

Genome-wide genetic diversity, population structure and admixture analysis in Eritrean Indigenous Cattle

S. Goitom¹, M.G. Gicheha², F.K. Njonge² & N. Kiplangat^{3#}

¹Department of Animal Sciences, Hamelmalo Agricultural College, Keren, Eritrea.

²Department of Animal Sciences, Jomo Kenyatta University of Agriculture & Technology, Nairobi, Kenya.

³Department of Animal Sciences, Egerton University, Njoro, Kenya.

(Received 20 July 2018; Accepted 9 March 2019; First published online 19 January 2020)

Copyright resides with the authors in terms of the Creative Commons Attribution 4.0 South African License.

See: <http://creativecommons.org/licenses/by/4.0/za>

Condition of use: The user may copy, distribute, transmit and adapt the work, but must recognize the authors and the South African Journal of Animal Science.

Abstract

Indigenous cattle play a vital role in subsistence and livelihood of pastoral producers in Eritrea. In order to optimally utilize and conserve these valuable indigenous cattle genetic resources, the need to carry out an inventory of their genetic diversity was recognized. This study assessed the genetic variability, population structure and admixture of the indigenous cattle populations (ICPs) of Eritrea using a genotype by sequencing (GBS) approach. The authors genotyped 188 animals, which were sampled from 27 cattle populations in three diverse agro-ecological zones (western lowlands, highlands and eastern lowlands). The genome-wide analysis results from this study revealed genetic diversity, population structure and admixture among the ICPs. Averages of the minor allele frequency (A_F), observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficient (F_{IS}) were 0.157, 0.255, 0.218, and -0.089, respectively. Nei's genetic distance (D_s) between populations ranged from 0.24 to 0.27. Mean population differentiation (F_{ST}) ranged from 0.01 to 0.30. Analysis of molecular variance revealed high genetic variation between the populations. Principal component analysis and the distance-based unweighted pair group method and arithmetic mean analyses revealed weak substructure among the populations, separating them into three genetic clusters. However, multi-locus clustering had the lowest cross-validation error when two genetically distinct groups were modelled. This information about genetic diversity and population structure of Eritrean ICPs provided a basis for establishing their conservation and genetic improvement programmes.

Keywords: genetic variability, molecular characterization, population differentiation

#Corresponding author: aarapngeno@gmail.com

Introduction

The concept of 'breed' was developed in the eighteenth century, and is meant to describe the differences (morphological and genetic) between populations. It has been estimated that over 6379 documented breeds from 30 species of livestock have been developed globally since domestication (FAO, 2000). Sub-Saharan Africa accounts for 145 cattle breeds/strains. These comprise two taurine Longhorns, 15 taurine Shorthorns, 75 Zebu (*Bos indicus*), 30 Sanga, 8 Zanga (Zebu-Sanga), nine breeds derived from interbreeding, and six composite breeds (Rege, 1999). This wide range of the breeds represents unique sets of genetic diversity (Dackson, 2008). The diversity among breeds of cattle resulted mainly from their coexistence with human beings who used artificial selection for various functions (Porto-Neto *et al.*, 2013). Cattle may have also diverged genetically as a result of natural selection and drift.

Indigenous cattle populations (ICPs) in Eritrea are distributed widely throughout diversified geographical and agro-ecological zones. Geographically isolated ICPs are subject to local climatic conditions and each may have unique characteristics that help the breed to survive and reproduce the harsh environments. Cattle around Red Sea coast are one example in which they have the ability to withstand the very hot climate (above 40 °C) and salty environment. However, the ICPs of Eritrea have not been characterized to establish their degree of genetic uniqueness objectively. Cluster analysis, based on single linkage agglomerative hierarchical and non-overlapping (SAHN) technique using morphological features, clustered Eritrean ICPs into two groups, namely Arado (highlands and eastern coast populations) and Barka

(western lowland populations) (Goitom *et al.*, 2016). However, Parker *et al.* (1998) recognized the need to confirm the phenotypic distinction through molecular characterization as this eliminates the effects of selection. Characterization based on a large number of molecular markers may provide unbiased estimates of similarities and differences of cattle populations. Properly chosen markers can be exploited comprehensively to access genetic variability across the entire genome. Specifically, the use of single nucleotide polymorphisms (SNPs) is one of the most powerful means of studying genetic diversity and admixture (Putman & Carbone, 2014). High-throughput sequencing technologies have improved the ability to discover and genotype SNPs. Genotyping-by-sequencing (GBS) is a cost-effective (De Donato *et al.*, 2013) and powerful genomic method for genetic diversity analysis owing to its ability to identify genomic variations throughout the genome (Baral *et al.*, 2018). However, utilization of high-throughput sequencing technology SNP markers to study the genetic variations and population structure of Eritrean ICPs has not been done. Therefore, the objective of this study was to assess the genetic variation, population structure and admixture of EIC, which were sampled from three diverse agro-ecological zones.

Materials and Methods

Eritrea lies between latitudes 12° 42' N and 18° 2' N and longitudes 36° 30' E and 43° 20' E. It is bordered by the Red Sea to the east, Sudan on the west, Ethiopia to the south, and Djibouti on its south-eastern side. The data used in this study were collected from different geographically identified cattle populations (n = 27) that were located within three agro-ecological zones in the country, namely the western lowlands (WLL), highlands (HL) and eastern lowlands (ELL) (Table 1).

Blood samples were collected from a total of 188 animals. Sampling criteria were based on about 60 km geographical distances between the sub-regions within the agro-ecological zone. The names of cattle populations were assigned based on the names of the sub-region.

Blood samples were collected from the jugular vein using 5 mL vacuette tubes coated with EDTA. The whole blood samples were preserved at -20 °C until the DNA was extracted according to the standard protocol described by Sambrook *et al.* (1998), which involves the use of proteinase K digestion and phenol-chloroform extraction procedures. The quantity and quality of the extracted genomic DNA were determined using gel electrophoresis to visualize the quality of bands produced. A Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA) was used to determine the amount and concentration of the DNA. This was to ensure that the samples contained sufficient high-quality DNA, that is, concentration greater than 50 ng/μL and quality greater than 1.8-2.0 A_{260/280}. For each sample, 30 ng/μl of DNA was diluted using double distilled water and stored at 4 °C before being shipped to Beijing Genomic Institute (BGI) laboratory for sequencing using a genotype-by-sequencing (GBS) approach. The GBS protocol follows the procedures described by Elshire *et al.* (2011).

Libraries were created with a unique barcode for each animal. Multiplexed libraries containing 96 samples each (including controls) were sent to BGI in Hong Kong for GBS. De-multiplexing of the barcodes was done by BGI and de-multiplexed barcoded reads were separated into individual files. After the de-multiplexing, raw FastQ data were subjected to quality control checks. QualityTrim software (Robinson, 2015) was used to trim sequences with a minimum quality set at 20, a minimum length of 50 and the maximum number of poor bases and N bases was set at 3. The resultant quality sequence reads were aligned against *Bos taurus* reference genome (UMD_3.1.1/BosTau8) using Borrow Wheel Aligner (BWA) (Li *et al.*, 2009). SAMtools v.1.3.1 (Li *et al.*, 2009) was used to convert the SAM files to BAM format, while implementing sorting, removing duplicates, merging and indexing procedures.

Variant calling was done by invoking the SAMtools mpile-up function with the called variants being stored in variant call format (VCF) files. Sex chromosomes, insertion and deletions (indels) and SNPs with quality below 20 were filtered using VCFtools v.0.1.15 (Danecek *et al.*, 2011). Additional quality control for SNPs was done by removing SNPs with calling rate below 98% and minor allele frequency (A_F) less than 5%. Linkage disequilibrium (LD) pruning (50 5 0.2) and verification of Hardy-Weinberg equilibrium (rejected at $P > 0.001$) was carried out using PLINK v1.07 software (Purcell *et al.*, 2007). The remaining autosomal SNP were used in making inferences of genetic diversity and relationships among the cattle populations.

PGDSpider software version 2.1.1.3 (Lischer & Excoffier, 2012) was used to convert PLINK PED and MAP files to Arlequin software format for population differentiation analysis. TASSEL 5.2.43 software (Bradbury *et al.*, 2007) was used to calculate Nei's (1972) standard genetic distance between pairs of cattle populations. Assessment of divergence among the cattle populations was done using the F_{ST} fixation index calculated by Arlequin software version 3.5 (Excoffier *et al.*, 2005).

Allele frequency, H_O and H_E and inbreeding coefficient were calculated using PLINK software version 1.07 (Purcell *et al.*, 2007). Visualization of population relationships as a phylogenetic tree was done by cluster analysis using the unweighted pair group method with the arithmetic mean bottom-up hierarchical clustering method (UPGMA, Sneath & Sokal, 1973). Patterns of population classification and genetic

structures were assessed using principal components analysis (PCA) in TASSEL 5.2.43 (Bradbury *et al.*, 2007) and multi-locus clustering of the populations was done using ADMIXTURE 1.2.3 (Alexander *et al.*, 2009). In making inferences on the number of clusters (K), one to six clusters were initially assumed, and the optimum value of K was determined as the number of clusters that showed the least cross-validation error. An analysis of molecular variance (AMOVA) was conducted to test the significance of variability between and within cattle populations, where cattle populations were nested within the agro-ecological zones (AEZs).

Table 1 Agro-ecological zones, sub-regions and their abbreviations from which the Eritrean indigenous cattle used in this study originated

Agro-ecological zones	Populations	Code
Western lowlands (WLL)	Shambuko	SHE
	Cambo ten	CAM
	Barentu	BAR
	Awgaro	AW
	Goluj	GOL
	Tekombia	TEKO
	Keru	KER
	Tekreret	TEK
	Akordet	CF
Highlands (HL)	Mai-Alba	MA
	Galaneffi	GAL
	Sheha	SHEH
	Afelba	AFE
	Senafe	SEN
	Serejeka	SER
	Habero	HAB
	Shieb-Seleba	SHES
	Hamelmallo	HAM
Eastern lowlands(ELL)	Shieb-Mensheb	MENS
	Dongolo	DOGO
	Shieb-Gedged	GHE
	Emberemi	EMB
	Foro	FOR
	Enghel-Eila	ENG
	Menkaneli	MENK
	Rekumbedin	REK
	Bada	BADA

Results

A total of 109 612 783 reads were called from the 188 sampled animals. After the quality control had been implemented, 1.22, 0.94, and 0.91 million SNPs remained for the populations in the WLL, HL, and ELL AEZs, respectively. These SNP were used to obtain the results that are described below. The number of quality variants within autosomal chromosomes varied by threefold among the populations that were sampled, ranging from 0.06 to 0.18 million (Table 2).

Table 2 Number of single nucleotide polymorphisms (millions) for all cattle populations from the three agro-ecological zones of Eritrea

Agro-ecological zones					
Western lowlands		Highlands		Eastern lowlands	
Barentu	0.12	Afelba	0.08	Bada	0.12
Cambo ten	0.16	Galanefhi	0.08	Dongolo	0.11
Awgaro	0.11	Habero	0.10	Emberemi	0.06
Keru	0.11	Hamelmallo	0.17	Enghel-Eila	0.10
Akordet	0.14	Mai-Alba	0.11	Foro	0.08
Shambuko	0.16	Serejeka	0.11	Shieb-Gedged	0.11
Goluj	0.18	Senafe	0.07	Menkaneli	0.11
Tekombia	0.12	Sheha	0.08	Shieb-Mensheb	0.12
Tekreret	0.12	Shieb-Seleba	0.14	Rekumbedin	0.10

Genetic diversity indicators, including H_O , H_E , F_{IS} and A_F averaged across all the loci and the estimates, are presented in Table 3. Across populations, H_E ranged from 0.190 to 0.343 with the smallest value being recorded in Foro cattle from the ELL and the greatest levels in Barentu and Awgaro cattle from the WLL. Observed heterozygosity ranged from 0.224 in Afelba cattle from the HL population to 0.316 in the Barentu and Awgaro populations. Deviations between the H_O and H_E were relatively small, averaging 0.037. The mean F_{IS} for the studied populations was -0.089, but ranged from -0.175 in the Menkaneli cattle from the ELL to 0.077 in Awgaro population. An average of 0.157 A_F was obtained from an analysis, with the Awgaro and Bada cattle having the highest A_F .

Greater heterozygosity indicates more genetic variation, which is important for adaptation to anticipated future changes in environmental and market conditions. The mean value of H_E of 0.218 obtained for the Eritrean *Bos-indicus* cattle is lower than the levels of 0.39 to 0.41 observed in indigenous Ethiopian cattle (Edea *et al.*, 2013), which are believed to be the cattle that are most related to the EIC populations. Similarly high levels of H_E (0.40) were reported in Sukuma, Tarime and Maasai in Tanzanian indigenous zebu populations (Msalya *et al.*, 2017). The within-population genetic variation as measured by H_O was greatest in the BAR and AW cattle populations. The high degree of genetic variability in these two populations may have resulted from uncontrolled breeding being practised in pastoral and agro-pastoral production systems. Intermixing with other indigenous populations, which is frequently observed with the BAR cattle population, is another possible explanation for the high degree of genetic variability. This intermixing occurs mostly around major towns, where animals from different parts of the agro-ecological zones are marketed. This movement of animals might have increased the gene exchange among the cattle populations around regional towns, resulting in low level of inbreeding. The results of the present study are consistent with findings in indigenous zebu populations of Tanzania that discovered high levels of admixture (Msalya *et al.*, 2017). The low heterozygosity detected in MENK cattle population was expected as the population is kept mainly in an area that is characterized by extremely hot environmental conditions, implying that the population is subjected to high natural selection pressure. This cattle population is found in the desert area of the southern Red Sea, where intermixing with other populations is rare owing to the environmental barrier. However, there is sufficient heterozygosity to imply low levels of inbreeding among ICPs.

Table 3 Observed and expected levels of heterozygosity, inbreeding coefficients and minor allele frequencies for Eritrean indigenous cattle populations

Population	Heterozygosity		Inbreeding	Minor allele frequency
	Observed	Expected		
Afelba	0.224	0.209	-0.037	0.147
Awgaro	0.316	0.343	0.077	0.227
Akordet	0.241	0.230	-0.010	0.164
Bada	0.299	0.321	0.070	0.227
Barentu	0.316	0.343	-0.008	0.154
Cambo te	0.225	0.220	-0.005	0.159
Dongolo	0.230	0.218	-0.044	0.154
Emberemi	0.230	0.215	-0.028	0.151
Enghel-Eila	0.247	0.189	-0.137	0.146
Foro	0.247	0.190	-0.162	0.142
Galaneffi	0.258	0.199	-0.137	0.149
Goluj	0.261	0.207	-0.122	0.154
Habero	0.257	0.200	-0.135	0.149
Hamelmallo	0.258	0.205	-0.121	0.153
Keru	0.249	0.196	-0.123	0.146
Mai-Alba	0.250	0.196	-0.127	0.145
Menkaneli	0.251	0.192	-0.175	0.145
Rekumbedin	0.256	0.219	-0.089	0.157
Senafe	0.256	0.219	-0.089	0.157
Serejeka	0.255	0.200	-0.117	0.150
Shambuko	0.257	0.203	-0.126	0.157
Sheha	0.244	0.198	-0.101	0.146
Shieb-Gedged	0.247	0.190	-0.137	0.149
Shieb-Mensheb	0.256	0.198	-0.148	0.145
Shieb-Seleba	0.255	0.201	-0.128	0.157
Tekombia	0.261	0.206	-0.125	0.154
Tekreret	0.255	0.201	-0.119	0.149
Grand mean	0.255 ± 0.01	0.218 ± 0.01	-0.089 ± 0.04	0.157 ± 0.04

Allele frequency was categorized into three groups; <0.05 , $\geq 0.05 - 0.1$ and ≥ 0.1 to <0.5 . Mean of these proportions of A_F were 34.13%, 13.89% and 56.15% respectively (Figure 1). In terms of monomorphic SNPs (<0.05), the lowest proportion (31.51%) was observed in the Goluj cattle from the WLL. Similarly, considering polymorphic SNPs ($\geq 0.1 - <0.5$), the highest frequency of polymorphic loci (59.73%) was observed in Akordet cattle also from the WLL.

Nei's standard pairwise genetic distances (D_s) were used to determine the genetic similarity of the cattle populations. These genetic distances ranged from 0.32 to 0.38. The smallest genetic distance (0.32) was recorded between Habero cattle from the HL and WLL Goluj cattle. Conversely, the greatest distances (0.38) were found between the Cambo ten population from WLL and Menkaneli cattle of ELL, and likewise separating Akordet cattle and the Barentu and Bada populations. Pairwise F_{ST} values between the populations ranged from 0.00 to 0.30, with the largest being recorded between Dongolo cattle of the ELL and Galaneffi cattle from the HL. The smallest F_{ST} value was obtained between the Awgaro and Habero cattle populations.

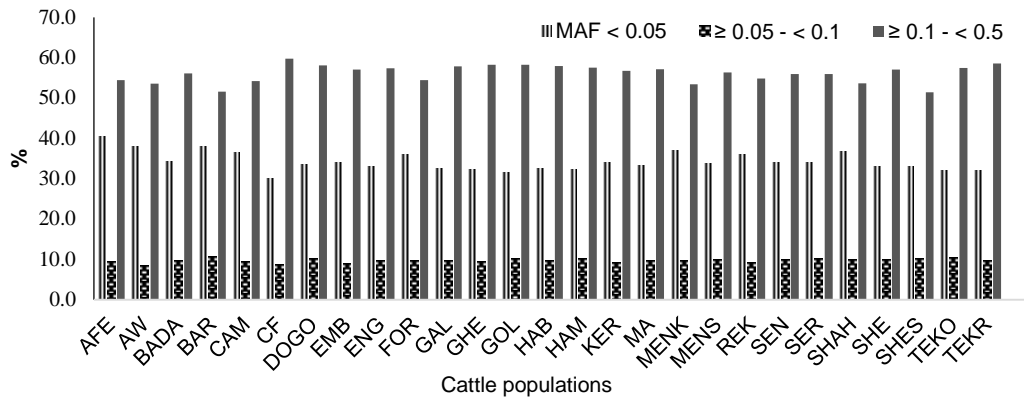


Figure 1 Distribution of minor allele frequencies in the various Indigenous cattle populations in Eritrea

AFE: Afelba, AW: Awgaro, BADA: Bada, BAR: Barentu, CAM: Cambo ten, CF: Akordet, DOGO: Dongolo, EMB: Emberemi, ENG: Enghel-Eila, FOR: Foro, GAL: Galanefhi, GHE: Shieb-Gedged, GOL: Goluj, HAB: Habero, HAM: Hamelmalo, KER: Keru, MA: Mai-Alba, MENK: Menkaneli, MENS: Shieb-Mensheb, REK: Rekumbedin, SEN: Senafe, SER: Serejeka, SHE: Shambuko, SHEH: Sheha, SHES: Shieb-Seleba, TEKO: Tekombia, TEKR: Tekreret

The UPGMA phylogenetic tree revealed two genetic groups with their own sub-groups (Figure 2). The first cluster was comprised mostly CF populations from WLL and the rest in a second cluster.

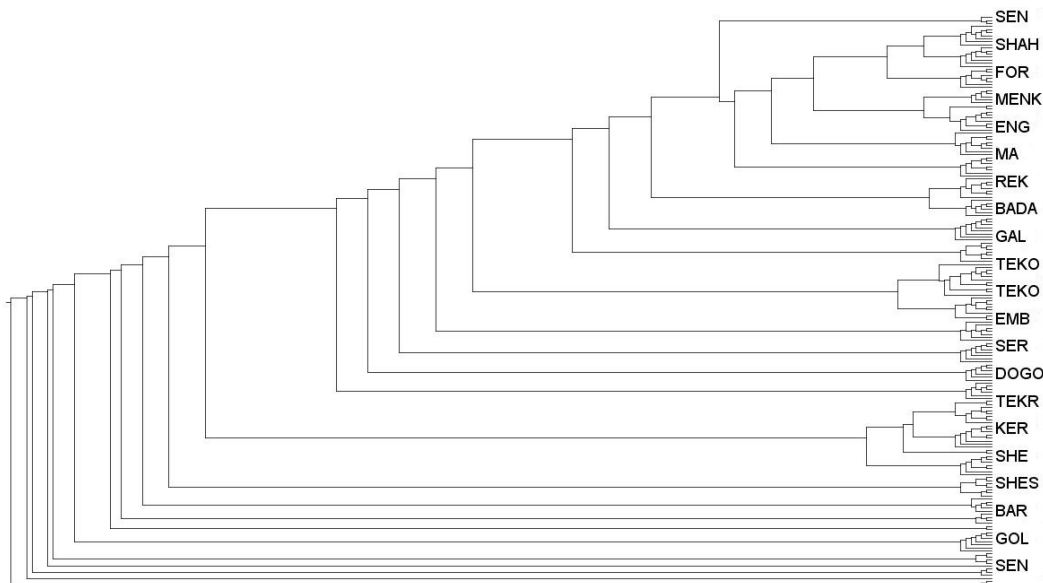


Figure 2 Phylogenetic tree produced with unweighted pair group method and arithmetic mean methodology to summarize genetic distance among Eritrean cattle populations

AFE: Afelba, AW: Awgaro, BADA: Bada, BAR: Barentu, CAM: Cambo ten, CF: Akordet, DOGO: Dongolo, EMB: Emberemi, ENG: Enghel-Eila, FOR: Foro, GAL: Galanefhi, GHE: Shieb-Gedged, GOL: Goluj, HAB: Habero, HAM: Hamelmalo, KER: Keru, MA: Mai-Alba, MENK: Menkaneli, MENS: Shieb-Mensheb, REK: Rekumbedin, SEN: Senafe, SER: Serejeka, SHE: Shambuko, SHEH: Sheha, SHES: Shieb-Seleba, TEKO: Tekombia, TEKR: Tekreret

Population clustering was further studied using PCA and admixture analysis. The PCA was performed using allele frequencies of the SNP markers. The first and the second principal components (PC1 and PC2) explained 23.1% and 19.73% of the total genetic variation, respectively (Figure 3). Generally, three clusters were revealed by PCA with the first cluster comprising Menkaneli, Mai-Alba and Keru cattle populations. The second cluster was large and comprised mostly cattle populations from the WLL and HL cattle populations. The third cluster included mostly cattle populations from the north of ELL (Emberemi, Foro, and Enghel-Eila

populations). However, the clusters indicated by PCA were quite divergent in the phylogenetic tree. Closer analysis indicated that PC1 resulted in a differentiation pattern between Emberemi, Foro and Enghel-Eila and Habero cattle populations from the rest. In PC2 the Rekumbedin and Menkaneli populations were separated from others.

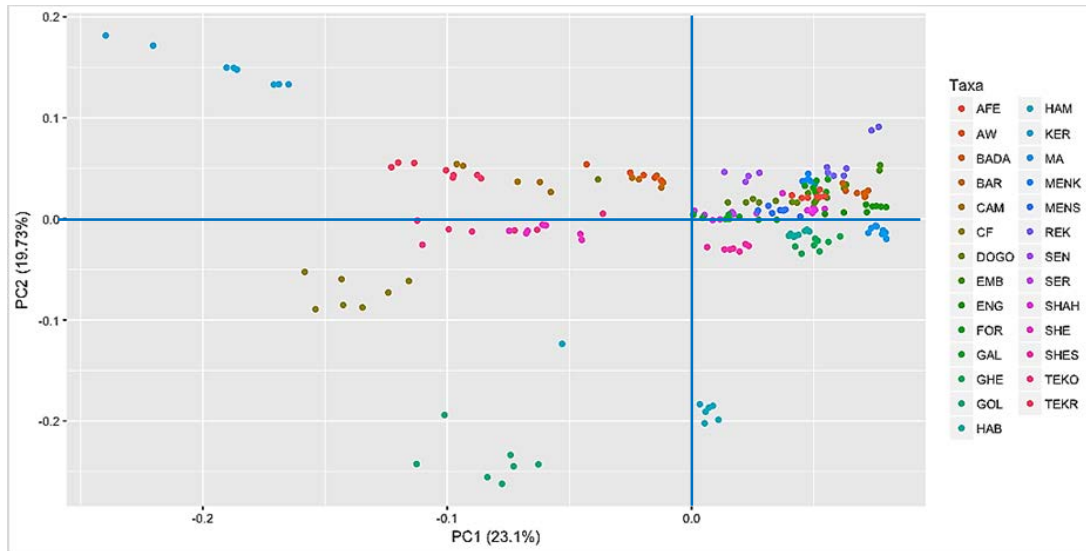


Figure 3 Graphic representation of diversity among cattle populations based on first two principal components (PC1 and PC2) from genotypes of autosomal SNP

AFE: Afelba, AW: Awgaro, BADA: Bada, BAR: Barentu, CAM: Cambo ten, CF: Akordet, DOGO: Dongolo, EMB: Emberemi, ENG: Enghel-Eila, FOR: Foro, GAL: Galanefhi, GHE: Shieb-Gedged, GOL: Goluj, HAB: Habero, HAM: Hamelmalo, KER: Keru, MA: Mai-Alba, MENK: Menkaneli, MENS: Shieb-Mensheb, REK: Rekumbedin, SEN: Senafe, SER: Serejeka, SHE: Shambuko, SHEH: Sheha, SHES: Shieb-Seleba, TEKO: Tekombia, TEKR: Tekreret

Model-based clustering revealed the population structure and levels of admixture among the populations. The cross-validation error was lowest for K=2 (0.669), indicating 2 as the most likely number of genetically distinct founder clusters (Figure 4). At K=2, a difference was observed between most admixed cattle populations with intermediate percentages of each founder and those populations that originated from one or the other of the founders.

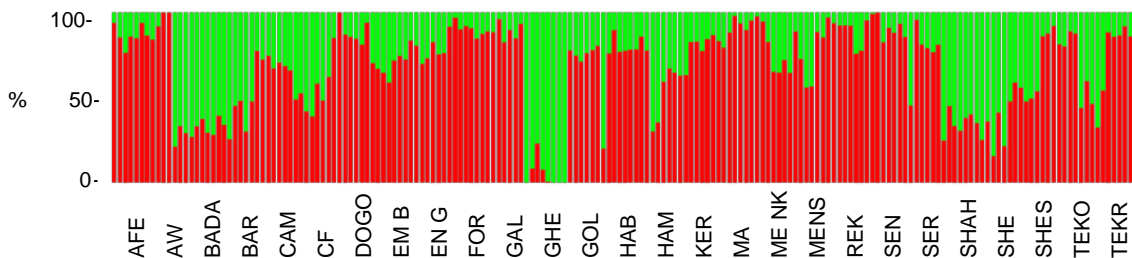


Figure 4 Clustering of the Eritrean cattle populations with two assumed founder groups. Each animal is represented by a single vertical line with colours indicating proportional membership in each cluster

AFE: Afelba, AW: Awgaro, BADA: Bada, BAR: Barentu, CAM: Cambo ten, CF: Akordet, DOGO: Dongolo, EMB: Emberemi, ENG: Enghel-Eila, FOR: Foro, GAL: Galanefhi, GHE: Shieb-Gedged, GOL: Goluj, HAB: Habero, HAM: Hamelmalo, KER: Keru, MA: Mai-Alba, MENK: Menkaneli, MENS: Shieb-Mensheb, REK: Rekumbedin, SEN: Senafe, SER: Serejeka, SHE: Shambuko, SHEH: Sheha, SHES: Shieb-Seleba, TEKO: Tekombia, TEKR: Tekreret

Analysis of molecular variance (AMOVA) was conducted based on genetic distances. Genetic variation between the AEZ and between populations within AEZ were both important sources of variation ($P < 0.001$) explaining 31% and 44% of the total, respectively (Table 4). Genetic variability within the populations explained 25% of the total variation.

Table 4 Analysis of molecular variance (AMOVA) based on genetic distances among cattle populations

Source of variation	DF	SS	MS	F	<i>P</i> -value
Between agro-ecological zones (AEZ)	2	0.057	0.029	72.5	<0.001
Between populations within AEZ	24	0.111	0.005	12.5	<0.001
Within populations	161	0.070	0.0004		

Discussion

Genetic characterization using SNP markers produced new information in terms of genetic diversity and differentiation of ICPs of Eritrea. Diversity analysis showed medium to high variation among indigenous cattle populations. Relatively, high A_F values (0.227) were observed in Awgaro and Bada populations. High differences in the distribution of A_F among the cattle populations indicated high variability among the cattle populations in the country. The overall mean (0.157) A_F of this study was greater than that of previous studies in indicine cattle (McKay *et al.*, 2008; Edea *et al.*, 2013; Kim *et al.*, 2015), but lower than was observed for Red Chittagong cattle (0.28) (Uzzaman *et al.*, 2014). Therefore, a high proportion of common A_F (56.15%) showed that there is genetic diversity within the Eritrean cattle populations though it is less than the observed 64% from a previous study of indigenous Zebu breeds in Pakistan (Hamid *et al.*, 2017).

High heterozygosity indicates a high degree of genetic variation within populations, which is important for adaptation and improvement of cattle in the light of the anticipated future fluctuation of environmental and market conditions. The mean expected value of heterozygosity of 0.218 that was obtained for the Eritrean *Bos-indicus* cattle was lower than the 0.39 to 0.41 that was found for indigenous Ethiopian cattle (Edea *et al.*, 2013), which are the closest comparable cattle population to Eritrea ICPs. Similarly, high levels of expected heterozygosity (0.40) were reported in Sukuma, Tarime, and Maasai in Tanzanian indigenous zebu populations (Msalya *et al.*, 2017). The high genetic variability in Barentu and Awgaro cattle could have resulted from uncontrolled breeding being practised in pastoral and agro-pastoral production systems. Another possible explanation could be because of frequent intermixing of Barentu cattle with other indigenous populations around major towns where animals from different parts of the WLL are marketed. This movement of animals might have increased gene exchange among the cattle populations and also resulted in a reduced level of inbreeding. The results of present study are consistent with findings in indigenous zebu populations of Tanzania, which also exhibited levels of admixture (Msalya *et al.*, 2017). The low heterozygosity detected in Menkaneli cattle was expected because the population is kept mainly in an area that is characterized by an extremely hot environment, which may imply a high level of natural selection pressure. This cattle population is found in the desert area of southern Red Sea, where intermixing with other populations is rare.

Results from AMOVA indicated that high proportions of variance existed between the AEZ and among populations within them. The result supports the findings obtained by Goitom *et al.* (2016), which showed that Eritrean ICPs had three distinct morphological categories. In almost all cases, these cattle populations had negative F_{IS} values, suggesting a high proportion of heterozygote genotypes. The estimates of Nei's genetic distances indicated that the Akordet cattle were relatively distant (0.38) from Bada, Barentu and Afelba cattle populations. This observation might be because of the long geographic distance between these populations or the tendency of farmers to discriminate against Akordet cattle. This discrimination might be based on the casual observation that they are considered aggressive and thus make routine management practices almost impossible to carry out. The relatively small genetic distances between cattle populations that are found in the HL, eastern coast (Arado) and north of the ELL (Arebo) regions may result from their close proximity and thus greater implied opportunity for interbreeding. The genetic differentiation among populations suggests a lack of gene flow between them, which has possibly resulted from physical barriers to migration such as escarpments and deserts.

Phylogeny and population structure analysis were done based on genome-wide SNPs from the 27 cattle populations. Multivariate PCA was performed to determine the cattle population structure. A

Multivariate analysis is essential in grouping correlated characters into uncorrected few new groups (Manly, 1994). This reduction of a set of original variables enables the maximum proportion of variance to be accounted for from a minimum number of new composite variables. The three clusters that were formed by UPGMA had consistency with the classification by PCA into three groups. However, the second group, which comprised most WLL and highland cattle populations were not distinctively separated by both principal components. The separation of group one (NEL) from the rest of the groups is possibly owing to the barrier of the eastern escarpment and the desert climatic conditions that may have restricted the movement of cattle from other populations. The third group comprised cattle populations of WLL and HL cattle. However, it is expected to form two separate groups because they have differences in their origin. The western lowland population (Barka type) is a member of the Large East African Zebu group, while the Arado type is from Zanga cattle. The proportion of genome that could be attributed to a common founder population was inferred by the ADMIXTURE analysis. The two clusters (K=2) that were inferred by the lowest cross-validation comprised mostly HL and western lowlands cattle populations, which formed the first cluster (red colour), and the Arebo type, which formed the second cluster (green colour). At K=3, three clusters were realized, which is consistent with the results of PCA and cluster analysis. The high value of admixture between Arado and Arebo types contradicts the result of the Eritrea ICPs morphological classification (Goitom *et al.*, 2016). Moreover, the high admixture among most cattle populations could be because of common ancestry and high gene flow owing to the migration of cattle populations because of the small land area that characterizes Eritrea.

Conclusions

The results revealed patterns of genetic variation in the ICPs. The model-based clustering and distance-based phylogenetic tree identified two clusters (eastern lowland cattle population and others), while the PCA seemed to indicate three groups. The observed high levels of variability between AEZ, among populations within the AEZ and among animals within the population will assist in the establishment of future genetic improvement and conservation programmes for the Eritrean ICPs.

Acknowledgements

The authors are grateful to National Commission for Higher Board (NCHB) of Eritrea, Japan International Cooperation Agency (JICA), Jomo Kenyatta University of Agriculture and Technology (JKUAT), and Hamelmalo Agricultural College (HAC) for the financial support.

Authors' Contributions

The authors designed the research project jointly. SG collected the samples, conducted the laboratory work and data analysis, and wrote the manuscript (as part of his PhD programme). MGG, FKN and NK assisted in conducting the laboratory work, data analyses and editing the manuscript. All authors read and approved the final draft of the manuscript.

Conflict of Interest Declaration

The authors declare that they have no conflict of interest.

References

- Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655-1664. <https://genome.cshlp.org/content/19/9/1655.full>
- Baral, K., Coulman, B., Biligetu, B. & Fu, Y., 2018. Genotyping-by-sequencing enhances genetic diversity analysis of crested wheatgrass (*Agropyron cristatum* (L.) Gaertn). 19, 2587. *Int. J. Mol. Sci.* 19(9), E2587. doi: 10.3390/ijms19092587
- Barcaccia, G., Felicetti, M., Galla, G., Capomaccio, S., Cappelli, K. & Albertini, E., 2013. Molecular analysis of genetic diversity, population structure and inbreeding level of the Italian Lipizzan horse. *Livest Sci.* 151(2-3), 124-33. <https://www.sciencedirect.com/science/article/abs/pii/S187114131200443X>
- Bradbury P.J., Zhang Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. & Buckler, E.S., 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633-2635. <https://academic.oup.com/bioinformatics/article/23/19/2633/185151>
- Dackson, N.Z., 2008. Genetic characterization of Zambian native cattle breeds. Thesis, Animal and Poultry Sciences, Faculty of Virginia Polytechnic Institute and State University.
- Danecek P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., ..., Durbin, R., 2011. 1000 Genomes Project Analysis Group. *Bioinformatics*, variant call format and VCFtools. <http://dx.doi.org/10.1093/bioinformatics/btr330>
- De Donato, M., Peters, S.O., Mitchell, S.E., Hussain, T. & Imumorin, I.G., 2013. Genotyping-by-sequencing (GBS): A novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS ONE* 8(5), e62137. <https://doi.org/10.1371/journal.pone.0062137>
- Edea, Z., Dadi, H., Kim, S., Dessie, T., Lee, T., Kim, H., Kim, J.-J. & Kim, J.-S., 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front. Genet.* 4, 35. <https://www.ncbi.nlm.nih.gov/pubmed/23518904>

- Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. & Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Excoffier, L., Laval, G. & Schneider, S., 2005. Arlequin version 3.0: An integrated software package for population genetics data analysis. Computational and Molecular Population Genetic Laboratory (CMPG), Institute of Zoology, University of Berne, Switzerland.
- FAO, 2000. World watch list for domestic animal diversity. 3rd edition. FAO, Rome, Italy. <http://www.fao.org/3/x8750e/x8750e00.htm>
- Hamid, M., Kim, E., Huson, J., Adeela, A., Riley, D., Talat, N., Afzal, A., Khalid, J. & Sonstegard, T., 2017. Genome-wide SNPs analysis of indigenous zebu breeds in Pakistan. *Biotechnology in Animal Husbandry* 33 (1), 13-25. <http://www.doiserbia.nb.rs/Article.aspx?id=1450-91561701013M#.XiF1Ey2B3q0>
- Goitom, S., Gicheha, M. & Teclehimanot, G., 2016. Morphological characterization of indigenous cattle in Eritrea. Fifth RUFORUM Biannual Conference, 17-21 October 2016 Cape Town, South Africa, 14(2), 787-791.
- Kim, E.S., Sonstegard, T.S. & Rothschild, M.F., 2015. Recent artificial selection in US Jersey cattle impacts autozygosity levels of specific genomic regions. *BMC Genomics* 6, 302. <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-015-1500-x>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078-2079. <https://www.ncbi.nlm.nih.gov/pubmed/19505943>
- Lischer, H.E.L. & Excoffier, L., 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28, 298-299. <https://academic.oup.com/bioinformatics/article/28/2/298/198891>
- Manly, F.J., 1994. Multivariate statistical methods. A primer. Second edition. Chapman & Hall, Boundary Row, London, UK.
- McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., ..., Moore, S.S., 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genomics* 9, 37. <https://bmcbgenet.biomedcentral.com/articles/10.1186/1471-2156-9-37>
- Msalya G., Kim, E-S., Laisser, E.L.K., Kipanyula, M.J., Karimuribo, E.D., Kusiluka, L.J.M., Chenyambuga, S.W. & Rothschild, M.F., 2017. 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* 12(1): e0171088. <https://doi.org/10.1371/journal.pone.0171088>
- Nei, M., 1972. Genetic distance between populations. *American Naturalist* 106, 283-292. <https://www.journals.uchicago.edu/doi/10.1086/282771?mobileUi=0&>
- Parker, P.G., Snow, A.A., Schug, M.D., Booton, G.C. & Fuerst, P.A., 1998. What molecules can tell us about populations: choosing and using a molecular marker. *Ecology* 79(2), 361-382.
- Porto-Neto, L.R., Sonstegard, T. S., Liu, G.E., Bickhart, D.M., Da Silva, M.V.B., Machado, M.A., Utsunomiya, Y.T., Garcia, J.F., Gondro, C. & Van Tassell, C.P., 2013. Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14, 876. <https://doi.org/10.1186/1471-2164-14-876>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A. & Bender, D., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *J. Hum. Genet.* 81, 559-575. <https://www.ncbi.nlm.nih.gov/pubmed/17701901>
- Putman, A.I. & Carbone, I., 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecol. Evol.* 4(22) 4399-4428. <https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.1305>
- Rege, J.E.O., 1999. Issues and current developments in the conservation of indigenous African domestic animal diversity. Proceedings 5th World Congress Genetics Applied to Livestock Production (Guelph, Canada) volume 21, 439-446. <http://www.wcgalp.org/system/files/proceedings/1994/issues-and-current-developments-conservation-indigenous-african-domestic-animal-diversity.pdf>
- Robinson, A., 2015. Quality trim. <https://bitbucket.org/arobinson/qualitytrim>
- Sambrook, J., Fritsch, E.F. & Maniatis, T., 1989. Molecular cloning: A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, New York, USA.
- Sneath, P.H.A. & Sokal, R.R., 1973. Numerical taxonomy. The principles and practice of numerical classification. W.H. Freeman & Co., San Francisco, California, USA. <https://www.jstor.org/stable/2412767?seq=1>
- Uzzaman, M.R., Edea, Z., Bhuiyan, M.S., Walker, J., Bhuiyan, A.K. & Kim, K.S., 2014. Genome-wide single nucleotide polymorphism analyses reveal genetic diversity and structure of wild and domestic cattle in Bangladesh. *Asian Australasian J. Anim. Sci.* 27(10), 1381-1386. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4150168/>