

Genetic characteristics and diversity of Korean Jeju Black cattle by whole genome SNP markers

Mohammad Zahangir Alam

Shahjalal University of Science and Technology

Hyo Jung Son

Yeungnam University

Yun Mi Lee

Yeungnam University

Lauren Hulsman Hanna

North Dakota State University

David Riley

Texas A&M University

Hideyuki Mannen

Kobe University

Shinji Sasazaki

Kobe university

Se Pill Park

Jeju National University

Jong Joo Kim (✉ kim_jong_joo@hotmail.com)

Yeungnam University <https://orcid.org/0000-0002-7730-8420>

Research article

Keywords: Jeju Black cattle, Hanwoo, genetic diversity, population structure, SNP

Posted Date: January 16th, 2020

DOI: <https://doi.org/10.21203/rs.2.21018/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Conservation and genetic improvement of cattle breeds requires to know the information about genetic diversity and population structure of animals. This study investigated the genetic diversity and population structure among the three breeds raised in Korean peninsula. Three popular breed found in Korea, i.e. Jeju Black, Hanwoo, Holstein with other six breeds such as Angus, Hereford, Brown Wagyu, Black Wagyu, Brahman and Nellore was examined in this study. Genetic diversity within the cattle breeds was analyzed using the popular measures of genetic diversity namely minor allele frequency (MAF), observed and expected heterozygosity (H_O and H_E), inbreeding coefficient (F_{IS}) and past effective population size (N_E). Molecular variance and population structure were performed among the nine cattle breeds using model-based clustering (ADMIXTURE) analysis. Genetic distances between breed pairs were evaluated using Nei's genetic distance (D_A) and with Weir and Cockerham's F_{ST} .

Results

This study revealed that Jeju Black cattle had lowest level of heterozygosity ($H_E = 0.21$) among the studied taurine cattle breeds ranging from 0.25 to 0.30, and low MAF of 0.16, while other breeds have MAF ranging 0.11~0.21. The level of inbreeding was -0.076 in case of Jeju Black as compared to other breed (-0.018 ~ -0.118). PCA analysis and neighbor-joining (NJ) tree showed a clear separation of Jeju Black cattle from other local and exotic cattle breeds. Model-based clustering also revealed a distinct pattern of admixture of Jeju Black cattle having no clustering with other studied populations. The F_{ST} value between Jeju Black cattle and Hanwoo was 0.106, which was lowest across the breeds ranging from 0.161 to 0.274, indicating some degree of genetic closeness of Jeju Black cattle with Hanwoo. The N_E of Jeju Black cattle was 38, whereas Hanwoo was 209 in the most recent 13 generation ago.

Conclusion

This study indicates an alarming trend of reducing effective population size in Jeju Black cattle. Thus, a sustainable breeding policy should be implemented to increase the population of Jeju Black cattle for the genetic improvement and future conservation.

Background

Cattle is an integral part of animal agriculture since 8000 BC, when it was thought to be domesticated in different parts of the world such as India, Middle East and North Africa [1]. Different cattle breeds are domesticated and adapted throughout the world due to variable geographical and climatic condition. Jeju Black Cattle (Jeju Heugu) is one of the indigenous cattle breeds found in the Jeju Island of Korean peninsula (Republic of Korea). Jeju Black cattle thought to be originated from the native Hanwoo cattle of mainland Korea according to island model of speciation [2]. Ancient cattle bones from the archaeological sites in Jeju Island suggesting the existence of the breed approximately 1100~2000 years ago. DNA analysis of bones recovered from Gonaeri and Gwakji-ri in Aewaleup, Jeju city discovered that ancestors of the present day Jeju Black cattle had been raised by human since prehistoric age [3]. Other ancient documents (the *Annals of the Joseon Dynasty*) and paintings found in the mural (Anak Tomb no.3, during Goguryeo Dynasty in 357 AD) also suggesting its existence. However, several reports from researchers showed their origin in a controversial way [4]. Whatever their origin, Jeju Black Cattle has been categorized as an endangered species due to a substantial shrinkage in its population size until 1980s.

This breed is well adapted in the subtropical environment of the Jeju Island. Beef from Jeju Black cattle have some distinct features such as high marbling and rich in oleic acid, linoleic acid, and unsaturated fatty acids which make it premium quality. Highly marbling beef has a great demand in Korean cuisine and culture like Japanese Wagyu beef (kobe beef). As the Korean economy become healthy on the basis of per capita GDP, the demand of premium quality beef is increasing day by day. But in the past decades, this indigenous breed paid little attention due to its impaired growth performances and priorities of rearing other local beef breed 'Hanwoo' by the government and farmers. However, in the recent decades South Korean government took initiative to conserve and genetic improvement of the breed. Genetic diversity of any given species is essential to conserve nature and future genetic improvement. Conservation and genetic improvement requires careful study of genetic characteristics and population structure of any species. Many useful parameter such as allelic richness (A_R), level of inbreeding (F_{IS}), effective population size (N_E) and genetic distance with other local and exotic breeds was studied along with the existing Jeju Black cattle population.

Genomic studies using high throughput whole genome sequencing data have become popular in the recent years. Although the cost of genotyping reduced but still it is not very cheap. Variation of the gene for a particular trait of interest is the raw material for animal breeder. If there is no genetic variation, improvement is not possible. Single nucleotide polymorphisms (SNPs) is one of the common types of genetic variant for any organism. However, genotyping using SNP microarray chip provides genomic data with an efficient and cost-effective manner. Many useful genetic parameters such as linkage disequilibrium (LD), effective population size (N_E), Inbreeding coefficient, levels of heterozygosity etc. can be estimated from SNP data. Commercial and custom-made Microarray SNP chips are available from low density to high density panels in different livestock species. However, the challenges are to calculation of various estimates from the SNP data with different statistical approach. Linkage disequilibrium (LD) which can be described as non-random association of alleles at two or more loci [5] is a powerful tool in genetics, evolutionary and conservation biology. Measuring LD might be useful in rare breeds like Jeju Black cattle where it can be used to calculate population genetic parameters in the absence of pedigree data. It is reported that registered Jeju Black cattle have a population size approximately 619 (Korea Seed stock Database), other sources mentioned their number might be 400 ~500 [6, 7]. Due to the small size of the remaining population of Jeju Black cattle only in the Jeju Island, it is essential to estimate the level of inbreeding as it is an important parameter to assess the genetic diversity of a given species. Inbreeding depression, a phenomena derived from the loss of fitness and reproductive performance in inbred population have deleterious effect in many cases. Inbreeding coefficient can be calculated both from pedigree and genomic data but in general, pedigree data information gives lower value than those obtained with genomic data. However, better accuracies of inbreeding coefficient (F_{SNP}) could be estimated from the genome wide SNP data [8]. Effective population size (N_E) is one of the important genetic parameters that used to determine the amount of genetic variation, genetic drift, and linkage disequilibrium (LD) in both cattle and human population [9, 10]. Marker based approach of N_E estimation gained popularity due to the availability of large amounts of genetic marker data derived from advanced DNA microarray chip technology.

Development of sustainable breed improvement strategies is dependent on the precise characterization of animal genetic diversity [11]. Although several studies have investigated the diversity pattern of Korean cattle along with Jeju Black Cattle [2, 6, 12] but it is still controversial whether Jeju Black formed as a separate breed or as a varieties of Hanwoo. Accurate refinement to decipher the origin of Jeju Black cattle is necessary for their future conservation and improvement program. Genetic diversity study using microsatellite-based marker often reported larger genetic differentiation values than SNP-based marker [13, 14] which is not desirable. Moreover SNPs markers using admixture analysis gives more accurate estimates than pedigree analysis [15]. In this study, we used high density Illumina SNP50K v.3 BeadChip to generate SNP data in order to explore genetic characteristics of Jeju Black cattle and compared with other eight cattle breeds. We chose Hanwoo as it is the main cattle breed of Korea, with two *Bos indicus* breed namely Brahman and Nellore, and other five *Bos taurus* breed namely Angus, Brown and Black Wagyu (Japanese breed), Hereford and Holstein.

Results

Within breed genetic diversity

Table 1 presents three measures of within breed diversity across the studied population. The minor allele frequency across breed was in the range from 0.11 (Nellore) to 0.21 (Hanwoo and Angus). The MAF of Jeju Black cattle was estimated 0.16 which represents the value in between of studied breeds. However, the Nellore cattle observed to have the lowest level of expected heterozygosity ($H_E=0.15$) than Hanwoo ($H_E = 0.28$) which had the highest level of genetic diversity in the studied breeds. F_{IS} which measures the non-random mating (inbreeding) and found less inbreeding as F_{IS} values were negative for all the breeds ranging from -0.018 in Black Wagyu to -0.118 in Brown Wagyu cattle. Jeju Black cattle had F_{IS} value -0.076 which is lower than Hanwoo (-0.025).

Analysis of molecular variance and population differentiation

Livestock biodiversity can be estimated from the level of genetic variation amongst breeds. Variation in SNPs allele frequencies between breeds can be used to measure the degree of genetic differentiation (F_{ST}). F_{ST} is the measure of the genetic differentiation among individual within population to total and deviation from the Hardy-Weinberg proportions within population. The average Wright's F-statistics (table 2) estimated with 20000 bootstraps over loci were: $F_{ST} = 0.173$, $F_{IS} = -0.030$ and $F_{IT} = 0.148$. Whereas, F_{IT} measures the genetic differentiation within individuals of total population. The value for quantification of F_{ST} ranges from 0.00 ~ 0.25 which indicate, low to very high differences as the degree of genetic relatedness between breeds. This study indicates that there is

significant deviation ($P < 0.01$) between breeds. AMOVA indicated that almost 17 % of the variation was estimated for variation among the populations, while 85% of the variation was accounted due to within individual variation (table 3). Genetic differentiation between nine cattle breed was estimated using pairwise F_{ST} which showed in figure 1. The F_{ST} ranged from 0.085 (Hanwoo and Brown Wagyu) to 0.376 (Nellore and Brown Wagyu). The genetic distances, D_A between Hereford and Nellore (0.107) was relatively higher as compared to JJBC and Hanwoo (0.028) in the studied population. Jeju Black cattle had the highest distance with Nellore (0.084) and Brahman (0.078) and lowest distance with Hanwoo (0.028) and Black Wagyu (0.042) cattle breeds. The Nei's genetic distance D_A matrix of nine cattle breeds was used to construct phylogenetic trees with Neighbor- Joining (NJ) method [16]. Topological relationships between breeds, from NJ tree clearly separated *Taurine* and *Indicine* cattle into two groups. But the NJ trees shows a distinct pattern of phylogenetic trees as Hanwoo, Brown Wagyu and Black Wagyu breeds clustered together which might be due to their genetically closeness to each other as depicted in the figure 2. JJBC grouped separately in NJ trees.

Principal component analysis

PCA reduces the dimension of the number of possible correlated variables (SNPs) into a fewer number of uncorrelated variables (principal components), while retaining most of the variability in the genotypic dataset. The first and second principal components (PC) accounted for 27.87% and 24.89% of the total variation respectively (table 5). The third and fourth principal components accounted for 19.00% and 16.73% of the variation, respectively. The first five PC accounted about 100% variation, across all studied cattle population. Jeju Black cattle showed closeness to Hanwoo than other cattle breeds raised in Korean peninsula. In our analysis, Nellore and Brahman formed a distinct cluster which might be due to large variation among *taurine* and *indicine* cattle. Hanwoo cattle formed closer cluster with Brown Wagyu and Black Wagyu rather than Holstein breeds raised in Korea. Japanese Wagyu breeds had a distinct cluster which showed closeness to Korean Hanwoo. Hanwoo lied in between Jeju Black cattle and Wagyu. Holstein raised in Korea had formed a different cluster with other European *taurine* breed Angus and Hereford.

Population Structure analysis between eight cattle breeds

The proportion of individuals in each of the breeds in our cattle breeds inferred by the ADMIXTURE are presented in table 6. In the current study, nine cattle population was tested, so we expected lowest cross validation error values when $K=9$ but lowest CV error estimator was found when $K=11$ (figure 4). Thus $K=11$ was taken as the most probable number of inferred populations.

Effective population size over the past generations

As effective population size (N_E) estimation is necessary to determine the accuracy of genomic selection [17]. We studied N_E in nine cattle breeds showed in figure 6 and table 7 at t generation ago. The size of N_e differed between populations. Figure 6 illustrate the trends of N_E from 13 ~ 913 generations. All nine cattle breeds show a marked decrease over time as expected. N_e estimates of Hanwoo cattle in the most recent 13 generation ago was 209 which was highest value across nine cattle breeds studied. Jeju Black cattle has an effective population size of 38 at 13 generation ago. Other cattle breeds Angus, Brahman, Nellore, Hereford, Brown Wagyu, Black Wagyu and Holstein had a N_E of 78, 37, 39, 51, 29, 93 and 91 respectively in the studied population at most recent 13 generations ago. In the more distant past at approximately 847 generations ago, the effective population was estimated to approximately 1168 and 2678 for the Jeju Black and Hanwoo cattle population respectively.

Linkage Disequilibrium

Extent of LD, measured as r^2 within autosomal SNPs markers were estimated in the nine cattle breeds were presented in the table 8. Different cattle breeds showed different pattern of LD at different inter-SNP distance (table 9) as breed is one of the factor influence LD estimates. The average r^2 value for Jeju Black cattle is 0.71 at 0 to 1 kb distance bin, whereas other breeds Hanwoo, Brown Wagyu, Black Wagyu, Brahman, Nellore, Angus, Hereford and Korean Holstein has the r^2 value is 0.63, 0.65, 0.54, 0.55, 0.58, 0.71, 0.63 and 0.67 respectively. This study showed that LD was highest for SNPs pairs within close proximity, i.e. 0 to 1 kb interval distance. Then mean r^2 value tended to decrease rapidly with increasing distance between pair of markers in all population. The most rapid decline being seen over the first 40 kb. The estimates showed almost no change from 600 to 1000 kb of marker distance. The average r^2

value across the cattle breeds in the studied population ranges from 0.13 to 0.63. Brahman and Nellore showed almost similar pattern of LD decay between the markers with respect to physical distance in studied population.

Discussion

Genetic characterization of animal on the basis of genomic data has become an attractive method to animal geneticists and biotechnologist due to easy access of high throughput data derived from microarray SNP chip technology. In genetic diversity analysis, SNP markers have many advantages over traditional microsatellite markers due to higher level of resolution, despite a set of microsatellites are being suggested by the FAO to assess genetic diversity of farm animals and endangered species [11, 18]. In the most recent years, genetic characterization using SNP markers have been studied in a variety of cattle breeds such as Irish Carry cattle [19], Tyrol Grey [20], Spanish beef cattle breeds [21], Canchim [22], Chinese Yiling yellow cattle [23] and many other indigenous and exotic cattle breeds raised in different countries worldwide [24–30]. Among Korean cattle breed, Hanwoo was given much more research interest due to its incorporation in the national breeding program since 1970s [2]. In this study, we emphasized on the genetic characterization of Jeju Black cattle in addition to other cattle breed Hanwoo, Japanese Wagyu, and Holstein raised in Korea to make a better and accurate comparison with other breeds.

Genetic diversity of cattle breeds can be estimated by various indices such as allelic richness (A_R), Heterozygosity level i.e., expected heterozygosity vs observed heterozygosity (H_E vs H_O) and Inbreeding level (F_{IS}). In our study all cattle breeds having a slightly higher observed heterozygosity label than expected. Jeju Black cattle showed lower level of genetic variability ($H_E = 0.21$) than Hanwoo and Holstein in Korea, both demonstrated heterozygosity value of $H_E = 0.28$. Sharma in her reports [12, 31] demonstrated different level of heterozygosity in JJBC ($H_E = 0.39$ and 0.25) while other researcher, Eva [2] reported $H_E = 0.29$. Heterozygosity level in our study is very close to the reported level of Sharma [12, 31]. However, different results might be due to the use of various genotyping platform, markers density and quality control criteria [2]. The lower level of heterozygosity in JJBC than Hanwoo and Holstein raised in Korea could be due to small population size in island, or few breeding males having chance to increase inbreeding. However, Makina [26] states that allele frequencies might be a poor estimate of inbreeding, thus to observe real status of inbreeding assessment should be done every five years to determine any unfavorable changes in inbreeding levels. Inbreeding level (F_{IS}) in JJBC was estimated to be -0.076 which are higher than other Korean breed Hanwoo (-0.025) and Holstein (-0.026) raised in Korea. Genetic variability in *B. indicus* breed was lowest in Nellore ($H_E = 0.15$) and Brahman ($H_E = 0.17$) breed in the studied population. Indicine breeds might have less genetic variability than taurine breed as observed by Lin [32].

Analysis of molecular variance also reveals the partitioning of genetic variation such as overall fixation indices (F_{ST}), within population inbreeding (F_{IS}) and total inbreeding (F_{IT}). Combining all nine breeds demonstrated that 85% of total genetic variation was within populations. This was lower than the within populations genetic variation observed in South African cattle populations (92%) [26] but higher than those reported for Iranian cattle (82.88%) [28] and Ethiopian cattle populations (83.96%) [24]. The overall F_{IS} value was negative (-0.03) and not significant ($P > 0.05$) probably because of less inbreeding level within populations. But total inbreeding estimate (F_{IT}) and estimate of population differentiation (F_{ST}) was 0.148 and 0.173 respectively, which showed significant ($P < 0.001$) in the studied population.

Pairwise F_{ST} was estimated to study the population differentiation in Korean cattle along with other breeds. JJBC and Hanwoo were found to be least differentiated in our study ($F_{ST} = 0.106$) compared to other breeds. Sharma [31] calculated F_{ST} values for Korean cattle breeds ranging from 0.02 to 0.06 and JJBC was found most differentiated with another Korean cattle breed Brindle Hanwoo (Chikso). Eva [2] estimated F_{ST} value 0.024 for JJBC to Hanwoo, 0.038 for JJBC to Chikso and 0.023 for Hanwoo to Chikso. Both studies confirmed that JJBC was least differentiated with Hanwoo in agreement with our study. Another Japanese breed Black Wagyu showed less differentiation with Hanwoo (0.085) than JJBC (0.164), might be a closer relation with Hanwoo. On the basis of F_{ST} value, we are in agreement with the concept that Hanwoo and Japanese Wagyu cattle breeds were genetically much closer to each other than indicine and European taurine breeds. PCA analysis (fig. 3A) also showed that Hanwoo and Wagyu breeds are much closer than JJBC. Unsupervised hierarchal clustering of our data implemented with ADMIXTURE software estimates general patterns of admixture and genetic relationship between cattle breeds. This revealed that 95% of Hanwoo breed were assigned to cluster ten whereas 39%, 38%, 23% and 1% of JJBC were assigned to different cluster five, one, six and ten respectively. This means that JJBC shared its genome only with Hanwoo cattle. On the other hand, rest of the genome (3%) of Hanwoo breed was observed to cluster in one, five and six. Japanese Brown Wagyu (100%) stands alone in cluster three but Black Wagyu (96%) clustered in eleven with 4% if its

genome assigned to cluster three, nine and ten. Angus (97%) were assigned to cluster nine with 2% of its genome assigned to cluster eight and ten. Brahman and Nellore only share 1% of their genome whereas Hereford and Angus were assigned to cluster seven and nine with 1~2 % of its genome assigned to these cluster. Besides this, 95% Holstein raised in Korea were assigned to cluster eight with 4% of its genome assigned to cluster in nine and ten. Figure 5 depicted that Brahman and Nellore populations showed the lowest level of admixture in the present study while JJBC showed highest level of admixture. JJBC shared few genetic links (1%) with the Hanwoo cattle, which might be due to co-ancestry regarding the origin of these two breeds. Admixture analysis clearly revealed that Korean and Japanese taurine cattle different from European taurine from our analysis.

Genetic relationships among nine cattle breeds revealed by phylogenetic tree, constructed based on Nei's genetic distance (D_A) showed in figure 2. JJBC were located at distinct branch whereas Hanwoo and Wagyu have some similar phenotype grouped together. Brahman and Nellore were indicine breed and shared some alleles thus grouped in a separate branch. The result of PCA also provided similar representation of the relationships among the studied cattle population. PCA results illustrated that Brahman and Nellore were in separate group and Hanwoo stands in between JJBC and Wagyu. Angus and Hereford showed closer association than Holstein raised in Korea but all European taurine grouped together. Jeju Black cattle population stands as separate breed depicted through PCA analysis.

Jeju Black cattle have an effective population size ($N_E = 38$) in the nearest 13 generation ago, whereas other studies performed by Sharma [12] showed their number to be 67. Sudrajad [6] calculated N_E in JJBC to be 60 until 11 generation ago and Eva [2] reported their number seems to be 11 at nearest generation. Other Korean cattle Hanwoo had a large N_E value of 209 among studied cattle breeds but till the differences is higher those reported by Li and Kim [9] who estimated N_E of Hanwoo was 630 at most recent 11 generation ago and Sudrajad [6] estimated N_E is about 531. Differences in the various reports might be caused by many factors such as sample size, SNP quality control measures and model used to study LD and N_E [6]. However, in our study the sample size of JJBC was 78 which was highest among all previous reports [2, 6, 12]. The N_E value of JJBC seems to be sufficient for maintaining genetic diversity for the short term as suggested by Frankham [33] but not enough for long term species management. Hanwoo had the lowest LD value (0.03) among all population, while JJBC had the lowest LD value of 0.15. The lowest value of LD in Hanwoo might be due to the artificial selection and breeding program initiated by the Korean government with twenty Korea proven bulls (KPN) throughout the country. Edea [34] and Sharma [12] noticed similar patterns in Korean cattle. The extent of LD also varied within autosomes among cattle breeds (table 9). The average r^2 values between adjacent syntenic markers were (0.15~0.27) for JJBC, Hanwoo (0.03~0.13), Brown Wagyu (0.19~0.26), Black Wagyu (0.07~0.23), Angus (0.08~0.23), Hereford (0.12~0.30) and for Holstein (0.07~0.23). Sudrajad [6] estimated higher value of LD for adjacent SNPs in JJBC which ranges from 0.10 to 0.33. Among indicine cattle, r^2 values between adjacent syntenic markers was found 0.16 ~ 0.26 for Brahman and 0.15 ~ 0.27 for Nellore breed. Distinct LD and N_E values were identified in the studied cattle breeds which reflect historical events and recent selection through the artificial breeding.

Conclusions

In conclusion, the present study confirms that a significant amount of genetic variation is still retained in the Korean cattle especially in JJBC population. Hence, we can speculate and suggested Jeju Black cattle as a separate breed. Although a genetic contribution in terms of allele sharing exists in JJBC with Hanwoo cattle. However, further in-depth study with whole genome resequencing and scanning using high density markers in a large population will help us to accurate measure of genetic parameters and establishment of JJBC as a separate breed.

Methods

Animals and genotypes

A total of 373 animals from nine breeds were chosen to study which including two indicine (Brahman = 15), (Nellore = 26), seven taurine (Angus = 27), (Holstein = 76), (Hereford = 25), (Hanwoo = 66), (Jeju Black = 78) and (Brown Wagyu = 10 and Black Wagyu = 50). Among the animals, Hanwoo DNA was extracted either from commercial AI bull semen straws or tail hair samples obtained from different farmers with the permission of the owners and Jeju Black cattle DNA was prepared either from AI bull semen straw or tail hair in Jeju Island, Korea and Holstein DNA was also collected from semen straw provided by Nunghyup Dairy cattle improvement

center. All other cattle DNA was provided by Texas A & M University, USA except Japanese Wagyu breed. No ethics statement was required for the collection of DNA samples from the Brahman, Nellore, Angus and Hereford cattle because DNA samples were provided by the two authors in USA under their rules and regulations. Japanese Wagyu breeds data was provided by the author in Japan which are previously used and published by other study. Genomic DNA purification and Genotyping was accomplished by DNALink, a commercial genome analysis service provider in Korea.

SNP genotyping and assembly of data sets

All of the breeds except Wagyu were genotyped using SNP chip by DNALink which includes common SNP from Bovine 50K v.3 BeadChip (Illumina, San Diego, CA, USA). Japanese Wagyu cattle breeds were also genotyped using Illumina bovine 50K SNP chip and data was provided by Kobe University, Japan. All the genotyped data was then merged and common SNPs on autosomal chromosomes were selected which resulted 45,526 SNPs in the final dataset across the breed.

Quality control and filtering of SNPs

SNP quality control and filtering based on recorded SNP genotypes were performed across nine cattle breeds to remove SNPs from further analysis if any SNPs with less than 95% call rate, and animals (sample) with less than 95% call rate. This left about 41,186 SNPs in 372 animals across the breeds according to Plink software program [35]. One Holstein breed failed to pass the QC criteria. SNP quality filtering was performed using PLINK version 1.9 and SNPs that were in high LD were pruned using the following parameter, -indep pair wise command 50 5 2 (SNP window size: 50, SNPs shifted per step: 5, r^2 threshold: 2). Pruning of SNPs that are in high LD have been shown to counter the effect of ascertainment bias and to generate meaningful comparison between breeds [36]. After pruning, a total of 18,524 SNPs was remained for analysis.

Estimates of within breed genetic diversity

Genetic diversity within cattle breeds can be estimated by three measures such as expected heterozygosity (H_E), observed heterozygosity (H_O) and inbreeding coefficient (F_{IS}). All of these parameter was calculated using R software package divRsity v1.9.9 [37].

Analysis of molecular variance (AMOVA) and population differentiation

Analysis of molecular variance to determine the partition of genetic diversity was performed in cattle breeds. The genetic variability among nine cattle breeds was estimated using the program ARLEQUIN v3.5 [38]. Population differentiation was calculated by pairwise F_{ST} estimates according to Weir and Cockerham's, 1984 [39] approach using R package divRsity v1.9.9.

Population structure Analysis

Population structure of the studied cattle breed was also carried out using the software ADMIXTURE v1.3. [40]. It is a very useful and popular tool to analyze SNP data. It performs an unsupervised clustering of large numbers of samples, and allows each individual cattle breed to be a mixture of clusters. ADMIXTURE uses a model-based estimation of individual ancestry for a range of prior values of K defined by the user. In order to elicit the true number of genetic populations (clusters or K) between nine cattle breeds, a cross validation (CV) approach was used to determine the most likely number of populations (K) in the SNP data. The best possible number of ancestral population (K) was inferred through 3 to 11 pre assumed populations. For each tested value of K , ADMIXTURE estimates the proportion of each individual's genotype derived from each cluster. A preferable value of K will show low cross validation error compared with other K values.

Principal component analysis (PCA)

Principal component analysis (PCA) was carried out to infer the relationship among the nine cattle populations by using PLINK v1.9. PCA determine breed relationships based directly on allele frequencies by using a multivariate method, which condenses the information from a large number of alleles and loci into a few synthetic variables known as principal components [34].

Genetic distance

Phylogenic analysis on the basis of SNP data has become an important tool for studying the evolutionary history of any organism. Many statistical methods and software developed for these purposes. In this study, we construct phylogenic tree by calculating Nei's Genetic distances (D_A) [16] using Poptree 2 [41] program. Neighbor-Joining (NJ) method applied for measuring Nei's genetic distance, D_A which is defined by

$$D_A = 1 - \frac{1}{r} \sum_j^r \sum_i^{m_j} \sqrt{x_{ij} y_{ij}}$$

where x_{ij} and y_{ij} are the frequencies of the i -th allele at the j -th locus in populations X and Y, respectively, m_j is the number of alleles at the j -th locus, and r is the number of loci used.

Linkage disequilibrium (LD)

Various statistical approach were practiced to measure the extent of LD. Among them, Lewontin's D' [42] and Hill's r^2 [43] were widely used to measure the extent of LD, although their functions are different. But the range of both measures is between 0 and 1. LD estimator, r^2 is preferred for association studies for its robustness, simplicity and not sensitivity to changing gene frequency and effective population size. On the other hand, estimates of D' are sensitive to allele frequency, especially when one allele is rare, and is inflated for small sample sizes [44]. Therefore, we focused on r^2 measure for further characterization of LD. The r^2 estimator represents the squared correlation coefficient (r) between two variables (alleles) at two separate SNP marker loci [45] and $r^2 = 1$ when only two haplotypes are present, which is usually a consequence of genetic drift or population bottlenecks. PLINK software [35] was used for estimation of r^2 parameter for post quality control SNPs with known genomic location of all autosomal chromosomes. Here, r^2 can be expressed by the following equation:

$$r^2(P_a, P_b, P_{ab}) = \left(\frac{(P_{ab} - P_a P_b)^2}{P_a(1-P_a)P_b(1-P_b)} \right)$$

Where, P_{ab} represents the frequency of haplotypes consisting of allele a at the first locus and allele b at the second locus [17]. Pairwise r^2 values were calculated for each chromosome in each breed, as well as across breed, with the $-r^2$ command, using the default settings. In this study, we used PLINK ' $-ld-window$ and $-ld-window-r^2$ commands' values, so that correlations between all possible SNPs were tested and that there was no r^2 threshold, in order to encapsulate all possible linkage interaction between SNPs per chromosome. To display the decay of LD, distances of pair-wise SNPs were binned into twenty types of intervals (0 to 1 kb, 10 kb intervals starting from 1 up to 100 kb and 100 kb interval starting from 100 up to 1 Mb). For each chromosome, r^2 values were then sorted by inter-SNP distance, and averaged across the afore-mentioned intervals to observe possible r^2 patterns for increasing inter-SNP distances.

Effective population size (N_E)

Effective population size (N_E) was estimated using the software tool, SNeP v1.1 by Barbato [46]. SNeP estimates N_E from linkage disequilibrium (LD) data, using the following formula suggested by Corbin [47],

$$N_{T(t)} = (4f(C_t))^{-1} (E[r_{adj}^2 | c_t]^{-1} - \alpha)$$

Where $N_{T(t)}$ represents the past effective population size estimated t generations ago, c_t represents the recombination rate t generations ago in the past for specific physical distance between markers calculated by the SNeP tool using default values, r_{adj}^2 represents the linkage disequilibrium (LD) value adjusted for sample size and $\alpha = \{1, 2, 2.2\}$ is the correction for the occurrence of mutation represents as constant. The recombination rate was calculated using the following equation suggested by Sved (Sved, 1971) [48],

$$f(c) = c[(1 - c/2)/(1 - 2)^2]$$

PLINK input files for quality filtered, autosomal SNP data sets were used for N_E calculation. Minimum and maximum inter-SNP distances of 0 and 1000kb were used respectively. The data sets for each sub-population, as well as the merged dataset were grouped into 20 distance bins of 10 ~ 100kb each. N_E estimates were subsequently calculated from the r^2 values obtained for the average distance of each distance bin.

Abbreviations

AG: Angus; **A_R:** Allelic richness; **AI:** Artificial insemination; **BM:** Brahman; **BRWG:** Brown Wagyu; **BWG:** Black Wagyu; **CV:** Coefficient of variation; **D_A:** Genetic distance; **DNA:** Deoxyribonucleic acid; **F_{IS}:** Inbreeding coefficients; **GDP:** Gross domestic product; **HF:** Hereford; **H_O and H_E:** Observed and expected heterozygosity; **HST:** Holstein; **HW:** Hanwoo; **JJBC:** Jeju Black Cattle; **LD:** Linkage disequilibrium; **MAF:** Minor allele frequency; **N_E:** Effective population size; **NJ:** Neighbor-joining; **NL:** Nellore; **PC:** Principal components; **PCA:** Principal components analysis; **SNPs:** Single-nucleotide polymorphisms

Declarations

Acknowledgements

We acknowledge the Nonghyup dairy cattle improvement center, Texas A&M University, North Dakota State University, USA and Kobe University, Japan for providing us sample and data in this analysis. We also appreciate the funder of this research.

Author' contributions

JJK conceived, design the experiment and revised the manuscript critically for important intellectual content. HJS carried out the experiments, laboratory analyses, statistical analyses and interpretation of the data, YML assisted statistical analysis and contributed as part of the laboratory team. MZA wrote the manuscript, equally contributed to the data analysis with HJS. All authors read and approved the final manuscript. All other authors including LHH, DR, HM and SPP contributed by providing DNA sample, carefully read and approved the final manuscript.

Funding

This research was financially supported by a grant (715003-07) from the Research Center for Production Management and Technical Development for High Quality Livestock Products through Agriculture, Food and Rural Affairs Research Center Program, Ministry of Agriculture, Food and Rural Affairs, Korea. The funding authorities had no roles in the design of the research nor publication of this manuscript.

Availability of data

The datasets used and/or analyzed during the current study are available via Figshare (DOI: <https://doi.org/10.6084/m9.figshare.9923873>).

Ethics approval and consent to participate

Not applicable as the Hanwoo DNA was extracted either from commercial bull semen straws or from tail hair samples obtained from different farmers with the permission of the owners. Both the hair and semen samples were collected following routine veterinary procedures and not explicitly for the purpose of this study. Holsteins DNA were prepared from commercial semen straw provided by Nunghyup Dairy cattle improvement center. Jeju Black cattle DNA were also extracted either from semen straw and tail hair provided by Jeju National University. No ethics statement was required for the collection of DNA samples from the Brahman, Nellore, Angus and Hereford cattle because DNA samples were provided by the two author in USA under their rules and regulations. Japanese Wagyu breeds data was provided by the author in Japan which were previously used and published by other study.

Consent for publication

Not applicable

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details

¹ School of Biotechnology, Yeungnam University, 280 Daehak-ro, Gyeongsan, Gyeongbuk 712-749, Republic of Korea.

² Department of Genetic Engineering and Biotechnology, Shahjalal University of Science and Technology, Sylhet-3114, Bangladesh.

³ Department of Animal Sciences, North Dakota State University, Fargo, North Dakota, USA.

⁴ Department of Animal Sciences, Texas A&M University, College Station, Texas, USA.

⁵ Graduate School of Agricultural Science, Kobe University, Kobe, Japan.

⁶ Faculty of Biotechnology, Jeju National University, Jeju, Republic of Korea.

References

1. Oldenbroek K WL. Textbook Animal Breeding and Genetics. Centre for Genetic Resources and Animal Breeding and Genomics Group, Wageningen University and Research Centre, the Netherlands; 2014.
2. Strucken EM, Lee SH, Jang GW, Porto-Neto LR, Gondro C. Towards breed formation by island model divergence in Korean cattle Phylogenetics and phylogeography. *BMC Evol Biol.* 2015;15:1–9. doi:10.1186/s12862-015-0563-2.
3. Kim JH, Oh JH, Song JH, Jeon JT, Han SH, Jung YH, et al. Molecular genetic analysis of ancient cattle bones excavated from archaeological sites in Jeju, Korea. *Mol Cells.* 2005;20:325–30.
4. Lee S-H, Park B-H, Sharma A, Dang C-G, Lee S-S, Choi T-J, et al. Hanwoo cattle: origin, domestication, breeding strategies and genomic selection. *J Anim Sci Technol.* 2014;56:2.
5. Qanbari S, Hansen M, Weigend S, Preisinger R, Simianer H. Linkage disequilibrium reveals different demographic history in egg laying chickens. *BMC Genet.* 2010;11.
6. Sudrajad P, Seo DW, Choi TJ, Park BH, Roh SH, Jung WY, et al. Genome-wide linkage disequilibrium and past effective population size in three Korean cattle breeds. *Anim Genet.* 2017;48:85–9.
7. Kim EY, Song DH, Park MJ, Park HY, Lee SE, Choi HY, et al. Post-death cloning of endangered jeju black cattle (Korean Native Cattle): Fertility and serum chemistry in a cloned bull and cow and their offspring. *J Reprod Dev.* 2013;59:536–43.
8. Zanella R, Peixoto JO, Cardoso FF, Cardoso LL, Biegelmeier P, Cantão ME, et al. Genetic diversity analysis of two commercial breeds of pigs using genomic and pedigree data. *Genet Sel Evol.* 2016;48:1–10.
9. Li Y, Kim JJ. Effective population size and signatures of selection using bovine 50K SNP chips in Korean native cattle (Hanwoo). *Evol Bioinforma.* 2015;11:143–53.

10. Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, et al. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 2007;17:520–6.
11. Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, et al. Genetic diversity in farm animals - A review. *Anim Genet.* 2010;41 SUPPL. 1:6–31.
12. Sharma A, Lim D, Chai H-H, Choi B-H, Cho Y. Demographic Trends in Korean Native Cattle Explained Using Bovine SNP50 Beadchip. *Genomics Inform.* 2016;14:230.
13. Ciani E, Cecchi F, Castellana E, D'Andrea M, Incoronato C, D'Angelo F, et al. Poorer resolution of low-density SNP vs. STR markers in reconstructing genetic relationships among seven Italian sheep breeds. *Large Anim Rev.* 2013;19:236–41.
14. Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, et al. Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics.* 2017;18:1–15. doi:10.1186/s12864-016-3459-7.
15. Frkonja A, Gredler B, Schnyder U, Curik I, Sölkner J. Prediction of breed composition in an admixed cattle population. *Anim Genet.* 2012;43:696–703.
16. Tateno Y, Nei M, Tajima F. Accuracy of estimated phylogenetic trees from molecular data - I. Distantly Related Species. *J Mol Evol.* 1982;18:387–404.
17. Hill WG RA. Linkage disequilibrium in finite populations. *Theor Appl Genet.* 1968;38:226–31.
18. Vargas J, Landi V, Martínez A, Gómez M, Camacho ME, Álvarez LÁ, et al. Molecular study of the amazonian Macabea cattle history. *PLoS One.* 2016;11:1–16.
19. Browett S, McHugo G, Richardson IW, Magee DA, Park SDE, Fahey AG, et al. Genomic characterisation of the indigenous Irish Kerry cattle breed. *Front Genet.* 2018;9 FEB:1–17.
20. Mészáros G, Boison SA, Pérez O'Brien AM, Ferenčaković M, Curik I, Da Silva MVB, et al. Genomic analysis for managing small and endangered populations: A case study in Tyrol Grey cattle. *Front Genet.* 2015;6 MAY:1–12.
21. Cañas-Álvarez JJ, González-Rodríguez A, Munilla S, Varona L, Díaz C, Baro JA, et al. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip. *J Anim Sci.* 2015;93:5164–74.
22. Buzanskas ME, Ventura RV, Chud TCS, Bernardes PA, De Abreu Santos DJ, De Almeida Regitano LC, et al. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. *PLoS One.* 2017;12:1–16.
23. Ling X, ZHANG W gang, LI J ya, ZHU D jiang, XU X cheng, TIAN Y zi, et al. Genetic background analysis and breed evaluation of Yiling yellow cattle. *J Integr Agric.* 2017;16:2246–56. doi:10.1016/S2095-3119(17)61679-4.
24. Edea Z, Bhuiyan MSA, Dessie T, Rothschild MF, Dadi H, Kim KS. Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. *Animal.* 2015;9:218–26.
25. Eusebi PG, Cortés O, Dunner S, Cañón J. Genomic diversity and population structure of Mexican and Spanish bovine Lidia breed. *Anim Genet.* 2017;48:682–5.
26. Makina SO, Muchadeyi FC, van Marle-Köster E, MacNeil MD, Maiwashe A. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front Genet.* 2014;5 SEP:1–7.
27. Msalya G, Kim ES, Laisser ELK, Kipanyula MJ, Karimuribo ED, Kusiluka LJM, et al. Determination of genetic structure and signatures of selection in three strains of Tanzania Shorthorn Zebu, Boran and Friesian cattle by genome-wide SNP analyses. *PLoS One.* 2017;12:1–18.
28. Karimi K, Strucken EM, Moghaddar N, Ferdosi MH, Esmailizadeh A, Gondro C. Local and global patterns of admixture and population structure in Iranian native cattle. *BMC Genet.* 2016;17:1–14. doi:10.1186/s12863-016-0416-z.
29. Kelleher MM, Berry DP, Kearney JF, McParland S, Buckley F, Purfield DC. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. *Animal.* 2017;11:15–23.
30. Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, et al. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. *Genet Sel Evol.* 2017;49:1–13.
31. Sharma A, Lee SH, Lim D, Chai HH, Choi BH, Cho Y. A genome-wide assessment of genetic diversity and population structure of Korean native cattle breeds. *BMC Genet.* 2016;17:1–9.
32. Lin BZ, Sasazaki S, Mannen H. Genetic diversity and structure in *bos taurus* and *bos indicus* populations analyzed by snp markers. *Anim Sci J.* 2010;81:281–9.

33. Frankham R, Ballou JD B DA. Introduction to conservation biology. New York: Cambridge University Press; 2002.
34. Edea Z, Dadi H, Kim SW, Dessie T, Lee T, Kim H, et al. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean hanwoo breeds using SNP markers. *Front Genet.* 2013;4 MAR:1–9.
35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
36. Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, et al. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One.* 2009;4.
37. Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA. DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol Evol.* 2013;4:782–8.
38. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10:564–7.
39. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Heredity (Edinb).* 1984;38:1358–70.
40. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64.
41. Takezaki N, Nei M, Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Mol Biol Evol.* 2010;27:747–52.
42. C. LR. The interaction of selection and linkage. I. general considerations; heterotic models . *Genetics.* 1964;49–67.
43. Hill WG. Estimation of linkage disequilibrium in randomly mating populations. *Heredity (Edinb).* 1974;33:229–39.
44. Visser C, Lashmar SF, Van Marle-Köster E, Poli MA, Allain D. Genetic diversity and population structure in South African, French and Argentinian Angora goats from genome-wide SNP data. *PLoS One.* 2016;11:1–15.
45. VanLiere JM, Rosenberg NA. Mathematical properties of the r^2 measure of linkage disequilibrium. *Theor Popul Biol.* 2008;74:130–7.
46. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet.* 2015;6 MAR:1–6.
47. Corbin LJ, Liu AYH, Bishop SC, Woolliams JA. Estimation of historical effective population size using linkage disequilibria with marker data. *J Anim Breed Genet.* 2012;129:257–70.
48. Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol.* 1971;2:125–41.

Tables

Table 1 Population size and genetic diversity within 9 cattle breeds

Breed	N	MAF(SD)	H _O (SD)	H _E (SD)	F _{IS} (SD)
Angus	270.21	(0.16)	0.30(0.20)	0.28(0.18)	-0.056
Brahman	150.12	(0.15)	0.18(0.22)	0.17(0.19)	-0.073
Brown Wagyu	100.16	(0.17)	0.25(0.24)	0.22(0.20)	-0.118
Black Wagyu	500.19	(0.16)	0.26(0.20)	0.26(0.19)	-0.018
Hereford	250.19	(0.16)	0.27(0.21)	0.26(0.19)	-0.045
Holstein	750.21	(0.16)	0.28(0.19)	0.28(0.18)	-0.026
Hanwoo	660.21	(0.16)	0.29(0.19)	0.28(0.18)	-0.025
Jeju Black	780.16	(0.16)	0.23(0.21)	0.21(0.19)	-0.076
Nellore	260.11	(0.15)	0.16(0.21)	0.15(0.19)	-0.059

N, sample size; MAF, minor allele frequency mean; H_O, observed heterozygosity; H_E, expected heterozygosity assumption in Hardy-Weinberg equilibrium; F_{IS}, inbreeding coefficient.

Table 2 Average F-Statistics over all loci (Weir and Cockerham, 1984)

Fixation indices	9 Breeds	P-values
F_{ST}	0.173**	0.00000
F_{IS}	-0.030	NS1.00000
F_{IT}	0.148**	0.00000

**p < 0.001. significant levels were obtained after 1,000 permutations; Fixation indices bootstrap over 20,000 bootstraps; F_{ST} = genetic differentiation among breed, F_{IS} = within population inbreeding, F_{IT} = total inbreeding, NS = not significant.

Table 3 Analysis of Molecular Variance among nine cattle breeds

Data set	Variance component (%)		
	Among populations	Among individuals within populations	Within individuals
All nine cattle breeds	17.31**	-2.48 NS	85.17**

**p < 0.001. Significant levels were obtained after 1,000 permutations; NS = not significant.

Table 4 Pairwise estimates of F_{ST} values according to Weir and Cockerham (1984) between 9 breeds populations (below diagonal) and Nei's D_A (1983) distances among 9 cattle breeds (above diagonal in bold)

	AG	BM	BRWG	BWG	HF	HST	HW	JJBC	NL
AG		0.09	0.07	0.046	0.047	0.035	0.034	0.054	0.1
BM	0.253**		0.1	0.083	0.098	0.08	0.071	0.078	0.037
BRWG	0.197**	0.328**		0.05	0.078	0.063	0.044	0.06	0.106
BWG	0.156**	0.245**	0.162**		0.057	0.04	0.025	0.042	0.091
HF	0.155**	0.292**	0.233**	0.192**		0.048	0.044	0.062	0.107
HST	0.122**	0.222**	0.183**	0.139**	0.161**		0.027	0.044	0.09
HW	0.118**	0.194**	0.126**	0.085**	0.152**	0.099**		0.028	0.079
JJBC	0.196**	0.274**	0.226**	0.164**	0.230**	0.161**	0.106**		0.084
NL	0.295**	0.165**	0.376**	0.275**	0.334**	0.249**	0.221**	0.296**	

**P < 0.001. significant levels were obtained after 1,000 permutations; AG (Angus), BM (Brahman), BRWG (Brown Wagyu), BWG (Black Wagyu), HF (Herford), HST (Holstein), HW (Hanwoo), JJBC (Jeju Black Cattle), NL (Nellore).

Table 5 Each 5 principal component and their contribution rate of variances and cumulative rates within 9 cattle breeds

Importance of components	PC1	PC2	PC3	PC4	PC5
Standard deviation	25.4522	7.2173	4.1527	10.51	
Proportion of Variance	27.8724	8.9190	16.7311	1.51	
Cumulative Proportion	27.8752	7.6717	68.49	100	

Table 6 Proportion of membership of the analyzed 9 cattle in each of the 11 clusters inferred in the ADMIXTURE program

Predefined populations	Inferred clusters											N
	1	2	3	4	5	6	7	8	9	10	11	
AG	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.97	0.01	0.00	27
BM	0.00	0.01	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15
BRWG	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10
BWG	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.96	50
HF	0.00	0.00	0.00	0.00	0.00	0.00	0.97	0.00	0.02	0.00	0.00	25
HST	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.02	0.02	0.00	75
HW	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.95	0.00	66
JJBC	0.38	0.00	0.00	0.00	0.39	0.23	0.00	0.00	0.00	0.01	0.00	78
NL	0.00	0.99	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	26

Table 7 Effective population size (N_E) across nine cattle breeds

Generation Ago	AG	BM	BRWG	BWG	HF	HST	HW	JJBC	NL
13	78	37	29	93	51	91	209	38	39
15	85	42	33	102	55	97	235	42	43
17	94	47	37	113	59	106	264	47	49
20	102	53	41	126	64	115	306	53	53
23	114	59	47	142	70	128	347	61	60
27	128	67	54	161	77	142	401	68	68
32	142	78	61	183	87	158	459	79	78
38	161	93	71	208	97	179	536	90	88
45	182	108	83	238	111	205	628	107	105
54	210	125	99	275	128	236	732	127	121
66	242	150	120	318	147	273	880	151	141
80	284	178	144	369	170	321	1020	181	166
98	336	217	174	427	202	371	1205	216	201
121	393	263	212	503	239	438	1383	260	238
150	471	313	255	578	290	515	1595	326	271
187	562	385	308	671	344	604	1839	386	336
234	652	460	394	789	416	723	2035	467	402
293	753	554	488	917	484	840	2188	551	474
367	881	639	586	1014	574	937	2318	668	575
454	1016	771	707	1121	683	1068	2386	775	646
552	1141	971	823	1272	781	1203	2532	891	787
658	1286	1092	933	1408	863	1300	2653	1031	903
760	1364	1151	1093	1492	962	1386	2686	1125	-
847	1544	1200	1162	1703	1088	1514	2678	1168	-
913	1653	-	1297	1708	1200	1567	2978	-	-

Gen Ago, Generation ago

Table 9 Mean pairwise linkage disequilibrium (LD), r^2 estimates for different inter-SNP distance estimated with Plink v1.9

Figures

Table 8 Linkage disequilibrium (r^2) as a correlation between r^2 and genetic distance between SNPs in the 9 cattle breed

GenAgo	AG		BM		BRWG		BWG		HF		HST		HW		JJBC		NL	
	r^2	SD																
13	0.08	0.10	0.16	0.18	0.19	0.19	0.07	0.09	0.12	0.14	0.07	0.09	0.03	0.04	0.15	0.18	0.15	0.17
15	0.08	0.11	0.16	0.18	0.19	0.20	0.07	0.10	0.12	0.15	0.07	0.10	0.03	0.04	0.16	0.19	0.15	0.18
17	0.09	0.11	0.16	0.18	0.20	0.20	0.07	0.10	0.13	0.16	0.08	0.10	0.03	0.05	0.16	0.19	0.15	0.18
20	0.09	0.12	0.16	0.18	0.20	0.20	0.08	0.10	0.14	0.16	0.08	0.11	0.03	0.05	0.16	0.19	0.16	0.18
23	0.09	0.12	0.17	0.19	0.20	0.21	0.08	0.11	0.15	0.17	0.09	0.11	0.03	0.05	0.16	0.19	0.17	0.19
27	0.10	0.12	0.17	0.19	0.21	0.21	0.08	0.11	0.15	0.18	0.09	0.12	0.03	0.05	0.17	0.20	0.17	0.19
32	0.10	0.13	0.17	0.19	0.21	0.21	0.08	0.11	0.16	0.18	0.09	0.12	0.03	0.05	0.17	0.20	0.17	0.19
38	0.11	0.13	0.17	0.19	0.21	0.21	0.09	0.11	0.17	0.19	0.10	0.12	0.03	0.05	0.18	0.20	0.18	0.20
45	0.11	0.14	0.18	0.20	0.22	0.21	0.09	0.12	0.17	0.19	0.10	0.13	0.04	0.05	0.18	0.21	0.18	0.20
54	0.12	0.14	0.18	0.20	0.22	0.22	0.09	0.12	0.18	0.20	0.10	0.13	0.04	0.05	0.18	0.21	0.19	0.20
66	0.12	0.15	0.18	0.20	0.22	0.22	0.09	0.13	0.18	0.20	0.11	0.14	0.04	0.05	0.18	0.21	0.19	0.21
80	0.12	0.15	0.18	0.21	0.22	0.22	0.10	0.13	0.19	0.21	0.11	0.14	0.04	0.05	0.18	0.21	0.19	0.21
98	0.13	0.15	0.19	0.20	0.22	0.22	0.10	0.14	0.20	0.22	0.12	0.15	0.04	0.06	0.19	0.21	0.20	0.21
121	0.13	0.16	0.19	0.21	0.22	0.22	0.11	0.14	0.20	0.22	0.12	0.15	0.04	0.06	0.19	0.22	0.20	0.22
150	0.14	0.16	0.19	0.21	0.23	0.22	0.12	0.15	0.21	0.22	0.13	0.16	0.05	0.07	0.19	0.22	0.22	0.23
187	0.14	0.17	0.20	0.21	0.23	0.22	0.12	0.16	0.21	0.23	0.13	0.16	0.05	0.07	0.20	0.22	0.22	0.23
234	0.15	0.18	0.20	0.22	0.23	0.22	0.13	0.17	0.22	0.23	0.14	0.17	0.05	0.08	0.20	0.23	0.23	0.23
293	0.16	0.19	0.21	0.22	0.23	0.23	0.14	0.17	0.23	0.24	0.15	0.18	0.06	0.10	0.21	0.23	0.24	0.24
367	0.17	0.20	0.22	0.23	0.24	0.23	0.15	0.19	0.24	0.25	0.16	0.20	0.07	0.11	0.22	0.24	0.24	0.24
454	0.18	0.21	0.23	0.23	0.24	0.23	0.17	0.21	0.25	0.25	0.18	0.21	0.09	0.13	0.23	0.25	0.26	0.25
552	0.20	0.21	0.22	0.23	0.25	0.24	0.18	0.21	0.26	0.26	0.19	0.21	0.10	0.15	0.24	0.26	0.26	0.25
658	0.20	0.22	0.23	0.23	0.26	0.24	0.19	0.23	0.28	0.27	0.20	0.23	0.11	0.16	0.24	0.26	0.27	0.26
760	0.22	0.23	0.25	0.25	0.26	0.24	0.20	0.23	0.28	0.27	0.22	0.24	0.12	0.18	0.25	0.27	-	-
847	0.22	0.23	0.26	0.25	0.27	0.24	0.20	0.23	0.28	0.27	0.22	0.24	0.14	0.18	0.27	0.27	-	-
913	0.22	0.23	-	-	0.26	0.24	0.21	0.25	0.28	0.26	0.23	0.26	0.13	0.19	-	-	-	-
959	0.23	0.24	-	-	-	-	0.23	0.25	0.30	0.28	-	-	-	-	-	-	-	-

Distance Interval (kb)	r^2 Mean \pm SD									
	AG	BH	BRWG	BWG	HF	HST	HW	JJBC	NL	Total
0-1	0.71 \pm 0.36	0.55 \pm 0.42	0.65 \pm 0.4	0.54 \pm 0.42	0.63 \pm 0.42	0.67 \pm 0.36	0.63 \pm 0.36	0.71 \pm 0.34	0.58 \pm 0.39	0.63 \pm 0.39
1-10	0.43 \pm 0.38	0.40 \pm 0.39	0.48 \pm 0.4	0.41 \pm 0.38	0.47 \pm 0.40	0.39 \pm 0.38	0.30 \pm 0.34	0.42 \pm 0.40	0.39 \pm 0.37	0.41 \pm 0.38
10-20	0.33 \pm 0.35	0.34 \pm 0.34	0.42 \pm 0.37	0.32 \pm 0.36	0.38 \pm 0.37	0.31 \pm 0.34	0.23 \pm 0.29	0.33 \pm 0.36	0.36 \pm 0.39	0.34 \pm 0.35
20-30	0.31 \pm 0.32	0.32 \pm 0.35	0.43 \pm 0.37	0.28 \pm 0.33	0.38 \pm 0.36	0.28 \pm 0.32	0.20 \pm 0.27	0.30 \pm 0.34	0.31 \pm 0.35	0.31 \pm 0.33
30-40	0.27 \pm 0.30	0.30 \pm 0.33	0.37 \pm 0.35	0.24 \pm 0.30	0.33 \pm 0.34	0.24 \pm 0.29	0.16 \pm 0.23	0.26 \pm 0.32	0.30 \pm 0.34	0.27 \pm 0.31
40-50	0.25 \pm 0.29	0.31 \pm 0.34	0.36 \pm 0.34	0.22 \pm 0.28	0.31 \pm 0.33	0.23 \pm 0.28	0.15 \pm 0.21	0.24 \pm 0.31	0.28 \pm 0.33	0.26 \pm 0.30
50-60	0.23 \pm 0.28	0.28 \pm 0.31	0.34 \pm 0.33	0.20 \pm 0.26	0.30 \pm 0.32	0.20 \pm 0.26	0.12 \pm 0.19	0.23 \pm 0.30	0.27 \pm 0.32	0.24 \pm 0.29
60-70	0.22 \pm 0.26	0.28 \pm 0.32	0.33 \pm 0.33	0.19 \pm 0.25	0.28 \pm 0.31	0.19 \pm 0.25	0.12 \pm 0.18	0.22 \pm 0.29	0.26 \pm 0.31	0.23 \pm 0.28
70-80	0.21 \pm 0.26	0.27 \pm 0.31	0.33 \pm 0.32	0.17 \pm 0.24	0.27 \pm 0.30	0.18 \pm 0.24	0.10 \pm 0.16	0.21 \pm 0.27	0.26 \pm 0.32	0.22 \pm 0.27
80-90	0.19 \pm 0.23	0.26 \pm 0.30	0.30 \pm 0.31	0.16 \pm 0.23	0.25 \pm 0.29	0.17 \pm 0.22	0.09 \pm 0.15	0.19 \pm 0.26	0.24 \pm 0.30	0.21 \pm 0.25
90-100	0.19 \pm 0.23	0.26 \pm 0.31	0.30 \pm 0.30	0.15 \pm 0.21	0.26 \pm 0.29	0.16 \pm 0.22	0.08 \pm 0.14	0.19 \pm 0.25	0.23 \pm 0.30	0.20 \pm 0.25
100-200	0.16 \pm 0.21	0.24 \pm 0.29	0.28 \pm 0.29	0.13 \pm 0.19	0.22 \pm 0.26	0.14 \pm 0.19	0.06 \pm 0.11	0.17 \pm 0.24	0.21 \pm 0.28	0.18 \pm 0.23
200-300	0.14 \pm 0.18	0.21 \pm 0.27	0.26 \pm 0.27	0.11 \pm 0.16	0.20 \pm 0.24	0.12 \pm 0.16	0.05 \pm 0.08	0.15 \pm 0.22	0.18 \pm 0.24	0.16 \pm 0.20
300-400	0.13 \pm 0.17	0.20 \pm 0.26	0.25 \pm 0.26	0.10 \pm 0.14	0.18 \pm 0.22	0.11 \pm 0.15	0.04 \pm 0.07	0.15 \pm 0.21	0.17 \pm 0.23	0.15 \pm 0.19
400-500	0.12 \pm 0.16	0.20 \pm 0.25	0.24 \pm 0.26	0.09 \pm 0.14	0.18 \pm 0.22	0.10 \pm 0.14	0.04 \pm 0.06	0.14 \pm 0.20	0.16 \pm 0.22	0.14 \pm 0.18
500-600	0.12 \pm 0.15	0.19 \pm 0.25	0.24 \pm 0.25	0.09 \pm 0.13	0.17 \pm 0.21	0.10 \pm 0.14	0.04 \pm 0.06	0.14 \pm 0.20	0.15 \pm 0.21	0.14 \pm 0.18
600-700	0.11 \pm 0.15	0.19 \pm 0.25	0.23 \pm 0.25	0.08 \pm 0.13	0.17 \pm 0.21	0.09 \pm 0.13	0.03 \pm 0.05	0.14 \pm 0.20	0.15 \pm 0.21	0.13 \pm 0.18
700-800	0.11 \pm 0.14	0.19 \pm 0.24	0.23 \pm 0.25	0.08 \pm 0.12	0.16 \pm 0.20	0.09 \pm 0.13	0.03 \pm 0.05	0.14 \pm 0.20	0.14 \pm 0.20	0.13 \pm 0.17
800-900	0.11 \pm 0.14	0.19 \pm 0.24	0.23 \pm 0.25	0.08 \pm 0.12	0.16 \pm 0.20	0.09 \pm 0.13	0.03 \pm 0.05	0.14 \pm 0.20	0.14 \pm 0.20	0.13 \pm 0.17
900-1000	0.11 \pm 0.14	0.19 \pm 0.24	0.23 \pm 0.25	0.08 \pm 0.12	0.15 \pm 0.19	0.09 \pm 0.12	0.03 \pm 0.05	0.13 \pm 0.19	0.14 \pm 0.20	0.13 \pm 0.17
Total	0.22 \pm 0.24	0.27 \pm 0.30	0.33 \pm 0.31	0.19 \pm 0.23	0.27 \pm 0.28	0.20 \pm 0.24	0.13 \pm 0.16	0.23 \pm 0.27	0.25 \pm 0.29	-

Matrix of pairwise Fst

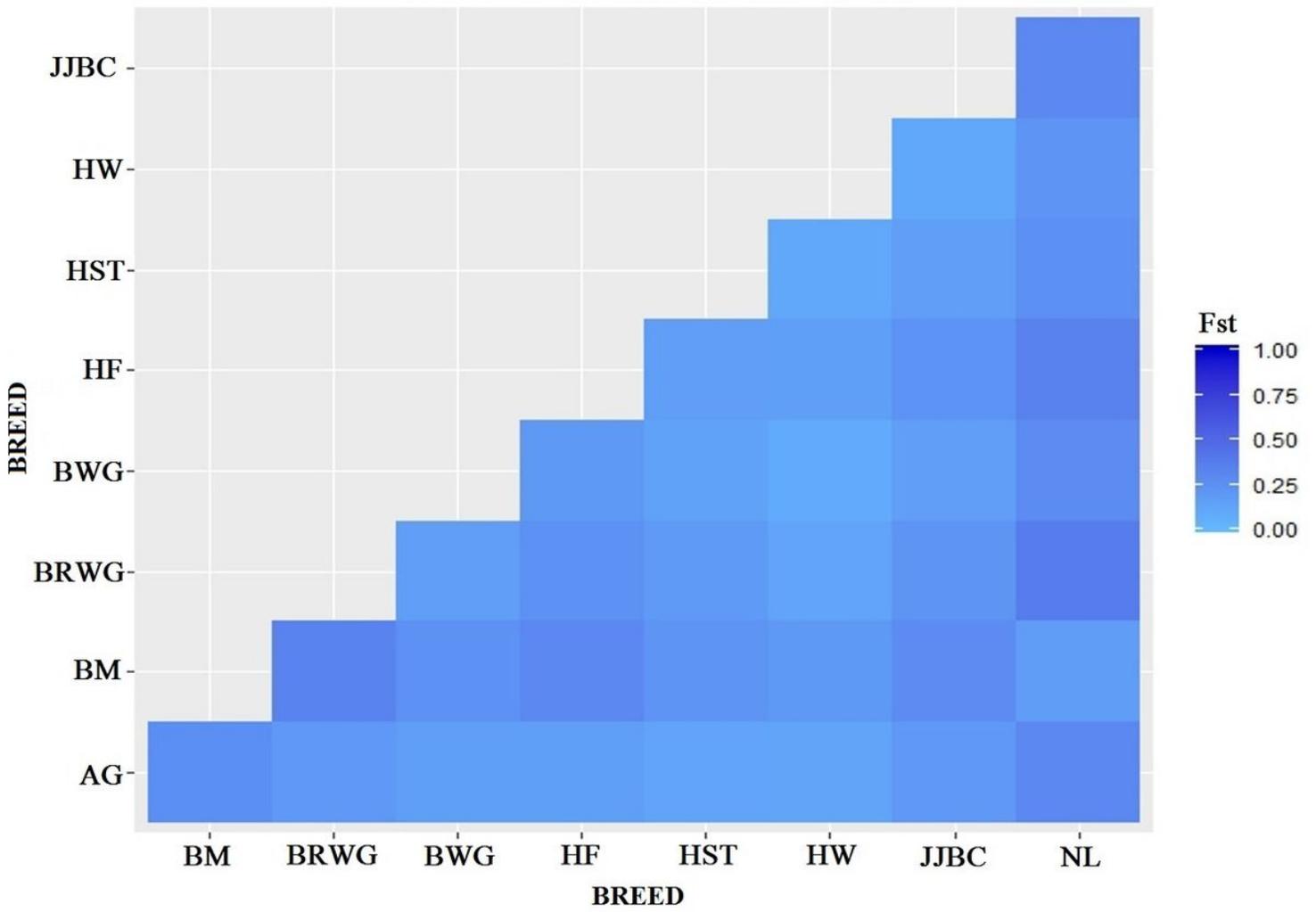


Figure 1

Graphical representation of pairwise Fst distance matrix

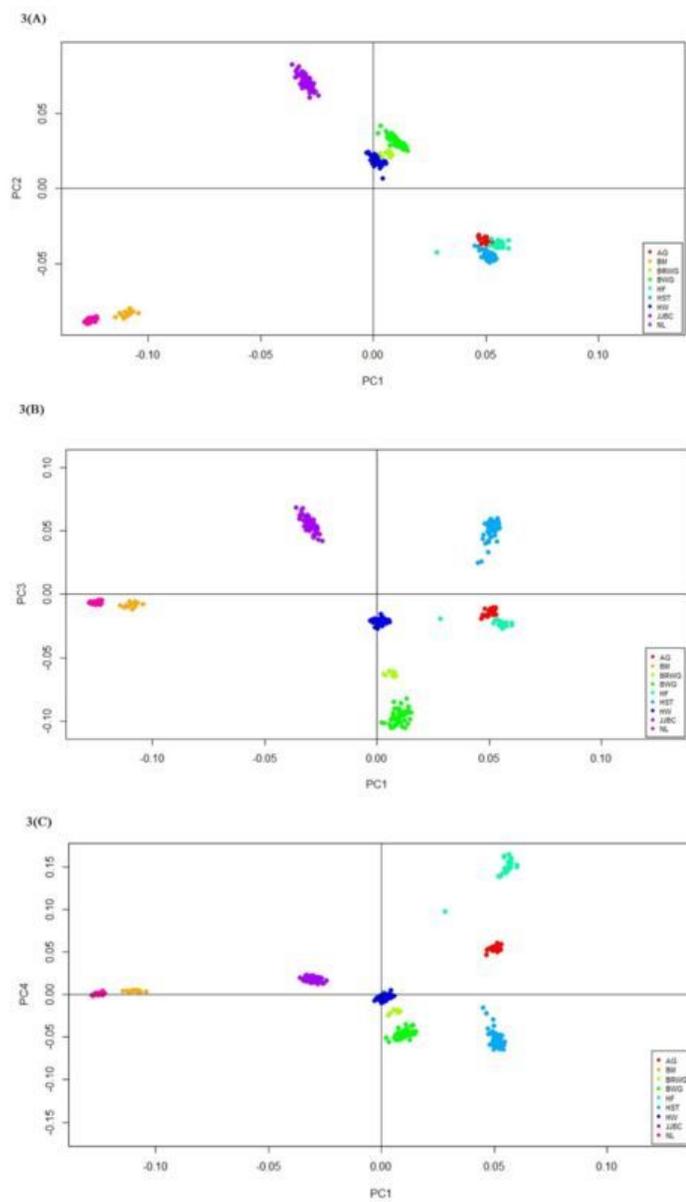


Figure 3

PCA analysis. First and second principal component (A), first and third principal component (B) and First and fourth principal component (C) analysis results from 18,524 SNPs in nine cattle populations; AG (Angus), BM (Brahman), BRWG (Brown Wagyu), BWG (Black Wagyu), HF (Hereford), HST (Holstein), HW (Hanwoo), JJBC (Jeju Black Cattle), NL (Nellore).

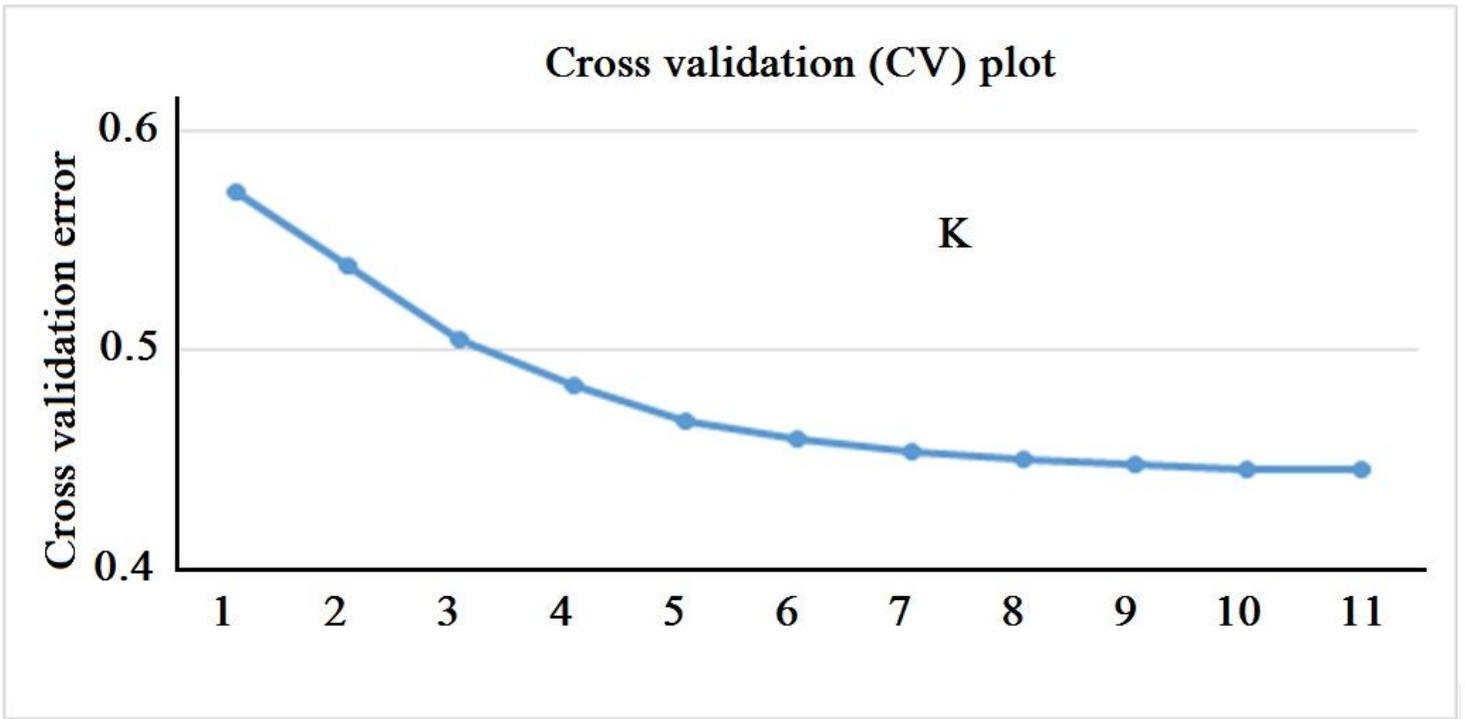


Figure 4

Cross validation plot. Cross-validation error was lowest for $k = 11$, which indicates that $k = 11$ is the optimal number of clusters.

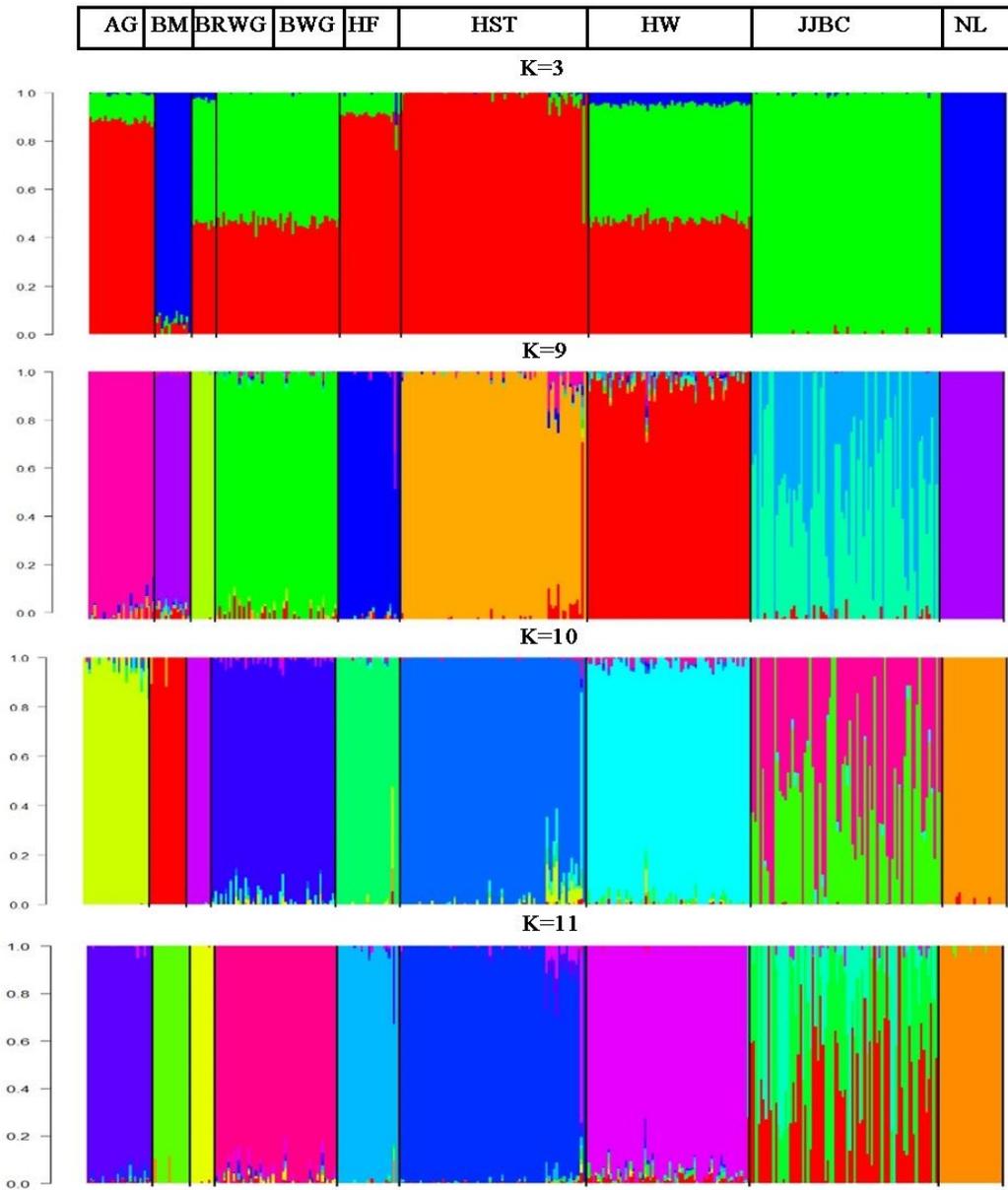


Figure 5

Clustering assignments of 9 cattle populations based on ADMIXTURE analysis at inferred K values ranging from 2 to 15; k=3, 9, 10 and K=11; AG(Angus), BM(Brahman), BRWG(Brown Wagyu), BWG(Black Wagyu), HF(Hereford), HST(Holstein), HW(Hanwoo), JJBC(Jeju Black Cattle), NL(Nellore). Each individual is represented by a single vertical line divided into K colored segments, where K is the number of the cluster assumed to have length proportional to each of the K inferred clusters. Black color separates the populations. Breeds are labeled by abbreviation at the top of figure.

EFFECTIVE POPULATION SIZE (N_E)

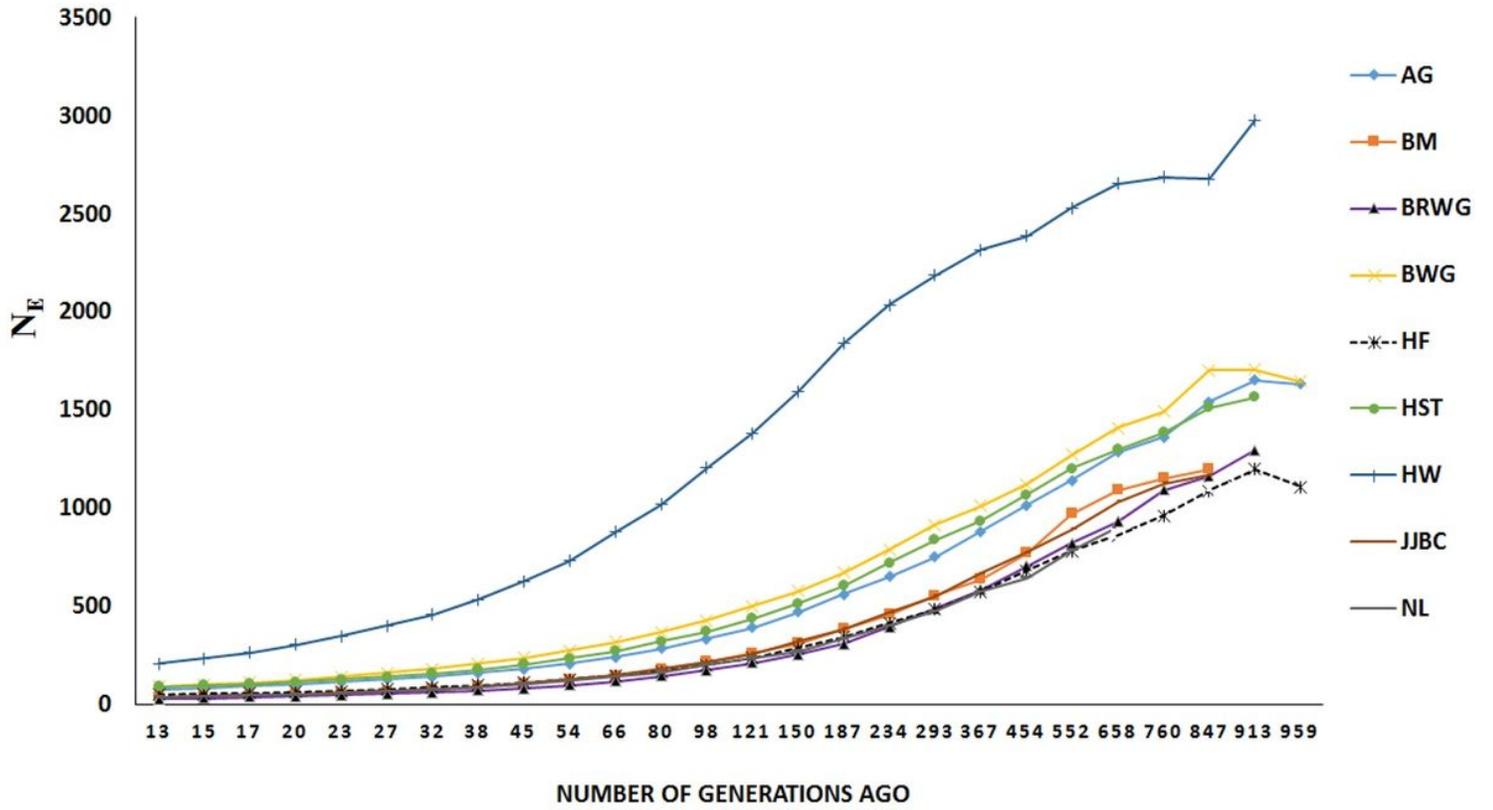


Figure 6

Effective population size (N_E) in 9 cattle breed.