 36568   | 36214  |

The highest and lowest values are in bold.

![Figure 2. The number of polymorphic markers on each chromosome in the studied cattle breeds.](image-url)
The population structure and genetic relationships among the four cattle breeds revealed by PCA and neighbor-joining tree were shown in Fig. 3 and Fig. 4. The PCA was used to explore the clustering of individuals from different populations. The analysis allowed the visualization of groups formed by individuals belonging to the same breed (Fig. 3). In the PCA analyses, after merging the Romanian cattle breeds and the Chinese cattle breeds, the Chinese Simmental population clustered together with the Romanian Simmental population, and the Chinese Holstein population overlapped with the Romanian Holstein population. Thereby, clearly separated clusters were observed for Simmental population when compared to Holstein population, which localized in clearly different areas of the plot. In order to assess the population structure, a neighbor-joining tree was also computed (Fig. 4). The architecture of the neighbor-joining tree showed a clear separation between Holstein and Simmental breeds. Individuals belonging to the Chinese Holstein and Romanian Holstein breeds were grouped together into a single cluster leading to a clear separation from Chinese Simmental and Romanian Simmental breeds that were grouped in another cluster. However, in a study aimed to identify novel selection signals in the two Chinese cattle populations (Holstein and Simmental) the PCA analyses showed that all Chinese Holstein individuals were assigned to the European Holstein population and breed assignment analyses confirmed that the Chinese Holstein and Simmental populations originated from Europe (Chen et al., 2016).

CONCLUSIONS

The results showed the level of single nucleotide polymorphisms and genetic variability among the four investigated cattle breeds from Romania and China, using 39,724 SNP markers. On average, 90.84% of the markers were polymorphic and the distribution shape of the polymorphic markers on autosomes was homogenous across all breeds. Our results indicate a high level of genetic diversity (\(H_o=0.422\), \(H_e=0.411\)) in all investigated populations. Both neighbor-joining tree and PCA analysis showed well-defined clusters between Holstein and Simmental breeds. The current study represents a first step for the implementation of future genomic selection schemes in Romanian dairy and dual-purpose breeds and in genome-wide association and genetic diversity studies in both Romanian and Chinese cattle. As hypothesized, results on genetic diversity showed that both Romanian and Chinese Holstein breeds are clustering strongly together and overlap, as a result of using extensively imported American Holstein genetics in both countries. The Simmental breeds, both Romanian and Chinese, were overlapping less than the Holsteins, indicating stronger differences in their genetic makeup, which is attributed to a wider set of selection indicators. In most European and Asian countries, the Simmental breeds are being selected for growth rates and carcass attributes, milk production and fitness-related indicators, unlike Holsteins which are selected almost exclusively for milk yields and to a lesser extent for reproductive efficiency.

REFERENCES

Boison S, Utsunomiya ATH, Santos DJA, Neves HHR3, Carvalheiro R, Mészáros G, Utsunomiya YT, do Carmo AS, Verneque RS, Machado MA, Panetto JCC, Gereza JF, Sölkner J, da Silva MVGB (2017) Accuracy of genomic predictions in Gyr (Bos indicus) dairy cattle. J Dairy Sci 100: 1–12. https://doi.org/10.3168/jds.2016-11811
