MOLECULAR CHARACTERIZATION OF SELECTED MILK GENES IN MAJOR CATTLE POPULATIONS OF ETHIOPIA

Behailu Samuel

A Dissertation Submitted to

The department of Applied Biology

School of Applied Natural Sciences

Presented in Fulfillment of the Requirement for the Degree of Doctor of Philosophy in Biotechnology (Specialization in Animal Biotechnology)

Office of Graduate Studies

Adama Science and Technology University

May, 2023

Adama, Ethiopia

MOLECULAR CHARACTERIZATION OF SELECTED MILK GENES IN MAJOR CATTLE POPULATIONS OF ETHIOPIA

Behailu Samuel

Supervisor(s)

Hunduma Dinka (Ph.D.,Prof.)

Hailu Dadi (Ph.D.)

A Dissertation Submitted to

The department of Applied Biology

School of Applied Natural Sciences

Presented in Fulfillment of the Requirement for the Degree of Doctor of Philosophy in Biotechnology (Specialization in Animal Biotechnology)

Office of Graduate Studies

Adama Science and Technology University

May, 2023

Adama, Ethiopia

DECLARATION

I hereby declare that this Dissertation entitled "MOLECULAR CHARACTERIZATION OF SELECTED MILK GENES IN MAJOR CATTLE POPULATIONS OF ETHIOPIA" is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this dissertation have been duly acknowledged through appropriate citations.

Behailu Samuel _____ _____

Name of the student Signature Date

RECOMMENDATIONS

I/we, the supervisor(s) of this dissertation, hereby certify that I/we have read and revised the dissertation entitled "MOLECULAR CHARACTERIZATION OF SELECTED MILK GENES IN MAJOR CATTLE POPULATIONS OF ETHIOPIA" prepared under my/our guidance by Behailu Samuel submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Biotechnology (Specialization in Animal Biotechnology).

Therefore, I/we recommend the submission of the dissertation to the department for further review and defense.

Hunduma Dinka (Ph.D., Prof.)		24/02/2023
Major Supervisor	Signature	Date
Hailu Dadi (Ph.D.,Asst.Prof.)		24/02/2023
Co-supervisor	Signature	Date

APPROVAL SHEET OF PH.D. DISSERTATION

I/we hereby certify that the recommendations and suggestions made by the board of examiners are appropriately incorporated into the final version of the dissertation entitled "**MOLECULAR CHARACTERIZATION OF SELECTED MILK GENES IN MAJOR CATTLE POPULATIONS OF ETHIOPIA**" by Behailu Samuel

Prof. Hunduma Dinka _____ _____

Major Supervisor Signature Date

Dr. Hailu Dadi _____ _____

Co-supervisor Signature Date

We, the undersigned, members of the Board of Examiners of the dissertation open defense by Behailu Samuel Hailu have read and evaluated the dissertation entitled "**MOLECULAR CHARACTERIZATION OF SELECTED MILK GENES IN MAJOR CATTLE POPULATIONS OF ETHIOPIA**" and examined the candidate during open defense. This is, therefore, to certify that the dissertation is accepted for partial fulfillment of the requirement of the degree of Doctor of Philosophy in Biotechnology (Specialization in Animal Biotechnology).

_____ _____ _____

Chairperson Signature Date

Dr. Ketema Tafess _____ _____

Internal Examiner 1 Signature Date

Dr. Kero Jemal _____ _____

Internal Examiner 2 Signature Date

Dr. Kefena Effa _____ _____

External Examiner 1 Signature Date

Dr. Bedaso Mammo _____ _____

External Examiner 2 Signature Date

Finally, approval and acceptance of the dissertation is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the candidate's Department Graduate Council (DGC) and School Graduate Committee (SGC).

_____ _____ _____

Department Head Signature Date

_____ _____ _____

School Dean Signature Date

_____ _____ _____

Office of Postgraduate Studies, Dean Signature Date

ACKNOWLEDGEMENTS

Above all, I would like to praise the Almighty God for giving his grace and endless love, giving me peace, health and strength in my work.

I wish to express my deepest gratitude and appreciation to my supervisors Prof. Hunduma Dinka from Adama Science and Technology University and Dr. Hailu Dadi from Bio and Emerging Technology Institute (BETin) for their unreserved support, guidance, encouragements and constructive suggestions. Their excellent teaching and constructive criticism and comments from the initial conception to the end of this work and their willingness to help at any time are beyond the author's expression. This Ph.D. dissertation would have never come up to end without the personal efforts made by Prof. Hunduma Dinka to ease the way for sequencing at the Konkuk University of South Korea and Dr. Hailu Dadi for bringing primers from abroad and hosting me to work most of my experiments at Bio and Emerging Technology Institute of Ethiopia.

Particularly, I would like to express my sincere appreciation to Prof. Chankyu Park and his team at the University of Konkuk South Korea for dedicating their time, resource and expertise for sequencing of PCR amplicons to conduct part of this work.

I would like to extend my gratitude to Adama Science and Technology University, Department of Applied Biology for the continuous assistance during my study. I would like to thank Bio and Emerging Technology Institute for hosting me as a graduate fellow.

My special thanks goes to BETin laboratory, genomics and bioinformatics section staff. Even the beginning seems confusing; you have brought me many steps forward in genomics. I convey special thanks to Dr. Zewdu Edea for reviewing the whole dissertation.

Finally, my deepest gratitude also goes to all my family and associates for their love, care, prayer and moral support during ups and down in the whole work that enabled me to succeed throughout my academic life.

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF TABLES

LIST OF TABLES IN APPENDICES

LIST OF FIGURES

LIST OF FIGURES IN APPENDICES

LISTS OF ACRONYMS AND ABBREVIATIONS

BG	Begait
BHC	Boran * HF cross
BR	Boran
BTA	*Bos taurus* autosome
CSA	Central Statistical Authority
DGAT1	*diacylglycerol O-acyl transferase* 1
DMY	Daily milk yield
DNA	Deoxyribo Nucleic Acid
dNTP	deoxy Nucleotide Tri Phosphate
EDTA	ethylene diamine tetra acetic acid
F1	First Generation Crossbred
FG	Fogera
HF	Holstein Friesian
HR	Horro
HWE	Hardy-Weinberg equilibrium
K	Lysine
MAF	minor allele frequency
MEGA	Molecular evolutionary genetic analysis
MJN	Median joining network
NCBI	National Center for Biotechnology information
PCR	Polymerase chain reaction
PIC	Polymorphic information content
PRL	*Prolactin*
QTL	quantitative trait loci
Rpm	Revolution per minute
SE	Standard Error
SNF	Solid not Fat
SNPs	Single Nucleotide Polymorphisms
SSCP	Single strand conformation polymorphism
TAGs	Triacylglycerols

ABSTRACT

*Milk genes are potential candidate for milk production due to their biological significance on the quantitative traits of interest, of which diacylglycerol O-acyl transferase 1 (DGAT1) and prolactin (PRL) genes are considered to be functional candidates with major effect on milk yield and composition traits. However, it has not been well investigated whether the known DGAT1 and PRL mutations are segregating and associated with milk yield and milk composition traits in zebu and their crossbred populations in Ethiopia. Thus, this study was undertaken to assess the genetic variation of DGAT1 exon-8 and PRL exon-4 genes, to estimate the allelic and genotypic frequencies and to assess the effect of DGAT1 K232A and PRL variants on daily milk yield and composition traits in Boran, Begait, Fogera, Horro and Boran * Holstein crossbred cattle populations. Blood samples for genomic DNA extraction and milk samples for analysis of milk components were collected from ninety two ((17 Boran, 16 Begait, 18 Fogera, 17 Horro, and 24 Boran * HF cross (75% F1 HF*(HF * Boran)) randomly sampled cattle populations. Gene specific primers were used to amplify 278bp and 294bp regions of DGAT1 exon-8 and PRL exon-4 genes, respectively. DNA sequencing was carried out using Sanger method to detect the genetic variation in selected gene regions. Unique accession numbers were obtained for each haplotype sequences of DGAT1 (ON262825 - ON262849) and PRL (ON677410 - ON677427) genes and deposited in the GeneBank. Heterozygosity, Polymorphic information content (PIC), allele and genotype frequency and test for Hardy-Weinberg equilibrium were calculated using Power Marker. Pairwise F_{ST} values, neutrality tests and number of haplotypes were computed using ARLEQUIN software. Phylogenetic analysis of coding region was carried out using MEGA 11 software and univariate analysis of variance was carried out through GLM procedure of SAS software to assess the associations of detected genetic variation with milk yield and composition. Results revealed that DGAT1 gene had 11 haplotypes; 0.615 haplotype diversity and 0.010 nucleotide diversity in the studied populations. Three genotypes AA, KA, and KK were detected in Boran * HF cross whereas, KA, and KK genotypes in the Boran, Begait, Horro and Fogera populations found to be in HWE. The frequency of K232 was ranged from* (0.50 in Boran * HF cross to 0.97 in Horro) populations. *In the Boran * HF crosses, the genotype KA (0.50) was prevalent, whereas in the Horro populations genotype KK (0.94) was more prevalent. Substitution effect of one copy of K allele leads to significant ($p<0.01$) increase of fat and protein content and, decrease of milk yield, lactose and solid not fat content in all breeds considered. Three SNPs were identified in the coding region of PRL and g.8323T>A, showed missense mutation with the presence of five different haplotypes. Substitution effect of one copy of G allele leads to significant ($p<0.05$) increase of fat, SNF and protein content and decrease of daily milk yield in the sampled populations. Heterozygosity, PIC and inbreeding coefficient revealed genetic variation in DGAT1 and PRL genes. The study indicated that DGAT1 K232A marker had significant effects on daily milk yield, milk fat and lactose content and whereas, SNP g.8398A>G only have significant effect ($p<0.05$) on average daily milk yield, fat and solid not fat percentage in all studied cattle populations. Overall, the results of this study revealed that DGAT1 and PRL genes were polymorphic in the sampled cattle populations, and these markers could be utilized to accelerate future molecular breeding of dairy cattle after validation in larger populations.*

Keywords: *DGAT1; Diversity; Ethiopian cattle; Frequency distribution; PRL*; SNPs

CHAPTER 1

1. INTRODUCTION

1.1. Background

Milk contains almost all the nutrients vital for life with high biological value and animal milk can play an important role in the diets of children in populations with very low fat intakes and limited access to other animal source (FAO, 2013; Haug *et al*., 2007). Even though, milk has essential contribution worldwide, particularly it provides 2, 5 and 6 % of dietary energy, protein and fat supply, respectively in Africa and account for an estimated 11% of food production by value (OECD-FAO, 2022). Milk volume increment has been the first breeding goals of dairy worldwide. Nowadays, newer breeding objectives, notably to improve milk composition traits, are a response to the demand for a healthier human diet, and the basis of the milk-pricing system has shifted from a standard amount to milk composition, with a direct effect on farms' financial performance (Krovvidi *et al*., 2013).

More than 95% of the cow milk proteins are constituted by caseins and whey proteins. Among the caseins, beta casein is the second most abundant protein and has excellent nutritional balance of amino acids. To date 12 protein variants of β-casein (β-CN) known and out of these A1 and A2 are the most common and differ at amino acid position 67 with histidine (CAT) in A1 and proline (CCT) in A2 milk as a result of single nucleotide difference (Caroli *et al*., 2009). This polymorphism leads to a key conformational change in the secondary structure of expressed β-CN protein. Gastrointestinal proteolytic digestion of A1 variant of β-CN (raw/processed milk) leads to generation of bioactive peptide, beta casomorphin 7 (BCM-7). Infants may absorb BCM-7 due to an immature gastrointestinal tract whereas adults gather the biological activity locally on the intestinal brush boarder. In hydrolysed milk with variant A1 of beta-casein, BCM-7 level is 4-fold higher than in A2 milk (Elliott *et al*., 1999). Tropical native *Bos indicus* dairy cows produces 100% A2 type of milk and hence are a source for safe milk than exotic cattle breeds like Holstein which produces around 60% A1 milk that is considered to be responsible for many human disorders like Type1 diabetes, autism and heart diseases (Mishra *et al*., 2009).

The largest and diverse livestock herd in Ethiopia, makes it a resource with potential to contribute significantly to national development, including ending hunger, attaining food security, enhancing nutrition, and encouraging sustainable agriculture. CSA (2020/21), report showed that 70 million cattle population are estimated to be found in the country. They sustain the livelihoods of millions of people in pastoral, agro-pastoral and crop-livestock production systems, through the provision of both foods (*e.g*. milk and meat) and non-food products (*e.g*. skin and traction power). The indigenous cattle hold great promise and potential for milk production and constitute about 97.4% of the total cattle population (CSA, 2020/21). The importance of indigenous cattle and their superiority over exotic cattle is owed to their great adaptive capacity to tropical climate, excellent nutritional benefits and resistance to diseases (Mirkena *et al*., 2010; Lemecha *et al*., 2006 and Bradley *et al*., 1996). Recent census has witnessed an increase of 1.97% in population of dairy animals from 19.7 million to 22.5 million (CSA, 2020/21).

Different studies on genes influencing the genetic variation in milk performance traits in indigenous cattle breeds have gained importance in the recent times (Gothwal *et al*., 2022; Elzaki *et al*., 2022; Shah *et al*.,2021; Brym *et al*., 2005; Winter *et al*., 2002). The genetic variation in genes affects the physiological pathways and phenotype (Mohammed *et al*., 2015). Candidate gene analysis approach can aid to identify mutations and potential association with respect to economic traits. This type of strategy will help in revealing marker for productivity which can be exploited in implementing future marker/gene assisted selection, conservation and genetic improvement strategies (El-Magd *et al*., 2014; Geldermann, 1990).

Bovine milk synthesis is a very complex and hormone-triggered process regulated by number of genes. Many candidate genes such as *prolactin* (*PRL*), *casein, Fatty acid synthase (FASN), growth hormone receptor (GHR), β-Lactoglobulin (LGB),* and *diacylglycerol O-acyl transferase* 1 (*DGAT1*) with different functions in metabolism have been proposed to affect milk yield and composition of which few candidate genes have been identified on the basis of knock out trials.

The candidate gene, *diacylglycerol O-acyl transferase* 1 (*DGAT1*) located on the centromeric end of the bovine chromosome 14 in the quantitative trait loci region (QTLs) region (Farnir

et al., 2002) is considered to be directly responsible for 50% variation in milk fat content and milk yield in dairy cattle (Banos *et al*., 2008). *DGAT1* has strong functional and positional role in milk traits such as milk yield and fat content in cattle (Gothwal *et al*., 2022; Elzaki *et al*., 2022; Valenti *et al*., 2019; Furbass *et al*., 2006; Winter *et al*., 2002). The *DGAT1* gene has become a candidate gene after experiments have shown reduced or inhibited milk secretion in *DGAT1* knock-out mouse lines (Smith *et al*., 2000). A di-nucleotide polymorphism (AA/GC) in exon eight of *DGAT1* gene harbours a non-conservative lysine to alanine substitution (K232A) with profound effect on milk fat content (Gothwal *et al*., 2022; Elzaki *et al*., 2022; Valenti *et al*., 2019; Grisart *et al*.,2004; Grisart *et al*.,2002; Winter *et al*., 2002) . The lysine variant, in particular, has been linked to higher fat yield and fat and protein percentages, whereas the alanine variant has been linked to higher milk and protein yield (Gothwal *et al*., 2022; Elzaki *et al*., 2022; Valenti *et al*., 2019; Gurcan *et al*., 2019; Ramkaran *et al*., 2019; Patel and Chauhan, 2017; Dokso *et al*., 2015).

Among the seventeen exons of *DGAT1* gene,exon-8 has previously been reported to be the most polymorphic and potentially affect milk composition and yield traits (Gothwal *et al*., 2022; Elzaki *et al*., 2022; Grisart *et al*., 2002; Spelman *et al*., 2002; Winter *et al*., 2002). The diallelic *DGAT1* effect on milk production traits could be explained in part by the presence of multiple alleles at the *DGAT1* locus or other mutations in closely related genes (Bennewitz *et al*., 2004; Kuhn *et al*., 2004)**.** The *DGAT1 K232A* genotypes showed significant effects on milk composition traits in Sudanese Kenana, Butana, Butana-Holstein and Benin cattle breeds (Elzaki *et al*., 2022; Houaga *et al*., 2017; Rahamattalla *et al*., 2015).

Genetic diversity parameters estimates for *DGAT1* exon-8 region indicated haplotype number, haplotype diversity and nucleotide diversity for *Bos indicus* cattle were 2, 0.536 and 0.003, respectively (Faraj *et al*. 2020) and gene diversity for *Bos indicus* cattle range from 0.02-0.50 (Kaupe *et al*., 2004). Low and negative estimates of F_{IS} (inbreeding coefficient) was observed in Holstein crossbred, Borgou and Creole cattle of India, Benin and Uruguay cattle, respectively (Krovvidi *et al*., 2021; Houaga *et al*., 2017; Rincon *et al*., 2006), whereas positive estimates of F_{IS} was reported for Jersey crossbred of India cattle (Krovvidi *et al*., 2021). Pairwise F_{ST} values for pooled subpopulations showed least divergence for *Bos indicus* breeds with high milk fat percentage for *DGAT1* gene (Kaupe *et al*., 2004).

Another candidate gene, prolactin (*PRL)*, also called as "versatilin or omnipotin" hormone (Bern and Nicoll, 1968) is a single chain polypeptide hormone primarily transcribed in the specialized lactotrophs cells of the anterior pituitary gland and tissues like mammary gland (He *et al*., 2006). The hormone is known to have over 300 separate biological functions (Bole-Feysot *et al*., 1998). The bovine *PRL* gene was mapped on the 23rd chromosome at 43 cM close to a chromosomal QTL. Although is not a part of a QTL, the gene of its receptor is part of QTL for milk traits (Weller, 2009; Bennewitz *et al*., 2004). The *PRL* gene invariably and directly affects growth and development of mammary gland (mammogenesis), synthesis of milk (lactogenesis) (Dahl, 2008) and maintenance of milk secretion (galactopoiesis) (Dong *et al*., 2013). In addition, it is known that the binding of the *PRL* gene product with its receptor initiates a signaling cascade that activates the transcription of a number of genes, including expression of milk protein, lactose and lipid genes (Othman *et al*., 2011) via Janus kinases/signal transducers and activators of transcription (JAK STATs) signal transduction pathway (Hu *et al*., 2009). Reduction in serum *PRL* level was also found to be associated with depressing effect on milk protein and lactose levels (Singh and Ludri, 1999).

Hu et al.(2009) detected genetic polymorphism of *PRL* exon-4 gene in Chinese Holstein cattle and confirmed the A>G mutation at position 8398 and frequencies of allele A and G were found to be 0.894 and 0.106, respectively. SNP located in position 8398 of *PRL* exon-4 is the most popular genetic marker for genetic characterization of cattle populations (Brym *et al*., 2005). A study conducted in Chinese Holstein cows by Dong et al.(2013) identified eight different SNP's in *PRL* exon-4 gene and genotypes vary significantly with respect to the milk performance traits (p<0.05). Cows with genotype AA had a higher milk yield at 305 days (8457 ± 938 kg) in comparison to those with the AG (7537 ± 1278 kg; p<0.01) and GG (7757 ± 1174 kg; p<0.05). However, the percentage of fat and protein in milk did not differ between the genotypes (p> 0.05). Genotype effect was found to be significant (p<0.05) on protein and fat percent in Crossbred Holstein–Friesian cows whereas in Jersey cows effect was non-significant (p>0.05) on all quality traits (Shah *et al*., 2021). The study corroborates the finding in Russian Red Pied cattle, where the AA genotype also had a positive effect on milk yield (Alipanah *et al*., 2007). In contrast, the study conducted by Ghasemi et al. (2009) revealed negative impact of genotype AA on milk yield in Montbeliarde cows and the highest milk yield in Holstein cows with genotype AG had been observed. Similarly, Das

et al.(2012), studied genetic polymorphisms of exon-4(8398A>G) of *PRL* gene in Deoni cattle breed in india, genotype AG was associated with highest fat (4.743 ± 0.12) and protein (3.287 ± 0.03) percentages.

Gene sequencing has been considered as powerful method for identification of nucleotide sequence variation in amplified DNA and is considered as the gold-standard approach for genotyping analysis and expected to have almost 100% sensitivity (Laurie and George, 2009). Single nucleotide polymorphisms (SNPs) are the most often abundant genetic variation in vertebrates' genome. Approximately one of 1000 nucleotides in animal genome is expected to be a SNP site, accounting for more than 90% of all differences between animals. SNP identification and functional assessment are becoming an increasingly more important approach in molecular diagnostics and biology (Zhou *et al*., 2005). In addition, nowadays the SNP marker is becoming marker of choice compared with other DNA markers (Matukumalli *et al*., 2006).

The effect of any identified polymorphism may differ across different populations or breed because of specific genetic and evolutionary backgrounds. Before the aforementioned markers is used for the genetic improvement of cattle populations of Ethiopia, their polymorphism should be clearly studied. The polymorphic status of *DGAT1* and *PRL* genes and their association with milk yield and milk components has been reported in different populations as mentioned above, however, there is no study undertaken for major cattle populations of Ethiopia that are critically important for future breeding programs.

In the light of the above backgrounds, the present study aimed to characterize the genetic diversity of *DGAT1* exon-8 and *PRL* exon-4 genes, estimate the allelic and genotypic frequencies of *DGAT1 K232A* and *PRL* variants and assess the effect of *DGAT1 K232A* and *PRL* variants on daily milk yield and composition traits in five cattle populations of Ethiopia.

1.2. Statement of the Problem

The genetic assessment of milk yield and composition of resilient native populations could increase their potential values and improve production levels. The potential value of indigenous livestock breeds for milk production traits must be analyzed and conserved to maximize the milk production so that they can become a self-sustainable resource. Particularly, variations within the *DGAT1* and *PRL* genes have been shown to have significant effects on milk traits in exotic cattle. However, it has not been well investigated whether the known *DGAT1* and *PRL* mutations are segregating and associated with milk yield and milk composition traits in Boran, Begait, Horro, Fogera and Boran * HF cross populations in Ethiopia.

The nucleotide sequence variation of the bovine *DGAT1* and *PRL* genes in cattle populations of Ethiopia was not reported till now and this requires critical analysis of the two important genes responsible for milk yield and composition traits. Haplotypes for *DGAT1* and *PRL* genes were not constructed for cattle populations of Ethiopia and the effect of *DGAT1* and *PRL* genotypes on milk yield and composition traits was also not studied. There is no effort carried out so far to characterize these economically important milk genes and analyze their effects on milk production traits in promising cattle populations of Ethiopia. Therefore, the present investigation was undertaken to fill the aforementioned gaps.

1.3. Significance of the Study

Milk and milk products, both their quantity and quality are important for the economy of the country. Population growth, urbanization and changes in diets are creating a strong demand for milk and milk products. To meet this demand, keeping genetic diversity and doubling productivity through improvement of milk yield and quality of cattle population of Ethiopia is mandatory. Intervention on effective genetic improvement requires genetic information about the genetic variability and their effects on milk yield and quality traits.

This dissertation analyzed for the first time the nucleotide sequence variation of the bovine *DGAT1* exon-8 and *PRL* exon-4 genes in cattle populations of Ethiopia. The result of the study will contribute to better understand the polymorphisms of *DGAT1* and *PRL* genes and

their effect on milk yield and composition in major cattle populations. It needs to study the frequency of significant desirable variants related with candidate genes in cattle population of Ethiopia for its possible use in future selection decisions. The detection of mutations affecting dairy traits will facilitate the application of marker assisted selection (MAS) and provides insights into the genetics of milk traits in tropically adapted cattle breeds. Besides, it provides baseline information to apply marker assisted selection. Raise the market value of studied populations cow's milk in terms of its quality and attract livestock keepers and investors to adopt Ethiopian cattle.

1.4. Objectives of the Study

1.4.1. General Objective

To molecularly characterize bovine *DGAT1* exon-8 and *PRL* exon-4 milk linked genes and evaluate their effect on milk yield and composition traits in major cattle populations of Ethiopia

1.4.2. Specific objectives

- To assess the genetic diversity of *DGAT1* exon-8 and *PRL* exon-4 genes regions in major cattle populations of Ethiopia
- To estimate the allele and genotype frequencies of *DGAT1 K232A* and *PRL* variants in major cattle populations of Ethiopia
- To assess the effects of *DGAT1 K232A* and *PRL* variants on daily milk yield and composition traits in major cattle populations of Ethiopia

1.5. Scope and limitation of the Study

The cattle populations included in the study were: Boran (Large East African zebu), Begait (Large East African zebu), Horro (Zenga), Fogera (Zenga) and 75%F1=Holstein* (Boran * Holstein) (Sanga) cattle populations. These populations represent the major cattle group in Ethiopia (Rege and Tawah, 1999).

DGAT1 exon-8 and *PRL* exon-4 genes were considered for investigation in this study, the candidate genes were selected by combining different sources of information available on databases and public literature. The chosen genes were those that met at least one of the following criteria, proposed by researchers (Khatib *et al*., 2009; Ogorevc *et al*., 2009); 1) the gene encodes a protein directly associated with milk components or milk metabolism that exists in different genetic variants; 2) markers on the gene have proven effect on milk traits in other breeds (Elzaki *et al*., 2022; Li *et al*., 2021; Shah *et al*.,2021; Ozdemir, 2020); 3) The gene expression profile is related to the phenotype under study; 4) the gene is located within cattle QTL for the studied phenotype and expressed in bovine mammary gland.

The study addressed genetic variation in exon-8 of *DGAT1* and exon-4 of *PRL* genes, estimated the allele and genotypic frequencies of *DGAT1* exon-8 and *PRL* exon-4 genes and observed the effects of genotypes of *DGAT1 K232A* and *PRL* on milk yield and composition traits in Boran, Begait, Fogera, Horro and Boran * HF cross cattle populations.

Financial and logistic constraints limited the sequencing of the whole length of *DGAT1* and *PRL* genes and testing both the haplotype and measures of Linkage Disequilibrium with more sample size and performing functional validation of studied genes.

CHAPTER 2

2. LITERATURE REVIEW

2.1. Importance of *DGAT1* and *PRL* genes on milk yield and composition traits

Milk is one of the most vital and complete nutrition sources for mammals. Milk is a highly diverse fluid composed of mainly water, lactose, fat, proteins, organic acids and minerals (Fox and McSweeney, 2009). Some of the domesticated animals for instance cattle were employed for milk production for the humankind in the last ten millennia.

Genes encoding *DGAT1* enzyme and *PRL* hormone are considered as key links in the gene networks constituting the hereditary component of milk productivity. Gene disruption experiments have proved the critical role of *DGAT1* and *PRL* genes in lactation during knock out trial (targeted deletions) in mouse lines established the features of reduced or inhibited milk secretion. Therefore, the two potential candidate genes, *DGAT1* and *PRL* having crucial role on milk yield and composition traits were included in the present investigation.

Along with traditional methods, gene level studies for marker development may provide basis for conservation and possible use of native breeds as genetic resources to meet potential future demand of adaptation to changing environment or production needs. In dairy farm animals, the principal goal of the selection is the improvement of milk yield and composition. Currently, there is a considerable interest in the application of molecular genetics technologies in the form of specific DNA markers that are associated with various productivity traits to promote more efficient and relatively easy selection and breeding of farm animals.

2.2. Genetic Markers

At the level of gene, identifying economically important traits and characterizations of responsible genes for the economic traits could be possible. The markers interpreting variabilities at the DNA level are referred to as the molecular markers. The advancement of livestock genetics became possible with the advent of molecular markers. Molecular markers are detectable DNA sequences at specific positions of the genome, and conveyed from

generation to generation. The use of gene marker tests allowed breeders to improve livestock using the marker assisted selection (MAS) as part of their overall improvement programs (Hayes, 2007). Among the molecular markers, DNA-based markers are preferred to study polymorphism and diversity of livestock populations. The MAS method was used to eliminate deleterious gene alleles or select for favorable conditions based on some marker information (Eggen, 2012).

A single nucleotide polymorphism marker is a single base change in a DNA sequence, with a usual alternative of two possible nucleotides at a given position. For such a base position with sequence alternatives in genomic DNA to be considered as an SNP, it is considered that the least frequent allele should have a frequency of 1 % or greater. Although in principle, at each position of a sequence stretch, any of the four possible nucleotide bases can be present; SNPs are usually biallelic in practice (Martinez-Arias *et al*., 2001). The probability of two independent base changes occurring at a single position is very low. Another reason is due to a bias in mutations, leading to the prevalence of two SNP types.

The most precise method for search of unknown SNP is PCR analysis followed by direct sequencing. This technique is based on the amplification of specific parts of the genome in the PCR reaction and sequencing of the obtained product. A comparison between electrophoregram images of sequencing products is performed, which allows determining whether a mutation in a given region has occurred.

In animal genomes, SNPs are the most common types of sequence variations (Jiang *et al*., 2016). SNPs are base substitutions within nucleotide sequences in a strict molecular context, and their high density in the genomes of eukaryotes, including animals, has been of great importance in population genomics (Goddard and Hayes, 2009). Even though, SNPs are bi-allelic co-dominant molecular markers their high density allows for more quantitative knowledge on genome dynamics to be elucidated within a sample population than any other technique (Zhou *et al*., 2005).

These characteristics, combined with a low error rate, are allowing SNP markers to be used in a broader range of applications in livestock genetic architecture, such as precise identification of genomic regions that regulate economically important traits, and ultimately genomic

selection (Yaro *et al*., 2017). It is worthwhile noting that MAS is likely to compliment rather than replace the traditional breeding programs. MAS enables increased rate of genetic improvement through improved selection intensity, reduction of generation interval, accurate prediction and ease in selecting individuals early in life or for sex limited traits (Gholizadeh *et al*., 2008).

SNP analysis, more than any other molecular marker currently available, is better suited for high throughput genotyping that is needed to elucidate greater molecular insights such as historical signatures of selection, phenotypic variations within livestock breeds, and linkage disequilibrium over short physical distances (Jiang *et al*., 2016).

Mutation mechanisms result either in transitions: purine-purine (Adenine↔Guanine, A↔G) or pyrimidine-pyrimidine (Cytosine↔Thymine, C↔T) exchanges, or transversions: purine-pyrimidine or pyrimidine-purine (A↔C, A↔T. G↔C and G↔T) exchanges. A study comparing rodent and human sequences indicated a transition rate equal to 1.4 times that of transversions, implying that each type of transitional change was produced 2.8 times as often as each type of transversional change (Collins andJukes, 1994). The great advantage of this polymorphism is its universality in the genome between particular species and highly efficient identification of polymorphism within the tested sequence. On the other hand, the high cost of the analysis represents a disadvantage. The high density of SNP markers in the genome leads to their extensive utilization in the genetic and population analysis.

2.3. Structure and role of *DGAT1* Gene

Bovine *DGAT1* gene located on the centromeric region of the bovine chromosome 14, spans 8.5 kb and comprised of 17 exons and 16 introns, encoding a 489 amino-acid (Winter *et al*., 2002) (Fig 1). Exons vary in length from 39 to 191 bp with an average of 86.5 bp, whereas, the first two introns are 3.4 and 1.9Kb long, the remaining 14 introns measured only 92.4 bp on average with range of 65-215 bp.

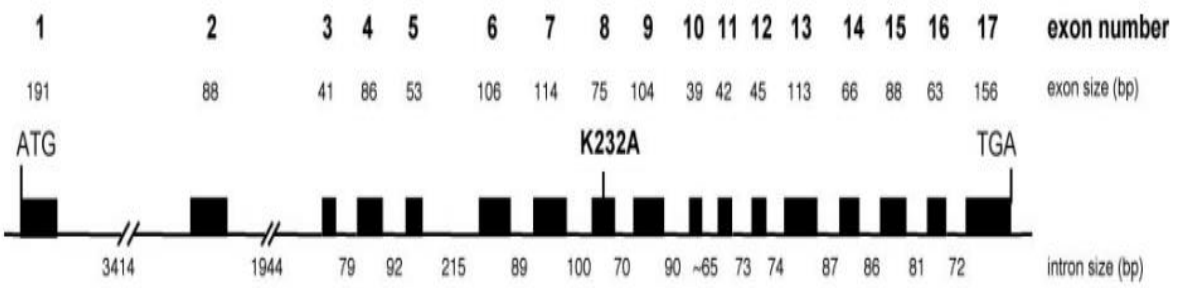


Figure 1. Structural organization of the bovine *DGAT1* gene

Exons are shown as rectangles. The sequence of intron 10 (65 bp) could not be completely resolved by sequencing. K232A indicates the only non-synonymous mutation. Source: Winter et al. (2002).

The *DGAT1* enzyme was found to play fundamental role in the metabolism of cellular triacylglycerol during physiological processes, such as intestinal fat absorption, lipoprotein assembly, adipose tissue formation and lactation (Cases *et al*., 1998). The *DGAT1* activity was first described by Weiss and Kennedy (1956). The final step in the synthesis in which diacylglycerol is transformed to triacylglycerol, is catalysed by the enzyme *DGAT1* (Yen *et al*., 2008).

Grisart et al.(2002) study showed a missense mutation *K232A* (Lys232→Ala) was shown to be significantly associated with variation in milk fat. Grisart et al.(2004) confirmed the functional and genetic causality of *K232A* in exon-8 of *DGAT1* gene for fat content through positional cloning approach and analysis of the expression of both alleles separately in virus expression systems.

The enzyme encoded by the K allele was characterized by a higher velocity rate (Vmax) in production of triacylglycerols than the A allele. The K allele of *DGAT1* gene have specific role in attachment of fatty acids to position 3, the only place on the triacylglycerol molecule where the C4 and C6 fatty acids are found whereas the A allele may be associated with a lower proportion of these short chained fatty acids. The mathematical prediction model of Shorten et al.(2004) also suggested an increase of 120% in acylation rate of *DGAT1* enzyme on substitution of alanine with lysine allele.

2.4. Genetic polymorphisms in *DGAT1* Gene

The haplotype analysis indicated that allele K of *DGAT1* gene appeared to be ancestral allele and the *K232A* substitution probably occurred early in the history of cattle domestication or even before domestication after the divergence of the *Bos indicus* and *Bos taurus* lineages (Winter *et al*., 2002). The reported *K232A* polymorphism in different breeds throughout the world is presented in Table 1. A wide range of distribution of the *K232A* polymorphism in 38 *Bos indicus* and *Bos taurus* breeds has been observed by Kaupe *et al*., (2004).

Table 1. Polymorphism of *DGAT1 K232A* gene in different cattle breeds

Breed	Genotype frequencies			Allele frequencies		References
	KK	KA	AA	K	A	
Sahiwal*HF	0.38	0.52	0.10	0.64	0.36	Ganguly et al.(2013)
Sahiwal	0.92	0.08	0.00	0.96	0.04	Ganguly et al.(2013)
Iranian HF	0.31	0.41	0.28	0.515	0.485	Ahani et al.(2015)
HF	0.09	0.30	0.61	0.24	0.76	Akyuz et al.(2015)
Crossbred HF	0.11	0.48	0.41	0.35	0.65	Molee et al.(2015)
Simmental	0.293	0.647	0.060	0.616	0.384	Dokso et al.(2015)
Brown Swiss	0.329	0.644	0.027	0.651	0.349	Dokso et al.(2015)
Polish HF	0.154	0.491	0.355	0.4	0.6	Komisarek and Kolenda(2016)
Kashmiri Jersey	0.44	0.33	0.23	0.6	0.4	Bhat et al.(2017)
Gir	0.85	0.15	-	0.925	0.075	Patel and Chauhan (2017)
Kankrej	0.79	0.21	-	0.895	0.105	Patel and Chauhan (2017)
Iranian HF	0.60	0.40	-	0.80	0.20	Gurcan et al.(2019)
Hardhenu	0.30	0.70	-	0.65	0.35	Ramkaran et al.(2019)
Iraq local cattle	0.40	0.30	0.30	0.6	0.4	Faraj et al.(2020)
Chinese HF	0	0.33	0.67	0.167	0.833	Li et al.(2020)
Kiwicross^{TM}	0.388	0.463	0.149	0.619	0.381	Li et al.(2021)
Butana	1	-	-	1	-	Elzaki et al.(2022)

HF: Holstein–Friesian

The information on allele frequency gives an indication on the genetic variance due to *DGAT1 K232A* polymorphism.

The nearly fixed nature of A allele of *DGAT1* gene was observed by Kaupe et al., (2004) in few of the *Bos taurus* breeds such as Belgian Blue, Gelbvieh, Hereford, Pinzgauer and Slavonian Syrmian breeds. Likewise, Kaupe et al.(2004) and Ganguly et al. (2013) reported very high *DGAT1* K allele frequency of 0.99 and 0.96 in Nellore and Sahiwal cattle, respectively.

2.5. Effects of *DGAT1 K232A* variants on milk traits

The effect of genes depends on genetic structure of the populations. The K to A amino acid substitution of a positively charged lysine residue with a neutral hydrophobic alanine residue in the *DGAT1* gene (Kaupe *et al*., 2007;Sanders *et al*., 2006) was reported to exert major effect on fat content and other milk characteristics (Koopaei *et al*., 2012; Farnir *et al*., 2002). The lysine coding allele K was found to enhance milk fat compared to alanine coding allele A (Gautier *et al*., 2007; Winter *et al*., 2002) and was reported to negatively affect the milk yield and protein yield (Berry *et al*., 2010). The *K232A* polymorphism was reported to have much larger additive effect for milk fat and sires selected for *DGAT1* KK genotype were observed to have higher breeding values for milk fat content and lower breeding values for milk yield and milk protein contents (Nowacka-Woszuk *et al*., 2008; Szyda and Komisarek, 2007).

The K allelic variant was found to negatively affect the amount of milk yield with concurrent increase in fat and protein content (Kaupe *et al*., 2004). This positive effect of K variant of *DGAT1* gene was reported by Thaller et al. (2003) with largest content of milk fat was observed for KK genotype (Leskova *et al*., 2013). A study in Butana and Friesian cattle indicated significantly higher milk and protein yield in heterozygous KA genotype than KK homozygous cows however non-significant variation for fat yield was reported for similar genotypes (Rahmatalla *et al*., 2015; Strzałkowska *et al*., 2005).

Although the effect of *DGAT1 K232A* alleles was established in the same direction in many cattle populations throughout the world but the magnitude of effects was variable between breeds (Suchocki *et al*., 2010), parity (Thaller *et al*., 2003) and during lactation (Szyda *et al*.,

2014). The seasonal effect of *K232A* genotype on some fatty acids in the milk was reported by Duchemin et al. (2013). Thus declined expression in summer season results in lesser polymorphism effect of *K232A* substitution (Komisarek and Kolenda, 2016).

In contrast, Dokso et al.(2015) observed the non-significant effect of K allele on the milk fat and milk protein levels in Holstein, Simmental and Brown Swiss animals. An opposite of the established findings was reported by Manga and Riha (2011) who demonstrated favourable effect of K allele on milk yield in Czech Holstein cows. Similar analogous results were reported in Hungarian Holstein (Anton *et al*., 2008) and Mexican Holstein (Hori-Oshima and Barreras-Serrano, 2003). The investigation of Rychtarova et al. (2014) in Czech Fleckvieh cattle demonstrated increase in estimated breeding value for protein and fat percentage for lysine variant of *DGAT1* gene.

Different genetic variations within the *DGAT1* gene can have major effects on milk yield and composition (Littlejohn, *et al.*, 2014). The effect of the mutations on milk yield and composition traits has been characterized in exotic cattle, however little is known on its effect on indicine cattle.

2.6. Bovine *prolactin* (*PRL*) Gene

The *PRL* hormone gene as one of the potential genetic marker and QTL region in dairy animals was found to be primarily responsible for the initiation and maintenance and regulation of lactation (Bernichtein *et al*., 2010) with key role in synthesis of milk components, including milk proteins, lactose, lipids (Dahl, 2008). Experimental trials in several animals revealed the expression of *PRL* gene in pituitary gland and several other sites including the central nervous system, immune system and the mammary gland (Ben-Jonathan et al., 1996; Sinha, 1995) which indirectly reflects its role in adaptation of animals to their surroundings.

2.6.1. Structure and role of *PRL* Gene

The bovine *PRL* gene has been cloned and mapped on chromosome 23 (Brym *et al*., 2005; Hallerman *et al*., 1988). Camper et al.(1984) detected five exons and four introns in *PRL* gene with an overall length of about 8.9 kb, encoding a 229 amino-acid *PRL* precursor (Fig 2). The

mature bovine *PRL* is composed of 199 amino acids along with a signal peptide of 30 amino acids (Cao et al., 2002;Wallis, 1974).

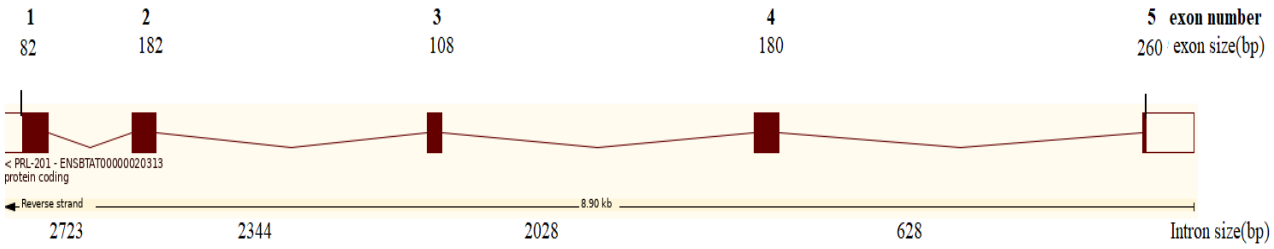


Figure 2. Structural organization of the bovine *PRL* gene
Exons are shown as rectangles shaded. Source: Modified from https://www.ensembl.org and Cao et al. (2002).

Subsequent research demonstrated that prolactin has two major roles in milk production: *PRL* induces lobuloalveolar growth of the mammary gland; alveoli are the clusters of cells in the mammary gland that actually secrete milk and stimulates lactogenesis or milk production after giving birth. *PRL* along with cortisol and insulin act together to stimulate transcription of the genes that encode milk proteins. The critical role of *PRL* in lactation has been confirmed in mice with targeted deletions in the prolactin gene. Female mice that are heterozygous for the deleted prolactin gene (and produce roughly half the normal amount of prolactin) show failure to lactate after their first pregnancy (Freeman *et al*., 2000).

It was observed in a trial that secretion, metabolism and distribution of *prolactin* hormone increases between days 30 to 150 of lactation (Collier *et al*., 1984). The *PRL* gene affects each stage of milk protein gene expression, i.e., transcription, translation, and post translational modifications of the proteins. The *PRL* gene is also a common mediator of the immune endocrine network, where nervous, endocrine and immune system communicate each other (Bole-Feysot *et al*., 1998).

2.6.2. Genetic polymorphisms in *PRL* Gene

Seven nucleotide substitutions based on the sequences of cDNA clones were detected in *PRL* gene (Sasavage *et al*., 1982) of which g.8398G>A nucleotide substitution has become the most

commonly used genetic marker for the genetic characterization of multiple cattle populations and associations with milk performance traits. However, Kaminski et al. (2005) identified five SNPs in *PRL* gene. Six SNPs were detected by Brym et al.(2005) through sequencing of three different SSCP patterns in exon-4 of *PRL* gene, one of which was shown to be associated with milk yield and fat content. SNP's g.8362T>C, g.8377G>C and g.8398G>A were observed in the exon-4 of *PRL* gene (Brym *et al*., 2005).

Sumantri and Jakaria (2015), studied genetic variability of *PRL* gene in Bali Cattle, frequencies of G allele were dominant to the A allele across all populations and heterozygosity values were found low in all population. Sonmez and Ozdemir (2017), studied *PRL* g.8398G>A gene polymorphism in East Anatolian Red cattle in Turkey, the population was in Hardy Weinberg equilibrium and heterozygosity was found at a medium rate as 0.338 and the calculated F_{IS} value was 0.072. Genotype and allele frequencies of the *PRL* gene across different cattle breeds throughout the world are presented in Table 2.

Table 2. Observed genotype and allelic frequencies of *PRL* g.8398A>G in cattle populations

Breeds	Genotype frequencies			Allele frequencies		References
	GG	GA	AA	G	A	
Deoni	0.1805	0.7222	0.0972	0.5416	0.4584	Das et al. (2012)
Czech Fleckvieh	0.77	0.22	0.01	0.88	0.12	Boleckova et al. (2012)
Sahiwal	0.72	0.18	0.1	0.81	0.19	Ishaq et al. (2012)
Achai	0.34	0.44	0.22	0.56	0.44	Ishaq et al. (2012)
Palestinian*HF	0.889	0.111	-	0.9445	0.0555	Khaizaran and Al-razem (2014)
Bali	0.75	0.25	-	0.875	0.125	Paramitasari et al. (2015)
EAR	0.59	0.34	0.07	0.76	0.24	Sonmez and Ozdemir (2017)
Vietnam HF	0.704	0.240	0.056	0.824	0.176	Thuy et al. (2018)
Turkey HF	0.56	0.38	0.06	0.75	0.25	Ozdemir (2020)
Kashmir Jersey	0.40	0.60	-	0.70	0.30	Shah et al. (2021)
Crossbred HF	0.26	0.50	0.24	0.51	0.49	Shah et al. (2021)

HF: Holstein–Friesian; EAR: East Anatolian Red

2.6.3. Effects of *PRL* gene variants on milk traits

Inconsistent and different results were observed by many workers across the globe for the effect of *PRL* gene on milk production performance which could be ascribed to breed or species difference (Brym et al., 2005). The inconsistent results might also be due to age dependent absorption of calcium under influence of *PRL*. SNPs occurring within the *PRL* gene may influence the chemical composition of milk or at least be an effective DNA marker of a sub region of dairy cattle genome (He *et al*., 2006).

Significant associations between *PRL* g.8398A>G polymorphisms and milk yield and constituents traits were observed by various scholars. Dybus et al.(2005) reported that different *PRL* genotypes were found to show variability in milk fat content and yield, and sum of fat and protein content during first lactation ($p \leq 0.01$). During third lactation, cows with AG genotype had higher fat and protein content (+0.44%) than AA individual which was incontrast with the study conducted by Dybus, (2002) who observed that cows with the AA genotype of the *PRL* RsaI gene had higher milk protein content than AG individuals. Chinese Holstein cows with genotype GG were observed to be high milk yielders than genotypes AA and AG ($p<0.05$) (Hu *et al*., 2009). The breed difference between Russian Black Pied and Red Pied cattle for *PRL* gene were reported by Alipanah et al.(2008).

2.7. Quantification of Genetic Diversity

Allele and genotype frequencies are considered as basic genetic parameters and their assessment could demonstrate the possible genetic differentiation of the population and their evolution over time in addition to comparison of one population with other populations in the sense that population with different genotype frequencies could have identical allele frequencies (Mayr, 1963). Such knowledge will be helpful in selection and mating strategy, development of knockout technology and understanding the structure, function and evolution of the gene. Some of the methods which help for the study of genetic diversity within a population are heterozygosity (expected and observed) estimates and allelic distribution; and they are good indicators of genetic polymorphisms within a population (Barker, 1994). Different parameters were suggested to measure the genetic variation within and between populations.

2.7.1. Heterozygosity/diversity estimators *[H_o, H_e]*

Observed heterozygosity [H_o]

Observed heterozygosity [H_o] is calculated directly from observed genotype frequencies for each locus and population/breed. H_o is less informative as a comparative measure of diversity since it is affected by inbreeding and other evolutionary processes violating the assumptions of Hardy Weinberg equilibrium.

$H_o = \sum n_i/N$, Where n_i is the number of heterozygous individuals at the i^{th} locus; N is the total number of animals sampled.

Expected heterozygosity [H_e]

Expected heterozygosity [H_e] is the expected proportion of heterozygous loci per individual. It is a composite measure that summarizes genetic variation at the allele level (Nei, 1978). The formulae for calculating H_e is derived right from the HWE. H_e, which is often referred to as genetic diversity is calculated as *i.e.*

$$He = 1-\sum_{i=1}^{n} p_{ik}^2$$

Where P_i is the frequency of allele i at locus k; n is the number of alleles at locus k

The heterozygosity is a measure of genetic variation within a natural population. High heterozygosity values within a natural population or breed denotes long term selection for adaptation, or historic mixing of different populations. *He* is also called as gene diversity (D) and accounts for both the richness and evenness of alleles in the population.

2.7.2. Nei's diversity index *[G_{ST}]*

Genetic distance describes genetic similarities and dissimilarities between two populations or individuals, and can be used for characterization of different breeds and for evaluation of the change in variation in species overtime (Naqvi, 2007). The genetic difference between populations is assessed based on the difference between allele frequencies distributions at several loci or based on allele size distributions (Laval *et al*., 2002).

Nei's (1973) statistic is one of the genetic distance measures with biological assumptions which are based on the evolutionary forces that could result in allelic variation. The two evolutionary forces, viz. mutation and genetic drift are the significant ones that are used to make assumptions or models. Nei's (1973) statistic, G_{ST}, is used to measure among population differentiation, relative to the total diversity: $G_{ST} = D_{ST}/H_T$, $0 \leq G_{ST} \leq 1$; $G_{ST} = 1\text{-}H_S/H_T$. The total genetic diversity at a locus ($H_T = 1 - \sum pi^2$, where pi is the mean frequency of the i^{th} allele in the pooled sample) is the sum of the mean genetic diversity within populations (H_s) plus the among-population genetic diversity (D_{ST}). G_{ST} values are often averaged over all polymorphic loci to estimate population divergence. All alleles are used in the calculation of H_T and H_s and, thus G_{ST} may be considered as a multiallelic form of Wright's F_{ST}, which was originally formulated for loci with two alleles. According to Schoen and Brown (1991), species with high interpopulation variation in allele frequencies with high G_{ST} are also likely to have high population to population heterogeneity (H_e).

2.7.3. Wright's F-statistics [F_{IS}, F_{ST}]

F-statistics developed by Curnow and Wright (1979) to summarize the genetic structure of a population and its subpopulations was used. The F-statistics (F_{IS} and F_{ST}) as measures of the fixation of alleles at different levels of organization (individuals, subpopulation, and population) is defined as follows:

Inbreeding coefficient (F_{is}): is the measure of reduction of heterozygosity of an individual as a result of non-random mating within its subpopulation (Widmer and Lexer, 2001). In another word, F_{IS} measures the extent of genetic inbreeding within sub-populations. The value of F_{IS} ranges from -1.0 when all individuals are heterozygous to 1.0 when there are no heterozygotes observed.

The value of (F_{IS}) fixation index indicates the level of inbreeding in the population probably as a result of long term selection and the maintenance of small group of closely related sires (Hansen *et al*., 2003). F_{IS} is a measure of the inbreeding coefficient, being low when genetic diversity was high ($F_{IS} < 0.05$) and medium or high when the heterozygosity was low ($F_{IS} > 0.10$). A bias towards heterozygote excess was inferred from the negative F_{IS} values that could be confirmed by the differences between observed and expected heterozygosity.

F_{IS} is estimated for populations which show significant deviation from the HWE and is significant for significant HWE estimation (Widmer and Lexer, 2001).

***F_{st}* fixation index**: is the mean reduction in heterozygosity of a sub-population due to genetic drift among sub-populations. It measures the extent of genetic differentiation among subpopulations. Its value can range from 0.0 when there is no differentiation to 1.0 whenever there is complete differentiation i.e. populations being fixed for different alleles.

$F_{ST} = H_T\text{-}H_S/H_T$, $0 \leq F_{ST} \leq 1$

Weir and Basten (1990) and Kalinowski (2002) had suggested the highest genetic distance (F_{ST}) to be higher than 0.25, moderate to be between 0.05 and 0.25 and the lowest estimate below 0.05. Frankham et al.(2002) also explained that a F_{ST} index of about 0.15 is considered to be an indication of significant differentiation among populations.

2.7.4. Polymorphic information content (PIC)

Initially derived by Botstein et al. (1980), polymorphic information content depicts the suitability of the markers and their primers used in the study for analyzing the genetic variability of a given population. Shete et al. (2000) also described polymorphism information content (PIC) value as most common measure of polymorphism for a marker locus for use in linkage analysis. Hence, microsatellite markers having greater than 0.5 PIC value are considered as highly informative and highly polymorphic (Marshall *et al*., 1998; Botstein *et al*., 1980).

2.7.5. Hardy-Weinberg equilibrium (HWE)

G.H. Hardy and Wilhelm Weinberg independently developed the concept that is known today as the 'HWE based on Mendel's principles of inheritance. The HWE states that both allele and genotype frequencies in a population remain constant or are in equilibrium from one generation to the next given that there are no disturbing influences such as non-random mating, mutation, limited population size, random genetic drift and gene flow. When this assumption of random mating is violated, the population will not have HWE. Violations of this provision include inbreeding, assortative mating and small population size and are known

to increase homozygosity. The remaining assumptions (selection, mutation and migration) affect the allele frequencies, but do not, in themselves affect random mating. If a population violates one of these, the population will continue to have HWE each generation, but the allele frequencies will change with these forces.

However, it is crucial to note that outside the lab one or more of these influences are in effect. Therefore, no real/natural population exactly satisfies all of the above stringent criteria simultaneously. The concept of genetic equilibrium is thus, an ideal state that provides a baseline to measure genetic change in populations across generations.

Testing for fit to HWE

Population is said to be in HWE for a given specific locus, when the observed and expected genotypic frequencies are not statistically different. There are different statistical tests used to test the allelic frequencies for fit to HWE. However, they can be grouped into three categories namely asymptotic, (Chi-square test and G-test, also known as log-likelihood test or a likelihood ratio test), exact (complete enumeration) and approximation (Markov Chain Monte Carlo). If the genotypic counts and/or the sample size are small, the asymptotic tests are not preferred as the distributions of the test statistics do not fit the chi-square distribution. In these situations, the exact test of HWE is chosen over the asymptotic tests (Guo and Thompson, 1992; McDonald, 2009).

CHAPTER 3

3. MATERIALS AND METHODS

3.1. Description of the study areas

The study was conducted on five cattle populations of Ethiopia from Oromia, Tigray and Amhara National Regional States (Fig 3).

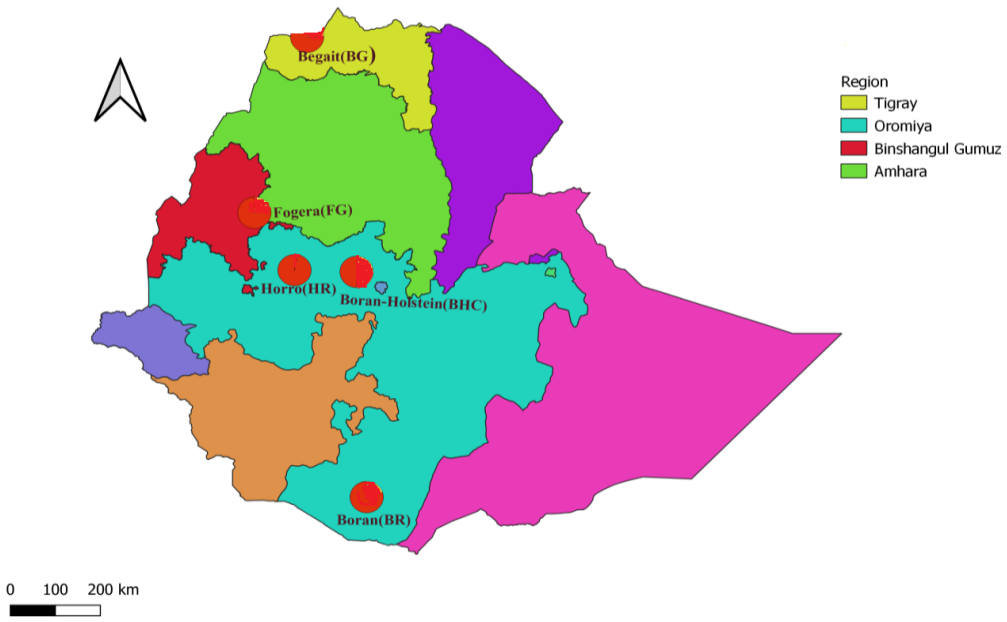


Figure 3. Map of Ethiopia showing the sampling sites of studied populations based on National Regional States. The map was constructed using the QGIS-OSGeo4W v. 3.8.0 software

The Boran cattle were sampled from Borena zone, Yabello district specifically from: Dida Tuyura Boran Cattle Breeding and Improvement Ranch, which is situated at about 550 km South of Addis Ababa and 20 km North of Yabello town with geographical coordinates of 04º53´N and 38º30´E. The mean annual temperatures vary from 19ºC to 24°C and show little seasonal variation. The area has an arid to semi-arid climate; with an average annual rainfall ranging from 110 to 600 mm. The rainfall distribution is bimodal, but erratic and unreliable.

About 59% of annual precipitation occurs from March to May and 27% from September to November (Haile *et al*., 2011).

Begait cattle are sampled from Humera Ranch of Begait Cattle Multiplication, Improvement and Conservation Center which is located in Kafta Humera districts of Western Zone of Tigray National Regional State, located at about 600 km West of Mekelle city and 954 km North of Addis Ababa. Humera Ranch lies within the coordinates of 14°28'N of latitude and 37°31'E of longitude and has altitude range of 515 to 1863 m.a.s.l. The annual rainfall of the area varied from 449 mm to 1100 mm and annual temperature of 33 to 41.7°C (Niguse and Aleme, 2015).

Horro cattle were sampled from Horro Guduru Cattle Breeding and Improvement Ranch, which is situated in Horro Guduru zone at about 300 km west of Addis Ababa, with geographical coordinates of 09º34´N and 37º06´E, and at an altitude of approximately 2296 m.a.s.l. The monthly mean temperature observed in the area was 14.9 ºC to 17.5 ºC. The area has one long rainy season extending from March to mid-October with annual rainfall ranging from 1000 – 2400 mm (Mwai *et al*., 2015).

Fogera cattle were sampled from Chagni Cattle Breeding and Improvement Ranch. The center is situated in the Guangua district of Awi Zone in Amhara National Regional State and is situated at about 505 km North-west of Addis Ababa. It lies at latitude of 10°55'N and 36°26' longitudes East of the meridian and an altitude of 1,829.5 m.a.s.l. The ranch receives an average annual rainfall of 1,730 mm; the temperature ranges from 13.7 to 29.5°C. The rainfall distribution is bi-modal (Almaz, 2012).

Boran * HF crosses were sampled from Holeta Agricultural Research Center which is located in the central highland of Ethiopia at 35km west of Addis Ababa at latitude of 9°3'N and 38°30'E longitudes and an altitude of 2400 m.a.s.l. The area has an average annual rainfall of 1100 mm and an average temperature of 15°C with minimum 6°C and maximum 24°C (Gojam *et al*., 2016; Gebreyohannes *et al*.,2013).

3.2. Study populations, sampling technique and sample size

Boran: Predominantly distributed in the semi-arid and arid areas of Southern Ethiopia, Northern Kenya and South Western Somalia and known for their fast growth, high milk yield and heavy muscling that needs calm environment (Haile *et al*., 2011; Rege and Tawah, 1999).

Begait: Reared in Tahtay Adiabo, Kaftahumera, Welqait, Asgedet Simbla districts of northwest Ethiopia. Mostly reared for milk and meat production (Teweldemedhn and Selam, 2020; Assefa and Hailu, 2018; Zerabruk *et al*., 2007).

Fogera: The breed inhabits the northwest highlands of Ethiopia in the surrounding area of Lake Tana and mainly reared for draft and dairy production (Mwai *et al*., 2015; Zerabruk *et al*., 2007; Rege, 1999).

Horro: Represents an ecotype adapted in the highland agro-ecological zones of the country with the main habitat in the Horro Gudru area of Wollega as well as neighboring parts of Western Showa and Illubabur and mainly characterized by a calm disposition and variable milk production in a mixed crop–livestock system (Mwai *et al*., 2015; Edea *et al*., 2012; Mekonnen *et al.*, 2012; Rege, 1999).

Boran * HF crosses: Boran * HF cross (75% HF: 25% Boran) which is the 50% F1 is back crossed with pure Friesian semen to produce the 75% HF. The Boran cattle used as a foundation stock for crossbreeding were brought from Borana pastoralists in the southern Ethiopia (their center of origin) and reared on station at Holetta Agricultural research center (Haile *et al*., 2011).

Animals were selected randomly from each population with exclusion of closely related individuals based on pedigree data. Lactating cows with a minimum of 120 days lactation and within parity three were included in the study. Apparently, healthy animals were identified based on their health record for the present study with no history of any chronic ailments. The sampled animals were raised on natural grazing without concentrate supplementation and they were sampled at the same season (March to May in 2020 G.C) eliminating the effect of season.

Hale et al. (2012) described that variability in allele frequency and expected heterozygosity among replicates decreased with increasing sample size, but this decrease was minimal above sample size of 25 to 30 for microsatellite based population genetic studies. However, in the current study as SNP markers which are considered to comprise more information are utilized and therefore; sample size could be decreased below suggested level for microsatellite markers, accordingly based on this hypothesis, samples taken below the range indicated to obtain all informative alleles at frequencies that are representative of the population allele frequencies. Ninety two animals comprising of five cattle populations of Ethiopia (Boran = 17; Begait = 16; Fogera = 18; Horro = 17 and Boran * HF cross = 24) were included in this study.

Based on the type of analysis employed and the sequence region covered different number of samples were used. For genetic analysis of *DGAT1* exon-8 gene, a total of eighty-nine animals comprising; Boran (*n* = 17); Begait (*n* =16); Fogera (*n*=17); Horro (*n* =16) and Boran * HF crosses (*n* = 23) were considered. To assess the effect of *DGAT1 K232A* on milk traits ninety-two animals including; Boran (n=17), Begait (n=16), Fogera (n=18), Horro (n=17) and Boran * HF cross(n=24) cows were used. To detect SNPs and assess their effect on average daily milk yield and milk components in exon-4 of *PRL* gene, eighty-seven animals including; Boran (n=18), Begait (n=17), Fogera (n=18), Horro (n=17) and Boran * HF crosses (n=17) were considered.

3.3. Blood sample collection and DNA extraction

Blood samples were collected from the selected animals representing the five cattle populations. The blood samples were drawn out from the tail head with a volume of 4 ml under aseptic conditions and gently mixed with ethylene diamine tetra acetic acid (EDTA.K3) vacutainer tube (Improvacuter®, Germany) and placed into an ice box containing ice. Collected samples were brought to Bio and Emerging Technology Institute laboratory at Addis Ababa, where it was stored at -20^{0}C until DNA was extracted. Extraction of genomic DNA was carried out using modified salting-out extraction procedure according to Nasiri et al. (2005).

Briefly, blood samples were first thawed at room temperature and about 500µl blood was poured into a 2ml centrifuge tube. Exactly 800µl of lysis buffer (0.3M sucrose, 0.01 M Tris

HCl, pH 7.5, $MgCl_2$ (5mM) and 1% Triton X100) were added to each centrifuging tube and blood samples were mixed gently by inversion. Then it was centrifuged for 5 min at 4725rpm. The supernatant was discarded carefully and the step was repeated until a white pellet was obtained. The pellet was vigorous vortex and re-suspended with the 60µl of 10mM trisHCl pH 8 and centrifuged again at 2500rpm for 2min. After discarding the supernatant, the pellet was re-suspended in 66µl 10mM tris HCl, 66µl laundry powder solution (30mg/ml laundry powder solution) and glass beads and vortexed for 2 min. Exactly 50µl of 6M NaCl was added and vortexed again for 20 seconds, then centrifuged for 5min at 11573rpm and the supernatant was transferred into 2ml Eppendorf tube and 150µl of 96% ethanol was added and the solution was mixed by inversion and centrifuged for 3min at 13,000rpm until the genomic DNA was precipitated. The precipitate was then rinsed twice with 100µl of 70% ethanol in 1.5 ml Eppendorf tube by inverting the solution for 5 minutes and centrifuged for 2min at 12,000rpm. Then ethanol was carefully poured off using tissue paper. The DNA was allowed to air dry for 30 minutes. Finally, the extracted DNA was allowed to dissolve in 60µl TAE buffer (10 mM tris-HCl pH 8; 0.1mM EDTA, pH 7.4) overnight and stored at 4^0C.

3.4. Genomic DNA Quantity and Quality Measurements

The quantities and quality of the extracted genomic DNA samples were measured by NANODROP spectrophotometry (GeneQuant) and agarose gel electrophoresis. Optical density (OD) at 260/280 ratio, an indicator of purity and quantity was observed in the range of 1.6 to 2.05 which indicated the suitability of the extracted DNA samples for hassle free *in vitro* amplification (Annex Fig 1). DNA samples above 1.6 were further analyzed by 1.5% agarose gel electrophoresis to check their qualities. Briefly, 2 µl of DNA samples were mixed with 2µl of 6X loading dye and loaded onto the gel and allowed to run for 40 minutes. After completion of electrophoresis, the gel was transferred and visualized under UV transilluminator (UV tech). Those samples showing thick single bands were considered as good quality DNA and were used for subsequent polymerase chain reaction and sequencing.

All samples employed in the present investigation were devoid of fragmentation as evidenced by the absence of smearing on gel and presence of intact bright genomic DNA band (Annex Fig 2).

3.5. Primers Employed

To amplify the 278bp product size of *DGAT1* exon-8 region, primers were designed with accession number AJ318490 as reference sequence using primer 3 plus software (Rozen and Skaletsky, 2000). The region was amplified by polymerase chain reaction (PCR) using the primers: Forward: 5'-AAGGCCAAGGCTGGTGAG-3' and Reverse: 5'-GGCGAAGAGGAA GTAGTAG-3'. While *PRL* exon-4 region of 294bp was amplified by PCR using previously published primers; Forward: 5'-CCAAATCCACTGAATTATGCTT-3'; Reverse:5'-ACAGAAATCACCTCTCTCATTCA-3'(Brym *et al*., 2005).

Primer sequences which were newly designed and adapted from publication were sent to south Korea for preparation at Macrogen company and returned back after preparation in lypholized powdered form and dissolved using PCR grade water to a final concentration of 100pmol/µl according to the manufacturer's instruction and required pmol/µl of primer stock solution was prepared for each gene and stored at -20^{0}C.

3.6. PCR Reaction and Amplification

A 278bp fragment of exon-8 of *DGAT1* and 294bp region of exon-4 of *PRL* genes were amplified by PCR in a final reaction volume of 25 µl separately (Annex Figures 3 and 4). The master mix was prepared on ice for ninety-two samples for each locus separately and 23 µl of the master mix was poured into each reaction tube and 2µl of template DNA was added to make the final volume.

For both *DGAT1* and *PRL* genes the PCR was carried out in a total volume of 25µl containing, 5X PCR buffer (5µl), 1.5mM MgCl2 (3µl), 10mM dNTP's mix (1µl), forward primer 10pmol/µl (0.5µl), reverse primer 10pmol/µl (0.5µl), genomic DNA 25ng/ µl (2µl), Taq DNA polymerase 5U/µl (0.3µl) and DNase free water (12.7µl). The optimized thermal profile (DW-K960-Thermal Cycler) include an initial denaturation at 94°C for 3 minutes, 30 cycles of denaturation at 94°C for 1 minutes, annealing at (*DGAT1*=57°C and *PRL*= 57.5°C) for 45 seconds, elongation at 72°C for 1 minute and a final extension at 72°C for 7 minutes. The size of the amplified DNA fragments was assessed by comparison with standard molecular weight marker (Annex Figures 3 and 4).

3.7. Electrophoresis and PCR product cleaning

To check amplification, 5 µl of the PCR product was loaded on 1.7% agarose stained with gel red prepared by dissolving 0.85g of agarose in 50ml 1XTAE buffer. Electrophoresis was carried out at 80 V for 45 minutes. After completion of electrophoresis, gel picture was taken under UV trans-illuminator by Biodoc Analyse 2.0 with a digital cannon camera.

Since the PCR reaction mix might contains primers, salts, dNTPs and proteins, it was cleaned-up from these constituents to avoid their possible interference during sequencing. Therefore, PCR products were cleaned using shrimp alkaline phosphatase (SAP) and exonuclease I (exo I; Thermofisher Scientific) treatment Kit to recover/concentrate DNA fragments.

3.8. DNA sequencing

Before sequencing, sequencing reaction was performed for the PCR products by using one of a pair of PCR primers used for amplification of *DGAT1* and *PRL* genes separately. A forward primer was used for *DGAT1* gene (278bp) while reverse primer was used for *PRL* gene (294bp) as their size was short. After completion of the reaction, reaction products of each gene PCR products were purified using a sodium acetate-ethanol purification method. The purified products of sequencing reactions of each gene were analyzed on an ABI3730 capillary genetic analyzer (Sanger Sequencing Machine) according to the manufacturer's instructions at Konkuk University of South Korea.

The quality of the sequences is assessed by Phred scores with CodonCode Aligner software. Phred quality scores are logarithmically linked to error probabilities. The phred quality score is the negative ratio of the error probability to the reference probability level of 1 expressed in Decibel (dB). Phred scores of more than 30 are considered as good quality. Quality sequences were obtained for all populations for the *DGAT1* and *PRL* genes (Annex Figures 5 and 6). The poor quality ends were trimmed and base calling was manually verified.

3.9. Submission of obtained sequences to database

Overall ninety two sequences were generated for *DGAT1* exon-8 region and eighty seven *PRL* exon-4 genes under different populations. After proper annotation of haplotypes/sequences of

both genes the online BankIt submission tool of GenBank database was used. Accordingly, accession numbers of *DGAT1* exon-8 gene (n=25) BHC (ON262825-ON262828), BR (ON262829-ON262833), BG (ON262834-ON262838), FG (ON262839-ON262844) and HR (ON262845-ON262849) and accession numbers of *PRL* exon-4 gene (n=18) BHC (ON677410-ON677412), BR (ON677413-ON677416), BG (ON677417-ON677419), FG (ON677420-ON677424) and HR (ON677425-ON677427) were deposited in the GenBank database.

3.10. Collection of milk samples and Analyses

Early morning and late afternoon 50ml of milk sample were collected from the ninety two lactating cows from which the blood was collected in clean, dry, grease free and labeled milk collection vials for milk composition analysis. The collection of milk samples was replicated on the same test day of each week for one months to obtain an average value of daily milk yield, milk fat and other compositional traits. Daily milk yield was collected for four days per month as the total of morning and afternoon milk production and its average was calculated per each individual cow. For milk composition analyses milk samples stored at 4°C cold chain were sent to Holeta dairy research center. Milk samples were initially kept at a constant temperature of 37ºC in water bath for 30 minutes. The samples were then properly mixed and homogenized before analysis. The different fractions of milk samples, viz., milk fat, protein, lactose and SNF were then analyzed through ultrasonic milk analyzer (Lactoscan SL 30, MB Ver.60).

3.11. Data management and statistical analysis

3.11.1. Analysis of sequence data and detection of SNPs

Prior to analysis, all the chromatograms were visualized and sequence fragments were edited using Bio-edit version 7.0.5.3. Sequence data obtained from the forward primer of the *DGAT1* and reverse primer of the *PRL* genes were analyzed to detect the polymorphic sites. Analysis of homology of samples sequences with database reference (acc no. AJ318490 and AF426315) for *DGAT1* and *PRL* genes, respectively and other mammalian species in gene bank was performed using Basic Local Alignment Search Tool (BLAST). Multiple sequence

alignment for each of the sequenced region, viz., exon-8 region of *DGAT1* gene and exon-4 of *PRL* gene was carried out using Clustal W software package.

Genetic diversity at sequence level was performed encompassing partial intron-7, exon-8 and intron-8 region of *DGAT1* gene. DNA polymorphism, Ne's genetic distance and Ka/Ks ratio were computed using DnaSP V6 (Librado and Rozas, 2009). To assess for past population expansion of the *DGAT1* gene, two statistical tests Tajima's *D* (Tajima, 1989) and Fu's *Fs* (Fu, 1997) were used. The analyses were implemented in the program ARLEQUIN software version 3.5.2.2 (Excoffier and Lischer, 2010) and *p*-values were generated using 1,000 simulations under a model of infinite site neutrality.

For both genes, values for the observed heterozygosity (Ho), expected heterozygosity (He), polymorphic information content (PIC), minor allele frequency (MAF), coefficient of inbreeding and test for HWE were calculated using Power Marker (version 3.25) (Liu and Muse, 2005). DNA sequence variants (SNPs) naming followed the Mutnomen nomenclature (http://varnomen.hgvs.org/.). Population differentiation due to the population genetic structure was also assessed from sequence data, population pairwise Wright's F_{ST} values (Weir and Cockerham, 1984) and number of haplotypes were calculated by using ARLEQUIN software version 3.5.2.2 applying 1000 replication values (Excoffier and Lischer, 2010). Pairwise F_{ST} graph and the average number of pairwise differences were generated and computed using the R package (https://cran.rproject.org/bin/windows/base/).

Moreover, fifteen *DGAT1* gene haplotypes/sequences of milk producing farm animals from Genbank including *Bos taurus* (AJ318490, EU077528, MF069174 and MF445056), *Bubalus bubalis* (MZ230553, MZ230553, MF069172 and KX965992), *Camelus dromedaries* (MF069170 and MF069171), *Capra hircus* (LT221856 and FJ415876) and *Ovies aries* (KJ918741, FJ415875 and EU178818) from Germany, Turkey, India, Iran and Benin were accessed at April 28, 2022; and included in the analysis.

3.11.2. Phylogenetic Analysis

Phylogenetic tree was constructed using the Neighbor-joining (NJ) method (Saitou and Nei, 1987) of bootstrap test of phylogeny in MEGA 11 (Tamura *et al*., 2021) software to evaluate the evolutionary relationships of studied populations with other farm animals including *Bos*

taurus, *Bubalus bubalis*, *Camelus dromedarius*, *Capra hircus* and *Ovies aries*. The evolutionary distances were computed using the Tamura-Nei model (Tamura and Nei, 1993) for *DGAT1* having minimum Bayesian Information Content (BIC) value of (2470.1764) was selected and Kimura 2-parameter model (Kimura, 1980) for *PRL* having minimum BIC value of (755.4788) was selected out of 24 different maximum likelihood substitution models via implementing 1000 bootstrap values (Nei and Kumar, 2000). Positions from DNA sequences containing gaps or missing data including identical sequences were excluded from the analysis.

Phylogenetic analysis of *PRL* gene haplotypes/sequences of milk producing farm animals, sequences from Genbank including *Bos taurus* (AF426315, AH013356 and JF826522), *Bubalus bubalis* (EU340420 and KX685940), *Capra hircus* (EU256165 and EU256167) and *Ovies aries* (KC764410 and LC718487) from China, Poland, Egypt, India, Cyprus and Iraq were included in the analysis.

For both genes median-joining network (MJN) tree was constructed using the NETWORK software. To evaluate the median network, the nucleotide sequences were first converted into binary data, while identical sites were omitted from the analysis. Each split was programmed as a binary character, satisfying the values of zero and one. The haplotypes were denoted as a binary vector in this method. The median vectors were estimated for each vectors until the construction of the median network was achieved (Bandelt *et al*., 1999).

3.11.3. Statistical and association Analysis

For both genes, allele and genotype frequencies were calculated using Power Marker (Liu and Muse, 2005). Standard error of allelic frequency was calculated as $[p\ (1\text{-}p)/2n]^{1/2}$ where n is the sample size and p is the frequency of allele (Spiess, 1977).

The General Linear Mixed Model (GLM) was used to estimate qualitative effect of *DGAT1 K232A* and *PRL* g.8398A>G genotypes on daily milk yield and milk components, Difference between means was separated using Tukey HSD test with SAS (SAS, 2011). The breed type and genotypes were fitted as fixed variables, while the observed traits: daily milk yield, fat%, protein %, SNF % and lactose % were fitted as dependent variables.

The effects of fixed factors were fitted in the statistical model described below:

$Y_{ijk}= \mu+ B_i + GD_j + GP_k + (B \times GD)_{ij} + (B \times GP)_{ik} + \varepsilon_{ijkl}$ Where:

Y_{ijk}= observed trait (daily milk yield, fat%, protein %, SNF % and lactose %) on the n^{th} cow of the i^{th} breed and the j^{th} genotypes

μ = is the population mean

B_i= fixed effects of i^{th} breed (i = 5)

GD_j= fixed effects of *DGAT1* genotypes (KK, KA and AA)

GP_k= fixed effects of *PRL g.8398A>G* genotypes (GG, GA and AA)

$(B \times GD)_{ij}$ = breed by *DGAT1* genotypes interaction effect

$(B \times GP)_{ik}$ = breed by *PRL* genotypes interaction effect

ε_{ijkl}= is the random residual associated with each record

Average allele substitution effects of one copy of K allele of *DGAT1 K232A* and one copy of G allele of g.8398A>G genotypes were also calculated. The linear model was used with the replacement of genotype effect by the linear regression on the number of desired allele: $Y_{ijk} = \mu + b.x_i + e_{ij}$

Where Y_{ij} is a phenotypic observation (milk traits) of the animal, μ is the overall mean, x_i is the number of desired alleles (0, 1, and 2) and b is the regression coefficient representing the allele substitution effect, and e_{ij} is a random residual effect.

Probability less than 0.05 was used to determine the level of significance. Bonferroni correction was performed to account for multiple comparisons. The results of the different effects are presented as least squares means ± standard error.

3.12. Ethics approval

Data exchange was in accordance with national and international regulations. In addition, Institutional Ethical Review Board of Adama Science and Technology University approved the experimental design and animal data collection for the present study (certificate reference number RECSoANS/BIO/07/2021). All methods were carried out in accordance with relevant guidelines and regulations.

CHAPTER 4

4. RESULTS

4.1. Genetic diversity of *DGAT1* gene in the studied cattle populations

4.1.1. SNP's analysis of *DGAT1* gene in the studied cattle populations

The multiple sequence alignment (MSA) of generated sequences was carried out to detect SNPs within the target *DGAT1* gene in amplified fragments through comparison with reference sequence of cattle (GenBank Accession No.AJ318490) (Fig 5). The *DGAT1* sequence fragment analyzed was confirmed to cover a part of intron-**7** (38 bp), the whole exon-8 (75 bp) and intron-8 (70 bp). Across the five study populations, ten SNP's which fulfill SNP criteria were detected at 10433th, 10434th, 10522th, 10525th, 10528th, 10530th, 10531th, 10536th, 10540th and 10543th positions of *DGAT1* region, where the majority of which (8) were located in the intron-8 region and only 2 of them in the coding exon-8 region of *DGAT1* gene (at 10433th and 10434th positions) (Fig 4).

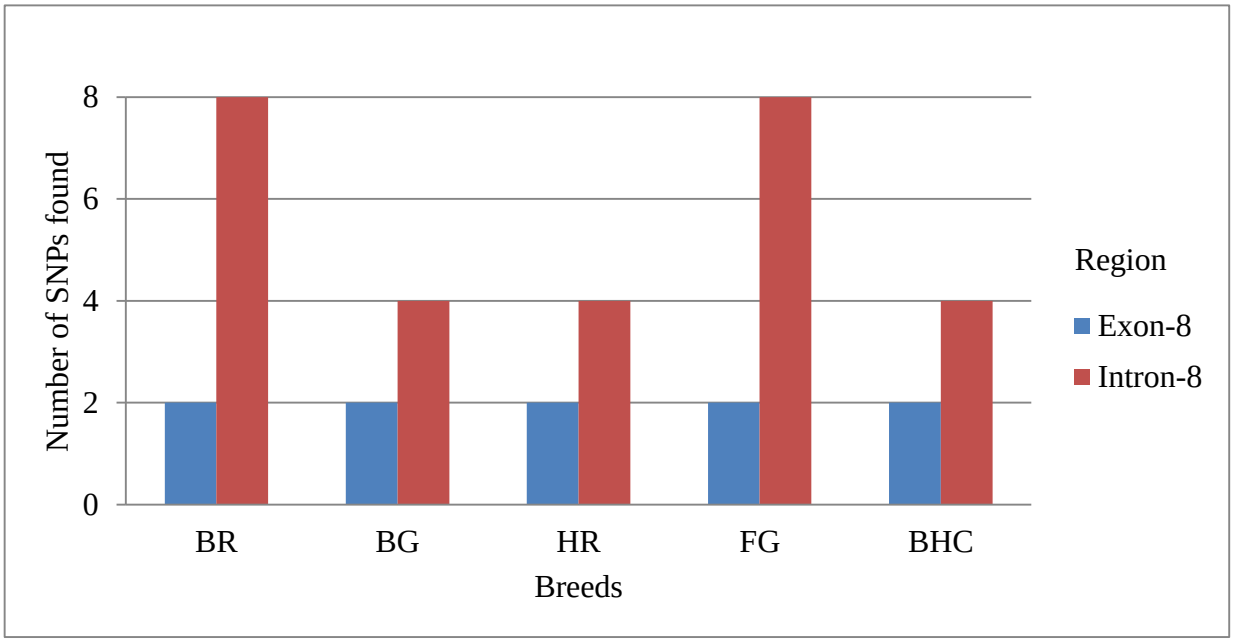


Figure 4. Proportion of exonic and intronic *DGAT1* SNPs in five cattle populations

All SNPs were seen in Boran and Fogera cattle populations. SNPs g.10531G>C, g.10536G>T, g.10540T>C and g.10543T>C were not seen in Begait, Boran * HF cross and Horro cattle populations. The SNPs observed at 10433th, 10540th and 10543th positions were transitions in nature, whereas the SNPs detected at 10434th, 10522th, 10525th, 10528th, 10530th 10531th and 10536th positions were transversion mutations (Fig 4).

Figure 5. *DGAT1* nucleotide sequences alignment of studied breeds with the reference sequence (Ac.No.AJ318490), dot represents similar nucleotides with the reference sequence ID from Genbank. Accession numbers: Boran * HF crosses (ON262825 - ON262828), Boran (ON262829 - ON262833), Begait (ON262834 - ON262838), Fogera (ON262839 - ON262844) and Horro (ON262845 - ON262849)

4.1.2. DNA polymorphism and genetic diversity parameter estimates

The number of polymorphic sites (*S*) which is the measure of usable loci that show more than one allele per locus was analyzed. Accordingly, the value of *S* was higher in Boran (10) and lower in Begait (4). Horro and Boran * HF crosses had the same number of polymorphic sites (6). The nucleotide diversities (π) were relatively higher in the Begait breed, whereas Horro breed had the lowest $\boldsymbol{\pi}$ value (Table 3). For all breeds, non-synonymous substitutions were less as compared to the synonymous substitutions (Table 3). The number of haplotypes detected ranged from 6 in Fogera to 4 in Boran * HF crosses. Boran, Begait and Horro cattle had the same number of haplotypes. The highest haplotype diversity (Hd) was observed in Begait, whereas Horro breed exhibited the lowest haplotype diversity (Table 3).

Table 3. Number of polymorphic sites (S), haplotypes per population (Hp), haplotype diversity (Hd), nucleotide diversity (π); and standard deviations of sampled populations for *DGAT1* gene

Populations	S	Hp	Hd ± SD	π ± SD
Boran*HF cross	6	4	0.636 ± 0.061	0.009 ± 0.004
Boran	10	5	0.507 ± 0.140	0.009 ± 0.006
Begait	4	5	0.700 ± 0.080	0.010 ± 0.003
Fogera	9	6	0.515 ± 0.145	0.009 ± 0.006
Horro	6	5	0.450 ± 0.151	0.005 ± 0.004

For analysis of genetic diversity within a population, minor allele frequency (MAF), polymorphic information content (PIC), average expected (H_e) and observed heterozygosity (H_o) values and, F_{IS} coefficient were estimated. The observed heterozygosity (H_o) ranged between 0.157 in Boran * HF cross to 0.412 in Begait cattle (Table 4). The expected heterozygosity (H_e) ranged from 0.120 (Boran * HF crosses) to 0.256 (Begait). Boran * HF cross had the lowest number of PIC, whereas the highest numbers of PIC was noted in Begait. MAF values ranged from 0.078 in Boran * HF crosses to 0.206 in Begait cattle. The different sequences of *DGAT1* gene were observed in heterozygote excess in studied breeds that ultimately lead to low and negative F_{IS} values (Table 4).

Table 4. Genetic parameters estimates for the considered breeds

Breeds	*N*	MAF	H_o	H_e	PIC	F_{IS}
Boran	17	0.129	0.259	0.176	0.148	-0.446
Begait	16	0.206	0.412	0.256	0.200	-0.591
Fogera	17	0.153	0.306	0.196	0.163	-0.538
Horro	16	0.119	0.238	0.146	0.118	-0.606
Boran * HF cross	23	0.078	0.157	0.120	0.101	-0.283

Number of sequences (*N*)

4.1.3. *DGAT1* Neutrality test

Fu's *Fs* was negative and non-significant (p>0.10) in Fogera and Horro cattle, while it is non-significant positive in Boran and Boran * HF cross. The Tajima's *D* value obtained was negative and non-significant (p>0.10) in Boran, Fogera, and Horro, while positive in Boran * HF cross. Both tests are positive and significant in Begait cattle (Table 5).

Table 5. Neutrality tests for the considered breeds

Breeds	*N*	Fu's *Fs*	Tajima's *D*
Boran	17	0.941^{ns}	-1.204^{ns}
Begait	16	0.461^{*}	2.364^{*}
Fogera	17	-0.631^{ns}	-0.956^{ns}
Horro	16	-1.037^{ns}	-1.214^{ns}
Boran * HF cross	23	2.463^{ns}	0.258^{ns}

Number of sequences (*N*), significant (*) at p<0.10, non-significant (ns)

4.1.4. *DGAT1* gene differentiation in the studied cattle populations

For the evaluation of population differentiation among the five cattle populations two main parameters (F_{ST} fixation index and exact *G* test) were estimated. The G_{ST} (coefficient of interpopulation genetic differentiation) test which is used to measure among population differentiation, relative to the total diversity value, was compared between cattle populations

and highest value was observed between Boran and Begait, whereas the lowest G_{ST} value was found between Boran vs. Fogera and Horro, and also Fogera vs. Horro (Table 6). The pairwise F_{ST} showed significant difference across all cattle populations (F_{ST} = 0.13; *p*-value ≤ 0.0001). The estimated F_{ST} values between *Bos taurus* breeds and our sampled cattle breeds ranged from 0.005 (*Bos taurus* vs. Boran * HF crosses) to 0.389 (*Bos taurus* vs. Begait).

Table 6. Genetic distances between pairs of populations based on F_{ST} below the diagonal and G_{ST} above the diagonal estimated

Breeds	BHC	BR	BG	FG	HR
BHC	-	0.03659	0.07786	0.06447	0.05644
BR	0.06655	-	0.09618	-0.01613	-0.02029
BG	0.34631	0.21648	-	0.07531	0.08628
FG	0.15855	-0.02984	0.13728	-	-0.02013
HR	0.12778	-0.03543	0.25748	-0.01595	-

Boran*HF cross (BHC), Boran (BR), Begait (BG), Fogera (FG), Horro (HR)

Among the sampled cattle populations, the F_{ST} value was the highest between Begait vs. Boran * HF cross, and a low F_{ST} value (no genetic subdivision) was noted between Boran vs. Fogera and Horro and also Fogera vs. Horro. Moreover, F_{ST} values were computed across species leading to the detection of higher genetic differences and *Bubalus bubalis* and *Camelus dromaderies* showed least differentiation and maximum differentiation was observed between cattle and other species (Fig 6).

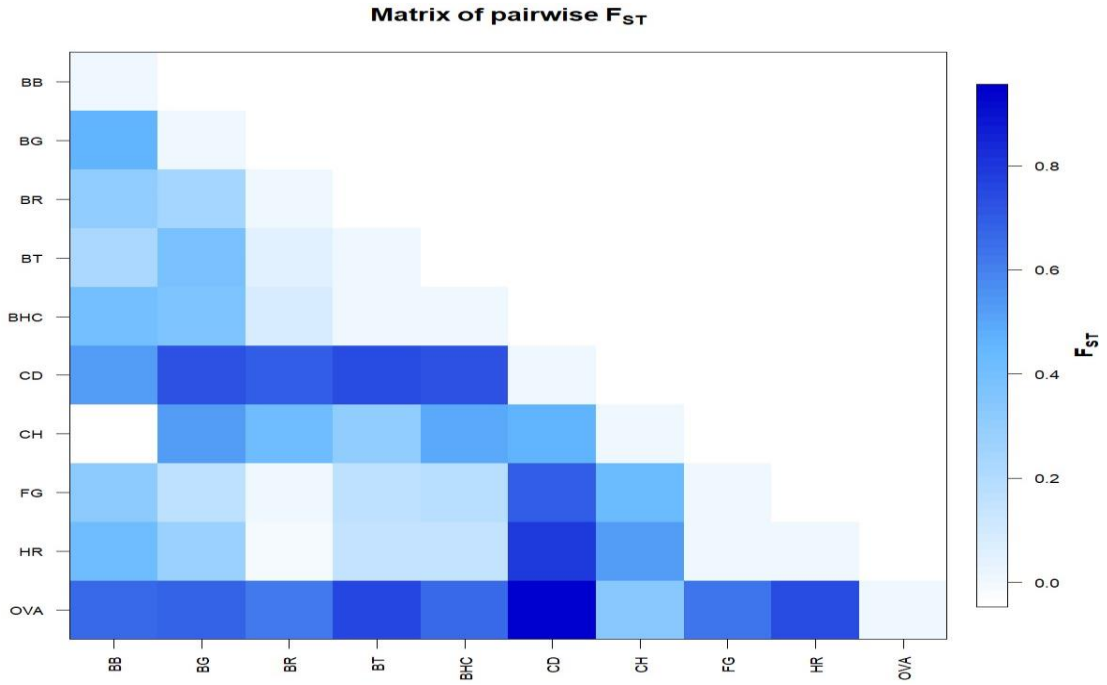


Figure 6. *DGAT1* gene graphic representation of calculated F_{ST} values between pairs of population. (*Bos taurus* (BT), *Camelus dromaderies* (CD), *Capra hircus* (CH), *Ovis aries* (OVA), and *Bubalus bubalis* (BB), generated by the R function: pairwise F_{ST} matrix.R

Analyses of molecular variance (AMOVA) were performed to examine the partitioning of genetic variation. Among groups variance was significant at 13.17% (F_{ST}; $p < 0.0001$), while 86.83% of the total genetic variation was within populations. Within-populations inbreeding (F_{IS}) was not found to be significant ($p > 0.05$) (Table 7).

Table 7. Analysis of molecular variance of *DGAT1* gene

Source	DF	Sum of squares	Variance components	Percentage of variation	F_{st}
Based on breeds					
Among population	4	16.682	0.1716Va	13.17	0.13***
Within population	84	95.094	1.13207Vb	86.83	
Total	88	111.775	1.3037		

F_{ST}; genetic differentiation *** $p < 0.0001$

4.1.5. Evolutionary Analysis

4.1.5.1. Evolution of *DGAT1* gene in the studied cattle populations

The ratio between non-synonymous (Ka) and synonymous mutations (Ks) was calculated to shed light on the evolution of *DGAT1* gene (Table 8). The ratio of non-synonymous substitution (Ka) to synonymous substitution (Ks) was the highest in Fogera cattle and was found to be the lowest in Boran, Begait and Horro cattle populations.

Table 8. Evolution (Ka and Ks) of *DGAT1* gene for considered breeds

Breeds	Ka	Ks	Ka/Ks
BR	0	0.014	0
BG	0	0.013	0
FG	0.0047	0.0062	0.758
HR	0	0.005	0
BHC	0.0037	0.01	0.37

The evolutionary divergence between different breeds/species was calculated in MEGA 11 algorithm program (Supplementary Table 2). The gene sequences of *DGAT1* region of different species or genus were retrieved from NCBI database and were aligned with the generated sequences of studied populations to reveal the divergence of *DGAT1* gene. The mean evolutionary distance estimated between pair of breed/species was observed as 0.0200±0.0100. The evolutionary distance between any two lysine variant in indigenous cattle ranged between 0.000 and 0.0348. Similarly, the evolutionary distance between indigenous and *Bos taurus* ranged from 0.0049 to 0.0348. The distance between different lysine variant of studied breeds and buffalo was slightly higher than mean value with range of 0.0049-0.0400.

4.1.5.2. Construction of Median-joining network tree

The ninety *DGAT1* gene sequences of major cattle populations of Ethiopia, were defined by eleven polymorphic sites that further collapsed into eleven haplotypes (Fig 7), among which five minor haplotypes, for instance, H_3, H_7, H_8, H_10 and H_11 were unique to BHC,

BR, BG, and FG populations, respectively. The number of individual sequences in H-1, H-2, H-3, H-4, H-5, H-6, H-7, H-8, H-9, H-10 and H-11 was 11, 53,1,12,2,4,1,1,2,1, and 1, respectively. H-2 was the most probable ancestral haplotype. Haplotypes H-5, H-7 and H-10 separated from the rest of the clusters. Fewer polymorphic sites that led to the generation of limited number of haplotypes suggested that Ethiopian cattle populations were characterized by narrow genetic diversity. Haplotype based analysis of the *DGAT1* gene indicated that from the total 11 haplotypes, only one haplotype is common for all cattle populations studied, while two haplotypes are shared by four populations and one haplotype is shared by three populations. For shared haplotype, the highest haplotype frequencies were obtained in Boran and Fogera (Supplementary Table 1). However, non-shared haplotypes have less than 10% haplotype frequency estimates.

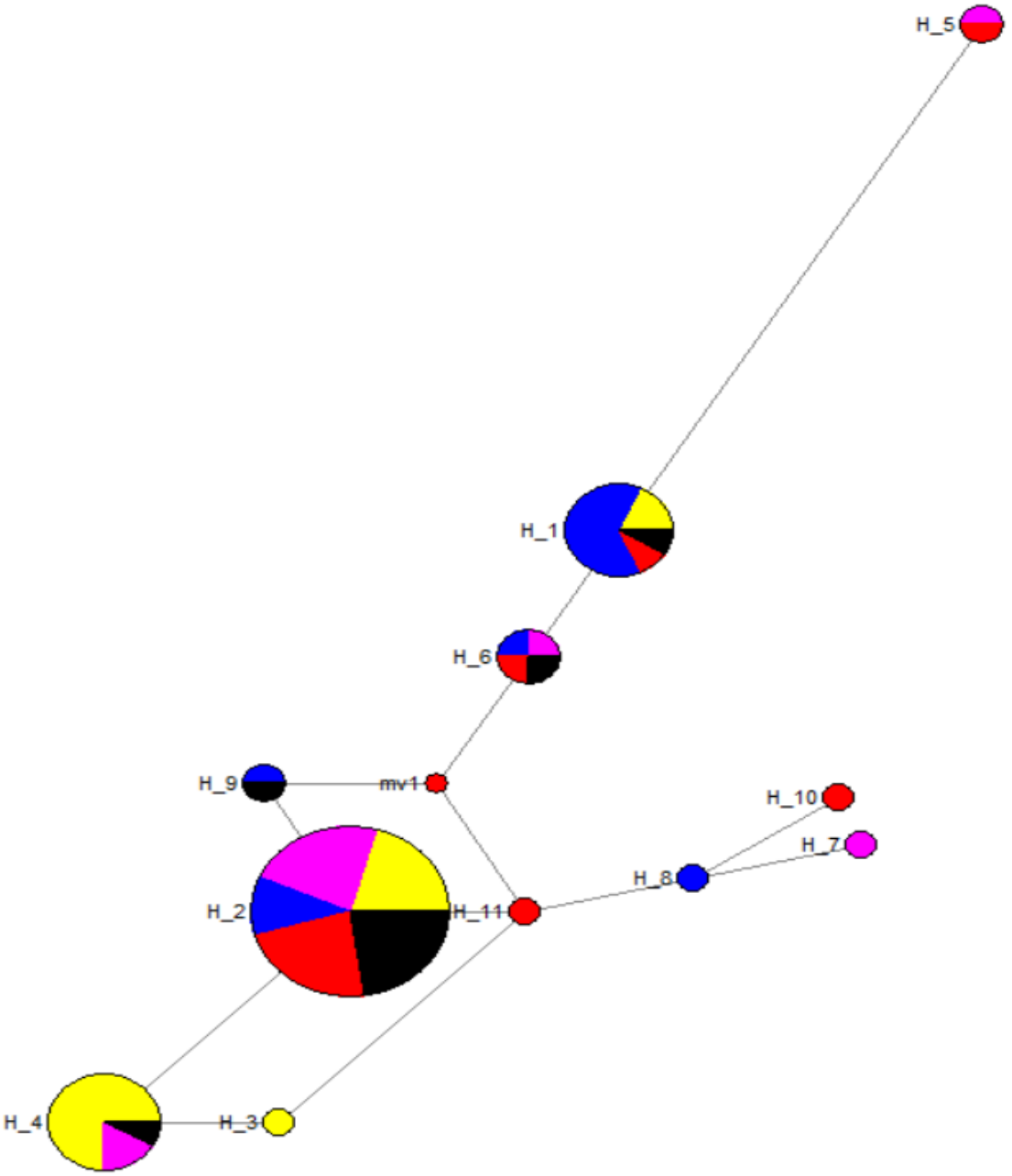


Figure 7. *DGAT1* Median-joining network of studied populations (BHC = Yellow; BG = Blue; BR = Pink; FG = Red; and HR = Black). Circles are proportional to the frequency of individual haplotypes. The branch length is proportional to the mutation rate of the haplotypes whereas the diamond with red color represents the median vector for extinct and unsampled extant ancestral sequences

4.1.5.3. Conserved regions and Homology analysis among breeds/species for *DGAT1* gene

The identification of conserved regions between the studied breeds and exotic cattle and other species was detected using Bio Edit software. Three conserved regions were identified among the studied cattle breeds whereas three different conserved regions were identified between studied breeds and exotic cattle and other species. In a nutshell, nucleotide positions 10381 - 10432, 10437 - 10521 and 10544 -10563 were identified as highly conserved among diverse range of species in reference to AJ318490 accession number (Fig 5).

Basic Local Alignment Search Tools (BLAST) found the regions of local similarity between nucleotide sequences. The result of analysis of *DGAT1* gene fragment of Boran, Begait, Fogera, Horro and Boran * HF cross cattle had 97.58% identity with cattle (*Bos taurus*) AJ318490, 96.62% with goat (*Capra hircus*) LT221856 and sheep (*ovis aries*) FJ415875 and 96.14% with camel (*Camelus dromedaries*) MF069171 (Supplementary Table 4).

4.2. Effect of breed and *DGAT1 K232A* on milk traits in the studied cattle populations

4.2.1. Effect of breed on milk yield and composition

Least square means of milk components across populations are reported in Table 9. Boran * HF crosses produced higher daily milk yield (DMY) than the others ($p<0.05$). The two lowland breeds Boran and Begait produced milk with higher fat and protein content, respectively, compared to the others ($p < 0.05$). On the other hand, Horro cattle had higher lactose content ($p < 0.05$). The Boran, Begait and Horro breeds produced milk with a higher content of SNF (Table 9).

Table 9. Effect of breed on DMY and milk composition traits in cattle populations of Ethiopia

Variables	Breeds				
	BR (17)	BG (16)	HR (18)	FG (17)	BHC (24)
DMY(L/day)	2.12 ± 0.12^{ab}	2.82 ± 0.12^{bc}	2.42 ± 0.26^{b}	2.68 ± 0.12^{bc}	7.21 ± 0.10^{d}
Fat (%)	6.21 ± 0.04^{d}	4.68 ± 0.042^{c}	3.47 ± 0.09^{b}	4.63 ± 0.041^{c}	3.18 ± 0.03^{a}
Protein (%)	3.54 ± 0.03^{bc}	3.63 ± 0.03^{c}	3.39 ± 0.08^{b}	3.04 ± 0.03^{a}	3.14 ± 0.03^{a}
Lactose (%)	4.72 ± 0.05^{b}	4.43 ± 0.05^{a}	5.31 ± 0.11^{c}	4.29 ± 0.05^{a}	4.70 ± 0.04^{b}
SNF (%)	8.99 ± 0.09^{c}	8.86 ± 0.09^{c}	9.35 ± 0.18^{c}	7.90 ± 0.08^{a}	8.57 ± 0.08^{b}
Ash (%)	0.71 ± 0.008^{a}	0.76 ± 0.008^{b}	0.74 ± 0.01^{b}	0.69 ± 0.008^{a}	0.70 ± 0.007^{a}

DMY: Daily Milk Yield; SNF: Solid Not Fat; ^{a,b,c,d} Values within a row with different superscripts differ significantly at $p<0.05$

4.2.2. Genotype and Allele frequencies of *K232A* protein variant

The genotypic and allelic frequencies of the *DGAT1* gene in Boran, Begait, Horro, Fogera and Boran * HF cross cattle are presented in Table 10 and Fig 10. The *DGAT1* 'K' allele, corresponding to lysine substitution was the predominant allele in all zebu populations (Fig 8).

Table 10. Genotypic and allele frequencies of exon-8 of *DGAT1* gene

		Genotypic frequencies			Allele frequencies		
Breeds	N	KK	KA	AA	K	A	SE
Boran	17	0.59(10)	0.41(7)	-	0.80	0.20	0.07
Begait	16	0.50(8)	0.50(8)	-	0.75	0.25	0.08
Horro	17	0.94(16)	0.06(1)	-	0.97	0.03	0.03
Fogera	18	0.56(10)	0.44(8)	-	0.78	0.22	0.07
Boran * HF cross	24	0.25(6)	0.50(12)	0.25(6)	0.50	0.50	0.07
Overall	92	0.54(50)	0.39(36)	0.07(6)	0.74	0.26	0.03

N; Number of animals, SE-Standard error of allele frequencies

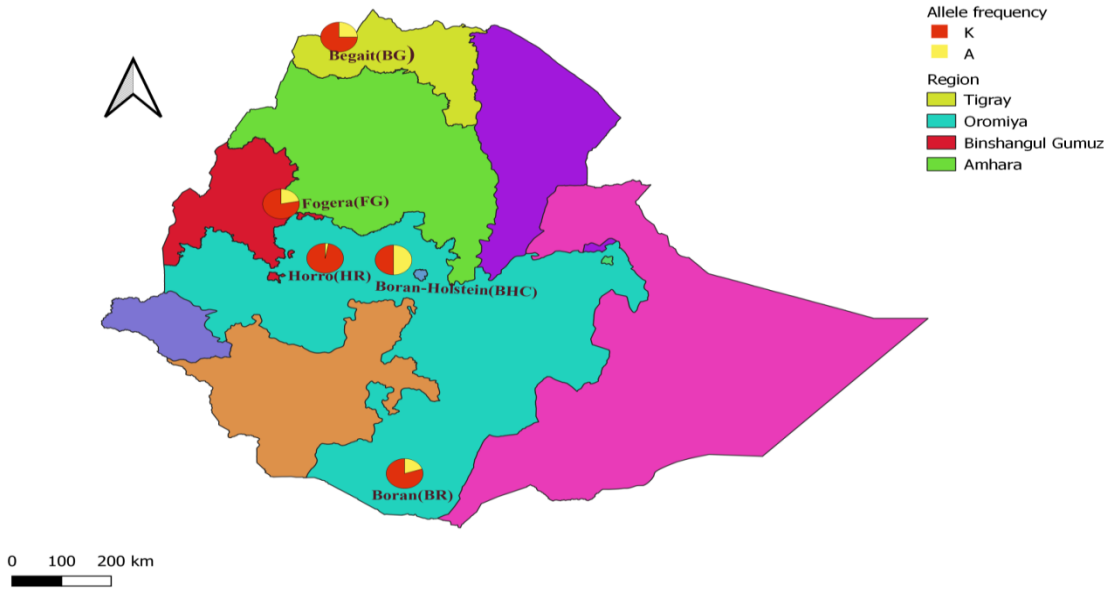


Figure 8. Map of Ethiopia showing the allele frequencies of *DGAT1* gene based on National Regional States. The map was constructed using the QGIS-OSGeo4W v.3.8.0 software

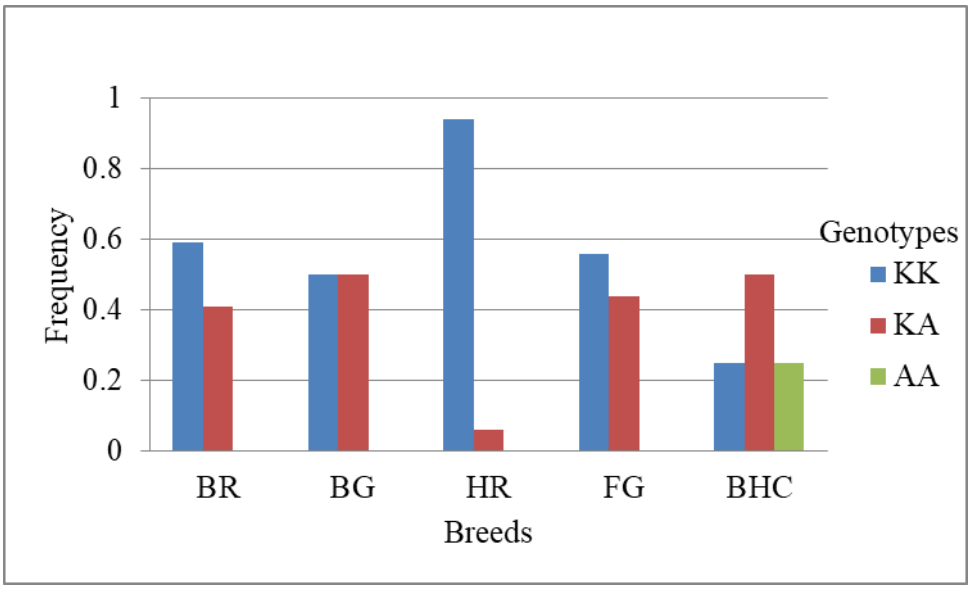
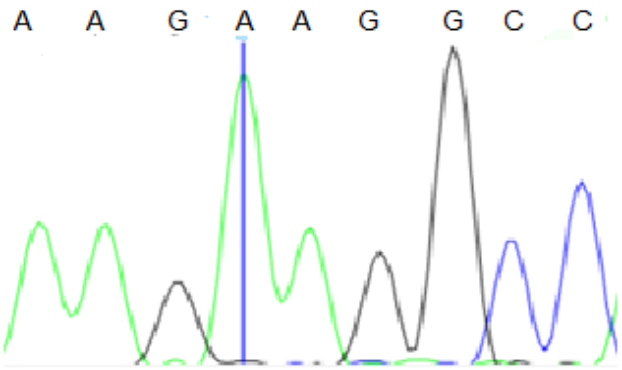
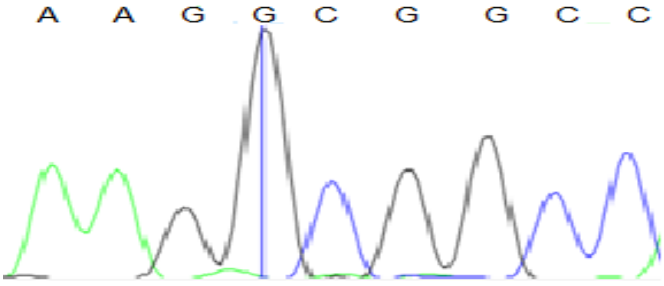


Figure 9. Genotypic frequencies of *DGAT1* K232A variants by breed

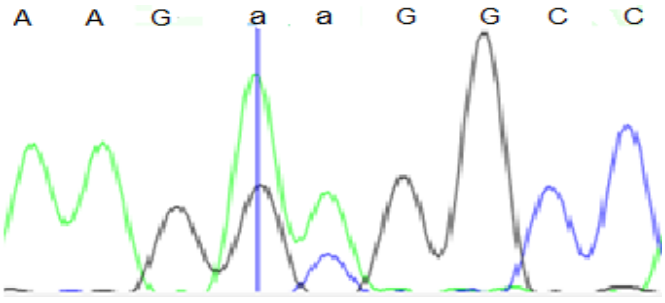
The three detected genotypes were shown on Fig 10. All three possible genotypes of *DGAT1 K232A* were observed in Boran * HF crosses (Fig 9), although the KA genotype predominated. The KK genotype predominated in the Boran and Fogera populations, but was even more prevalent in Horro. In Begait, the KK and KA genotypes have similar frequency distributions (Fig 9).



AA/AA (KK) Genotype



GC/GC (AA) Genotype



GC/AA (KA) Genotype

Figure 10. Sequence trace views of sequencing traces for positions 10433 and 10434. The vertical line indicates nucleotide position 10433. Positions 10433 and 10434 are responsible for the K232A substitution. The positions of individual mutations were named according to the sequence available in the GenBank (AJ318490)

4.2.3. Effect of *DGAT1 K232A* variants on daily milk yield and composition traits

Variability at the *DGAT1* K232A polymorphic site indicated that this locus can be considered for association studies only in Boran, Begait, Fogera and Boran * HF crosses. In Horro populations the minor allele frequency is less than (0.1) hence does not qualify for association studies.

The AA genotype was associated with more milk yield in Boran * HF cross, whereas KA genotype associated with more milk yield in the zebu populations (p<0.05). The fat and lactose content values of KA genotypes were lesser than the KK genotypes in all the genetic groups (p< 0.05). Milk protein and SNF percentage of KA genotypes were lesser than the KK genotypes in all the genetic groups (p > 0.05) (Table 11).

Table 11. *DGAT1* K232A genotypes effect on DMY and composition traits (Mean±SE)

		Genotypes for DMY(L/day)			Genotypes for fat (%)		
Breeds	N	AA	KA	KK	AA	KA	KK
BR	17	-	2.40 ± 0.19^{b}	1.84 ± 0.16^{a}	-	6.08 ± 0.06^{a}	6.35 ± .05^{b}
BG	16	-	2.91 ± 0.18^{b}	2.74 ± 0.18^{a}	-	4.46 ± 0.06^{a}	4.9 ± 0.06^{b}
FG	18	-	2.89 ± 0.18^{b}	2.47 ± 0.16^{a}	-	4.44 ± 0.06^{a}	4.83 ± 0.05^{b}
BHC	24	10 ± 0.21^{c}	7.62 ± 0.15^{b}	4 ± 0.21^{a}	2.41±0.07^{a}	3.12 ± 0.05^{b}	4.03 ± 0.07^{c}
		Genotypes for lactose (%)			Genotypes for protein (%)		
		AA	KA	KK	AA	KA	KK
BR	17	-	5.00 ± 0.08^{b}	4.45 ± 0.07^{a}	-	3.38 ± 0.06	3.69 ± 0.05
BG	16	-	4.49 ± 0.08^{b}	4.36 ± 0.08^{a}	-	3.60 ± 0.06	3.66 ± 0.06
FG	18	-	4.34 ± 0.08^{b}	4.23 ± 0.07^{a}	-	2.99 ± 0.06	3.11 ± 0.05
BHC	24	5.11 ± 0.09^{c}	4.59 ± 0.06^{b}	4.40 ± 0.09^{a}	3.12 ± .06	3.03 ± 0.05	3.28 ± 0.06

N, number of cows, DMY, daily milk yield, Predicted means and the SE of those means were derived from the GLM. ^{a,b,c} Values within a row with different superscripts differ significantly at p<0.05 for the traits DMY, fat(%), and Lactose content (%). In the case of protein content (%), p>0.05.

Substitution effect of K allele showed that the one copy of this allele leads to significant (p<0.05) increase of fat content upto 0.81% in Boran * HF crosses and protein content upto 0.32% in Boran breeds. Whereas, substitution effect of 'K' allele leads to decrease of daily milk yield up to 3L in Boran * HF crosses (Fig 11).

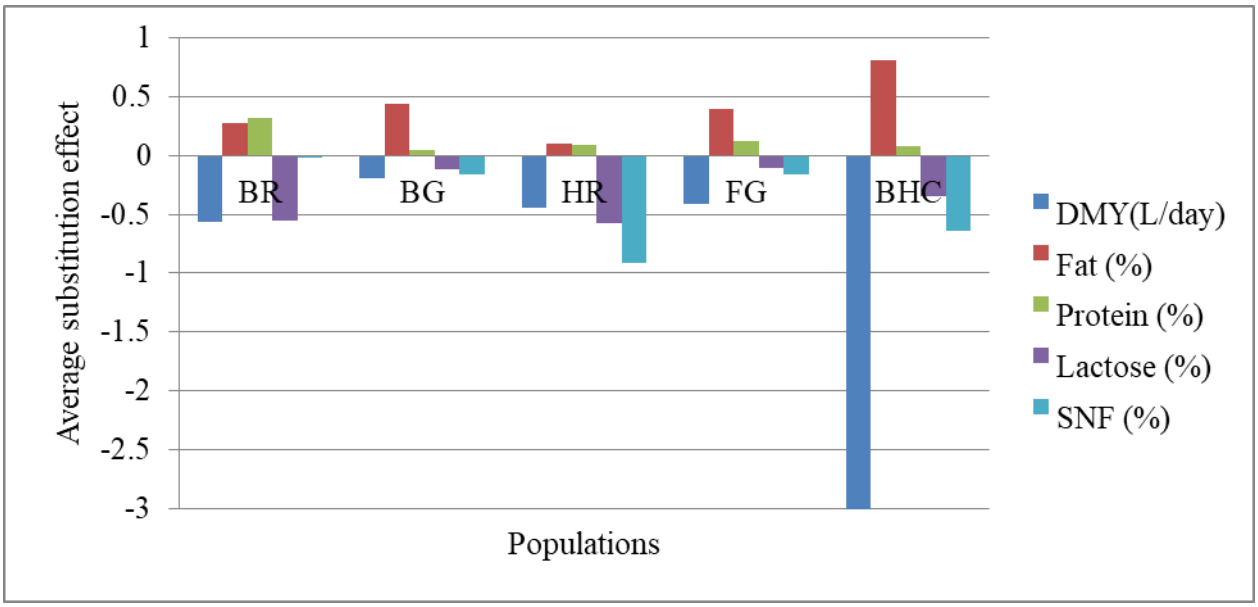


Figure 11. Estimates of the K allele average substitution effect

The phenotypic correlations between milk components in Boran * HF cross cows are presented in Table 12. The correlations of fat percentage were negative between all the traits (p<0.01). DMY showed a significant (p<0.01) positive correlation with lactose content and SNF and a negative correlation with fat content (p<0.01).

The phenotypic correlations between milk components in indigenous cows are presented in Table 12. The fat percentage showed significantly (p<0.05) positive correlation with protein content and negative correlation with DMY (p<0.05), lactose and ash content (p<0.01) (Table 12).

Table 12. Phenotypic coefficient of correlations (Pearson) for Boran * HF cross cattle (above the diagonal) and Indigenous cattle (below the diagonal)

Variable	DMY	Fat (%)	Protein (%)	Lactose (%)	SNF (%)	Ash (%)
DMY		-0.96**	-0.29	0.76**	0.76**	0.27
Fat (%)	-0.30*		0.38	-0.69**	-0.70**	-0.33
Protein (%)	-0.39**	0.29*		-0.16	-0.16	0.23
Lactose (%)	-0.08	-0.34**	0.04		1.00**	0.05
SNF (%)	-0.14	0.06	0.68**	0.68**		0.04
Ash (%)	0.03	-0.36**	0.56**	0.33**	0.56**	

DMY: Daily milk yield, SNF: Solid not fat *p<0.05; **p< 0.01

4.3. The effect of *PRL* gene on milk yield and composition in the studied cattle populations

4.3.1. *PRL* SNPs among the studied populations

The position of individual *PRL* SNP was identified through comparison with reference sequence (AF426315) available in NCBI GenBank database. The *PRL* sequence fragment analyzed was confirmed to cover a part of intron-3 (30 bp), the whole exon-4 (180 bp) and part of intron-4 (17 bp). A total of five SNP's were detected on alignment of representative sequences with reference sequence, of which two SNP's were observed in non-coding region. The exon-4 region was found to be highly variable than other regions of the studied fragment as sixty percent of detected SNP's were located in this region. Three SNP's were observed in coding region of exon-4 of *PRL* gene at position 8323th, 8377th and 8398th. Mutation g.8362T>C is not a SNP in the sampled populations, as the C allele frequency was 1, so this site is only different from reference sequence. Most of the identified mutation in amplified fragments were observed to be transition in nature except mutation at position 8323th, and 8377th which were transversion in nature. The SNP's identified at referred position 8323th, 8377th and 8398th were located in exon-4 region (Fig 12), whereas SNP's at referred position g.8307T>C noted in Begait, Horro and Fogera cattle only and g.8314C>T seen in all the populations were located in intron-3 of *PRL* gene. All the detected SNP's were

found in Begait and Fogera populations. In exon-4 region in Boran, Boran * HF cross and Horro cattle g .8377G>C and g.8398A>G SNP's were detected.

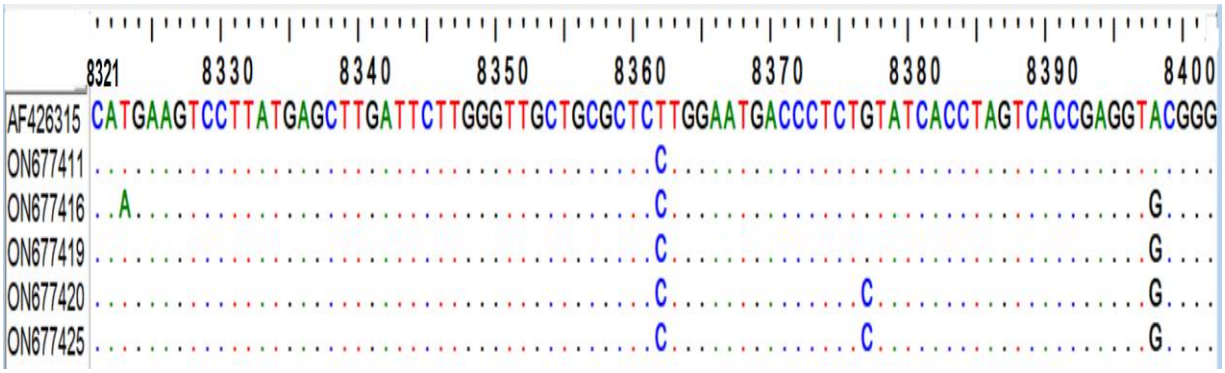


Figure 12. The four mutations identified at the coding region of *PRL* gene, positions shown relative to reference sequence ID from the Genbank (AF426315), dot represents similar nucleotides with the reference sequence ID from Genbank. Accession numbers: Boran* HF Cross (ON677411), Boran (ON677416), Begait (ON677419), Fogera (ON677420) and Horro (ON677425)

4.3.2. Distributions of *PRL* SNPs allele and genotype frequencies

Allele distributions pertaining to the different SNPs were shown in Table 13. As shown in Table 13, the frequencies of allele A and T in the pooled population was 0.046 and 0.954, respectively for g.8323T>A SNP and allele A was a rare allele only detected in FG breed. SNP g.8377G>C, allele C was more frequent in the BHC and BR compared to the rest of the breeds, while the G allele was more predominant in BG, FG and HR. SNP g.8398A>G, allele A was more frequent in BG and less frequent in BR breeds. G allele was more frequent in BR, BHC and HR cattle breeds.

Table 13. Distributions of *PRL* SNPs allele frequencies in cattle population of Ethiopia

SNPs	N	Allele(frequencies)	Alleles distribution in each breed				
			BG(17)	BR(18)	BHC(17)	FG(18)	HR(17)
g.8323T>A	83	T(0.954)	0.971	1	1	0.806	1
	4	A(0.046)	0.029	-	-	0.194	-
g.8377G>C	51	G(0.587)	0.706	0.417	0.352	0.778	0.677
	36	C(0.413)	0.294	0.583	0.648	0.222	0.323
g.8398A>G	39	A(0.449)	0.706	0.333	0.352	0.500	0.352
	48	G(0.551)	0.294	0.667	0.648	0.500	0.648

N (number of cattle heads), Begait (BG), Boran (BR), Boran * HF cross (BHC), Fogera (FG), Horro (HR)

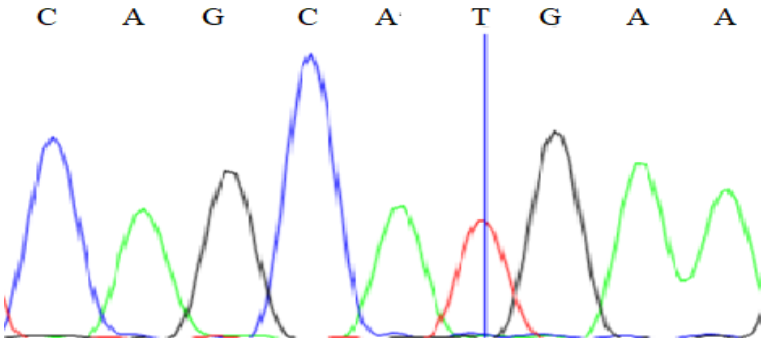
Distributions of SNPs genotypic frequencies in cattle population of Ethiopia are presented in Table 14. The genotype frequencies of SNPs g.8377G>C and g.8398A>G varied across the breeds (Table 14, Fig 13). SNP g.8323T>A genotype AT was a rare variant detected in BG and FG breed. For SNP g.8377G>C, the variant CC was more frequent in BHC and GC was more frequent in BR and lowest in FG. GG variant was frequent in FG. As far as SNP g.8398A>G is concerned, variant GA and GG showed similar pattern, while variant AA was not detected in BR breeds. In contrast, the frequency of GA was lowest in FG and frequent in BR.

Table 14. Genetic variability of exon-4 regions of *PRL* gene in the sampled populations

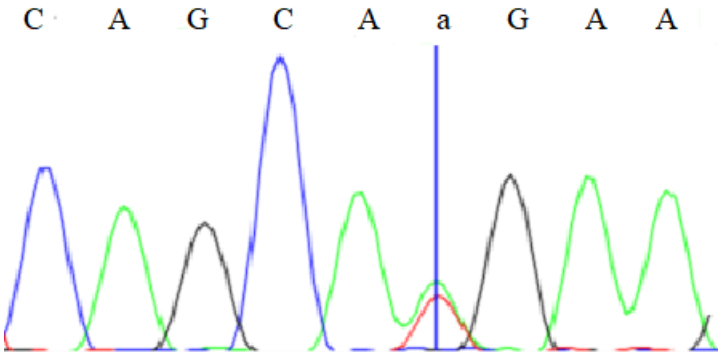
SNPs	N	Genotype Frequencies	Genotype distribution in each breed						
			BG	BR	BHC	FG	HR	*MAF	**HWE
	8	AT (0.092)	0.059	-	-	0.389	-	0.046	0.653
g.8323T>A	79	TT (0.908)	0.941	1.00	1.00	0.611	1		
	18	CC (0.207)	0.06	0.222	0.471	0.111	0.177		
g.8377G>C	36	GC (0.414)	0.47	0.722	0.352	0.222	0.294	0.413	0.170
	33	GG (0.379)	0.47	0.056	0.177	0.667	0.529		
	22	AA (0.254)	0.471	-	0.178	0.389	0.235		
g.8398A>G	34	GA (0.390)	0.470	0.667	0.352	0.222	0.235	0.448	0.147
	31	GG (0.356)	0.059	0.333	0.470	0.389	0.530		

N(number of cattle heads), Minor allele Frequency (*MAF), Hardy-Weinberg Equilibrium (**HWE)

g.8323T>A

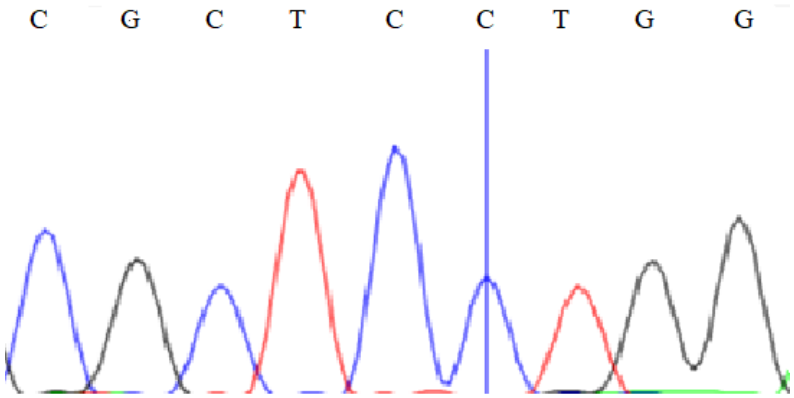


TT Genotype



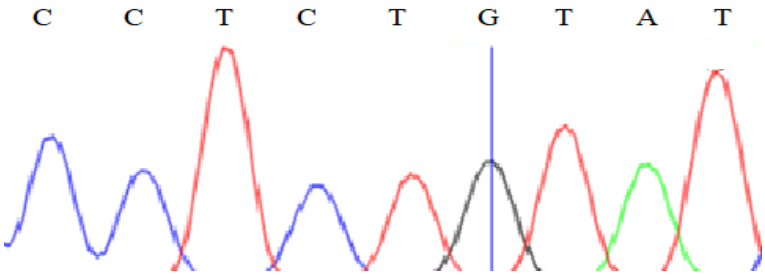
TA Genotype

g.8362T>C

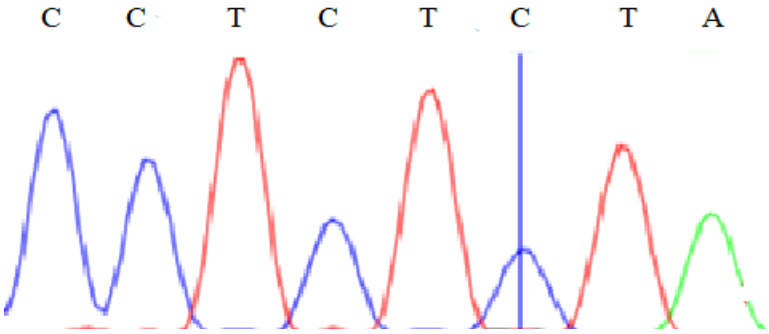


CC Genotype

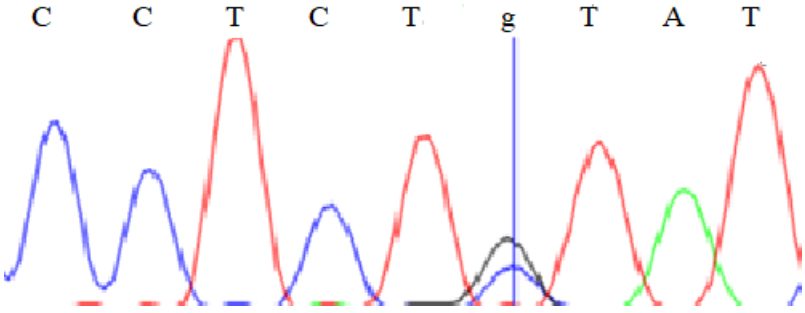
g.8377G>C



GG Genotype



CC Genotype



GC Genotype

g.8398A>G

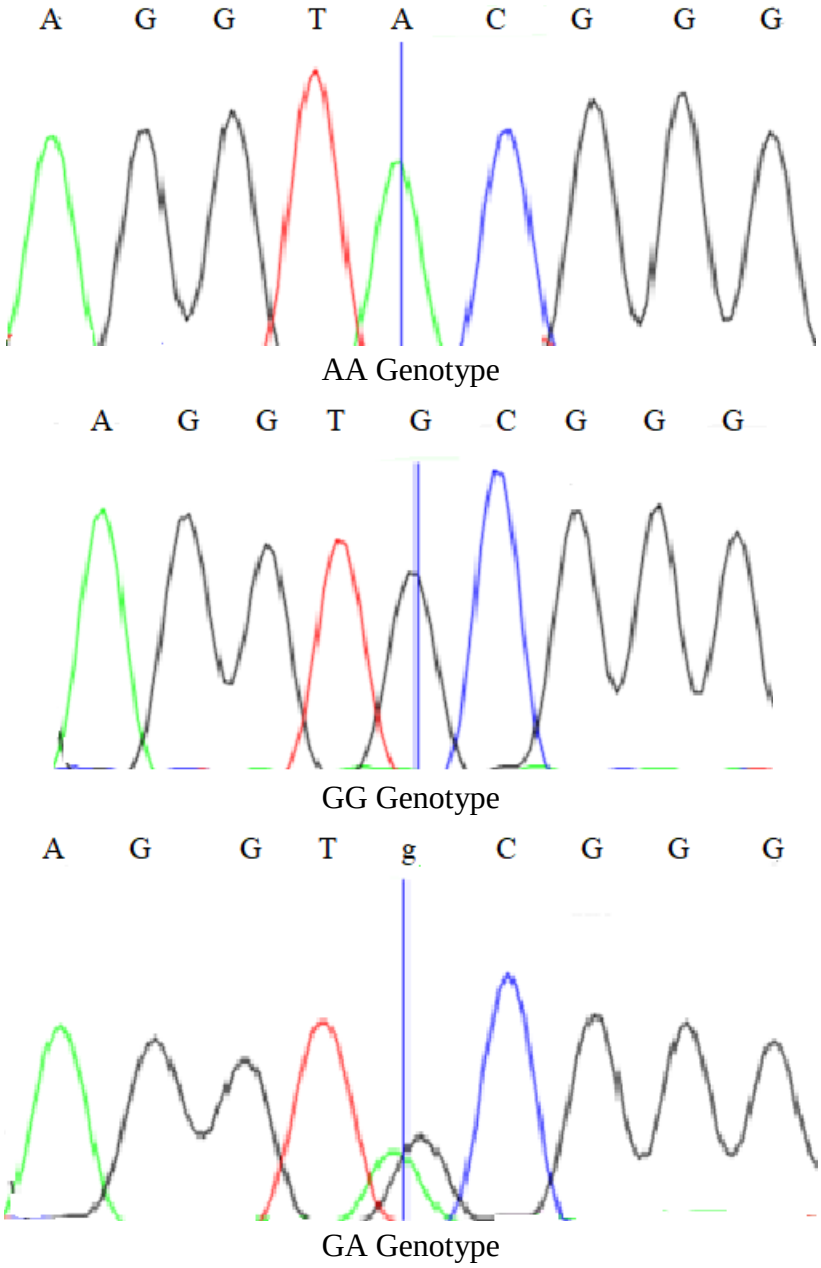


Figure 13. Sequence trace views of sequencing traces for positions 8323, 8362, 8377 and 8398. The vertical line indicates nucleotide position g.8323T>A, g.8362T>C, g.8377G>C and g.8398A>G. The positions of individual mutations were named based on the sequence available in the GenBank (Acc.No. AF426315)

The MAF value for SNP g.8398A>G was the highest as compared to the rest SNPs. The genotypes of the g.8323T>A, g.8377G>C and g.8398A>G SNPs fit HWE ($p > 0.05$) (Table 14). Four changes g.8323T>A, g.8362T>C, g.8377G>C and g.8398A>G in exon-4 were identified compared to the *Bos taurus* (BT) reference sequence (AF426315). The missense mutation detected at g.8323T>A led to amino acid alterations from Histidine to Glutamine (Fig 14). The SNP g.8323T>A was only observed in BG and FG cattle populations. The genotypes of observed SNPs revealed no significant deviations from HWE ($p > 0.05$) (Table 14).

SNPs g.8377G>C and g.8398A>G corresponded to the production of Leucine 125L>L and Valine 132V>V, respectively and generated silent mutation (Fig 14).

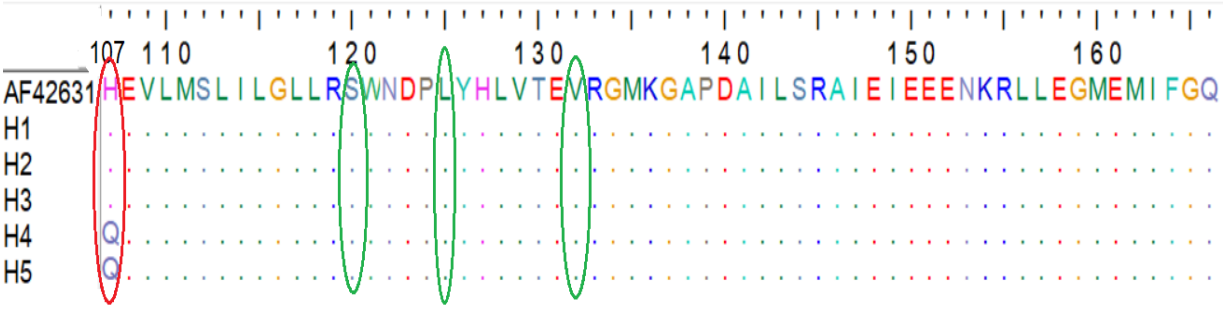


Figure 14. Mutations detected in exon-4 of *PRL* gene with corresponding amino acid changes circled by a red color, silent mutations identified circled with green and the dot represents similar amino acids with the reference sequence ID from Genbank (AF426315; protein_id=AAL28075.1) and H = haplotypes

4.3.3. Genetic diversity among the studied breeds

Haplotype based analysis of the *PRL* gene was carried out in this study. The results indicated that from the total five haplotypes, only three of them (H-1, H-2 and H-3) are common for all populations studied. Relatively highest haplotype frequencies were observed in Boran and Boran * HF cross cattle populations in the shared haplotypes (Table 15).

Table 15. Haplotype frequency distribution of *PRL* gene in the studied populations

Haplotypes	N	BG	BR	BHC	FG	HR
H-1	19	0.059	0.222	0.471	0.111	0.176
H-2	27	0.471	0.111	0.294	0.389	0.294
H-3	30	0.471	0.611	0.235	0.111	0.529
H-4	7	-	0.056	-	0.333	-
H-5	4	-	-	-	0.056	-

N (number of cattle heads)

Haplotype one was more frequent in the BHC and lowest in BG, while haplotype two was relatively high in BG and low in BR. Haplotype three was the most common in the BR and HR and lowest in FG and haplotype four was a rare haplotype only detected in FG. The overall gene (haplotype) diversity was 0.735 ± 0.020 and whereas the mean nucleotide diversity was estimated to be 0.006 ± 0.001 for all populations.

F_{ST} is the mean reduction in heterozygosity of a sub-population relative to the total population due to genetic drift among sub-populations. The estimated F_{ST} values ranged from 0.470 (*BT* vs. FG) to 0.631 (*BT* vs. BR) with respect to reference BT sequence (Fig 15). Fig 15 depicts the graphical representation of the calculated F_{ST} values between pairs of populations. F_{ST} values of greater than 0.800 distances were computed across species leading to the detection of higher genetic differences for all the species.

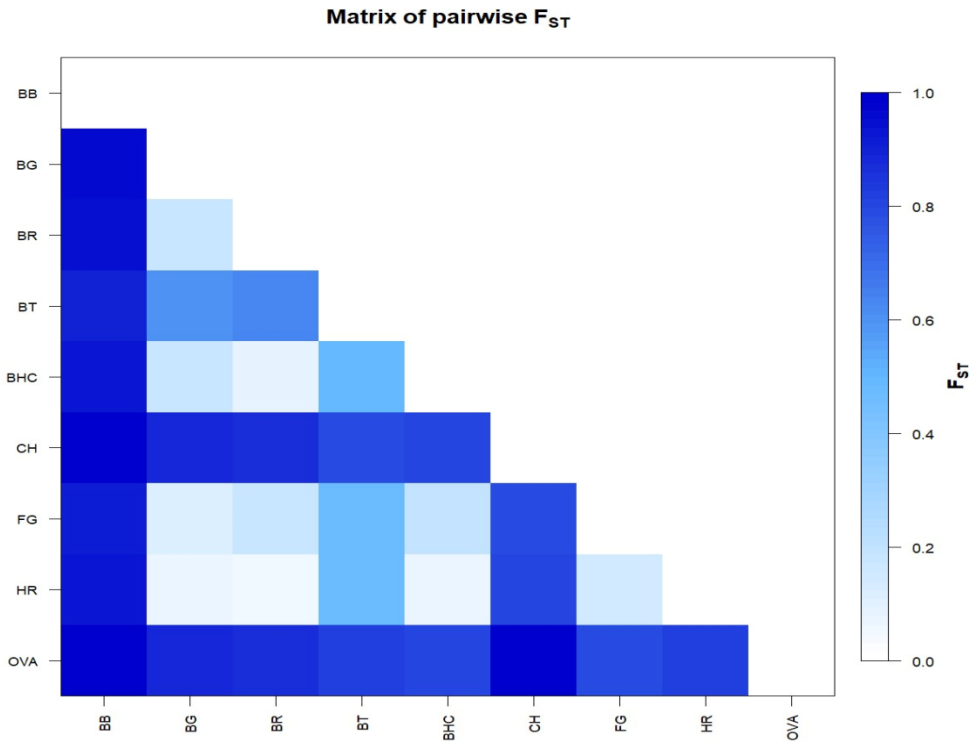


Figure 15. Representation of calculated F_{ST} values of *PRL* gene between pairs of population *Bos taurus* (BT), *Capra hircus* (CH), *Ovis aries* (OVA), and *Bubalus bubalis* (BB) graphs, generated by the R function: pairwise F_{ST} matrix.R

SNPs g.8377G>C and g.8398A>G have slightly greater expected heterozygosity (H_e) than the observed heterozygosity (H_o) (Table 16). The results showed slightly informative PIC for g.8323T>A, while reasonably informative PIC was noted for g.8377G>C and g.8398A>G. SNP g.8323T>A presented positive F_{IS} value. The rest of the SNPs had negative F_{IS} value for g.8398A>G and g.8377G>C (Table 16).

Table 16. Heterozygosity, PIC and F_{IS} estimates for *PRL* SNPs

SNPs	H_o	H_e	PIC	F_{IS}
g.8323T>A	0.092	0.088	0.083	0.042
g.8377G>C	0.413	0.485	0.368	-0.152
g.8398A>G	0.390	0.494	0.372	-0.216

4.3.4. Phylogenetic relationships

4.3.4.1. Evolutionary divergence among breeds/species

The mean evolutionary distance between different breeds or species was observed as 0.0300 ± 0.0100 (Supplementary Table 3). The distance between alleles of studied breeds ranged below the mean value and was observed as 0.000 to 0.0195. Buffalo diverged from different cattle breeds more with estimated distance of 0.0059 to 0.1216. Goats and sheep diverged from bovids animals with estimated distances of 0.0213 - 0.0321 and 0.0178 - 0.0398, respectively.

4.3.4.2. Construction of Phylogenetic Tree

An evolutionary analysis was conducted using MEGA 11 (Tamura *et al*., 2021) and the tree generated by power marker indicated that the two Zenga breeds (Fogera and Horro) were closely clustered as compared to Boran and Boran * HF cross cattle breeds (Fig 16).

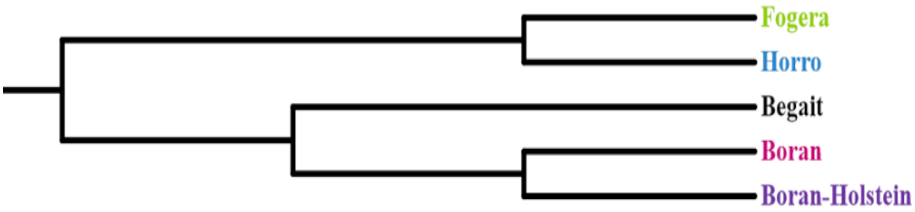


Figure 16. UPGMA tree for the five cattle populations, generated by Power Marker (version 3.25) based on PRL SNPs data with linked software, MEGA 11 with 0.02 nucleotide substitutions per site

To trace the evolutionary history of the *PRL* gene, haplotypes extracted from this study were compared to the reference sequences from *Bos taurus*, *Capra hircus*, *Ovis aries*, and *Bubalus bubalis* (Fig 17). The reference *BT* sequence clustered together with the haplotypes in our study. The sequences of the Buffalo *PRL* gene were clustered separately vis-a-vis the other mammalian species.

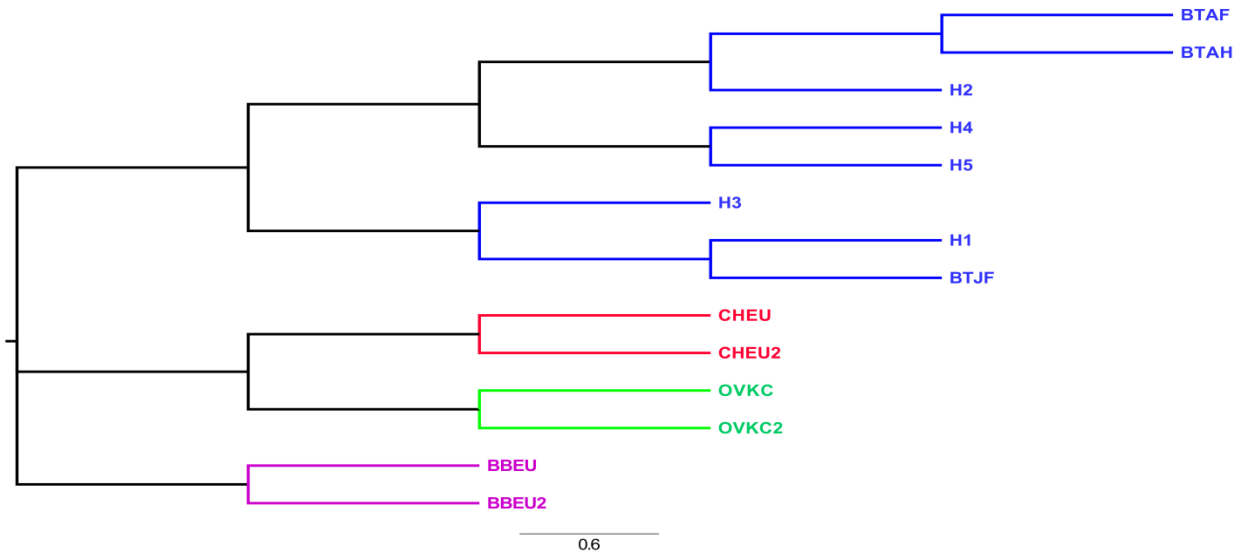


Figure 17. Molecular phylogenetic analysis by Neighbor joining method for comparison of the haplotypes of the cattle breeds with other species with 0.6 nucleotide substitutions per site of *PRL* gene. *Bos taurus* (BTAF, BTAH and BTJF), Haplotypes (H1 - H5), *Capra hircus* (CHEU and CHEU2), *Ovis aries* (OVKC and OVKC2), and *Bubalus bubalis* (BBEU and BBEU2)

Moreover, a median-joining network tree was constructed based on the haplotypes (Fig 18). The number of individual sequences in H-1, H-2, H-3, H-4 and H-5 were, 19, 27, 33, 7, and 1, respectively. The eighty seven *DGAT1* sequences of major cattle populations of Ethiopia, were further collapsed into five haplotypes (Fig 18), among which two minor haplotypes, for instance, H_4 was unique to only the FG and BR cattle populations while H_5 consisted of haplotype derived from Fogera populations. Fewer polymorphic sites that led to the generation of limited number of haplotypes suggested that Ethiopian cattle populations were characterized by narrow genetic diversity.

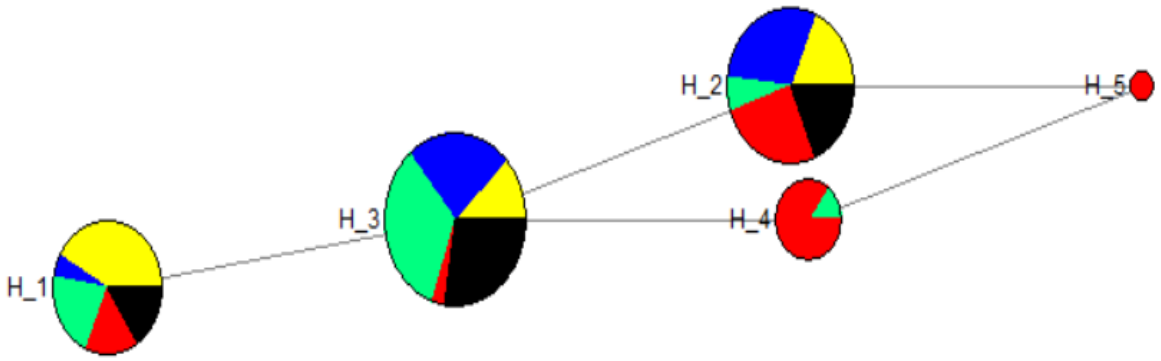


Figure 18. *PRL* Median-joining network of studied populations (BHC = Yellow; BG = Blue; BR = Green; FG = Red; and HR = Black). Circles are proportional to the frequency of individual haplotypes. The branch length is proportional to the mutation rate of the haplotypes

4.3.4.3. Conserved regions and Homology analysis among breeds/species for *PRL* gene

Three conserved regions were identified between the studied cattle breeds and the other species in the fragment region of 8354bp - 8406bp, 8408bp - 8427bp and 8429bp - 8497bp in reference to AF426315 accession number. The result of analysis of *PRL* gene fragment of Boran, Begait, Fogera, Horro and Boran * HF cross cattle revealed 98.1% identity with reference cattle (*Bos taurus*) AF426315, 96.89% with goat (*Capra hircus*) EU256167, 98.37% with sheep (*ovis aries*) LC718488, and 98.7% with Buffalo (*Bubalus bubalis*) MZ302991 (Supplementary Table 5).

4.3.5. Effects of *PRL* SNPs genotypes on milk traits

SNP g.8323T>A, MAF is less than (0.1) hence does not qualify for marker-trait association analysis. SNPs g.8377G>C and g.8398A>G, minor allele frequencies were greater than 0.1 and polymorphic thus qualifying for association analysis. However g.8377G>C did not show significant effect with any of the traits under investigation in the studied cattle populations. Therefore association data were only shown for SNP g.8398A>G. Analysis for the effect of g.8398A>G on daily milk yield and composition showed GG, GA and AA genotypes have significant effects on average daily milk yield, fat and SNF percentage in all studied breeds

(Table 17). Cows with genotype AA produced higher daily milk yield and milk with lower content of fat. Genotype AA cows had a higher (p<0.05) daily milk yield (8±0.32 L/day) than GA cows (7.67±0.18 L/day) and GG cows (4.62±0.16 L/day) in BHC cows, whereas, genotype GG was significantly associated with higher milk fat and SNF percentage. However, no significant difference (p>0.05) was observed for the protein (%), lactose (%) and ash (%) among the GG, GA and AA genotypes in all populations.

Table 17. *PRL* g.8398A>G genotypes effects on milk traits (Mean±SE)

		Genotypes for DMY(L/day)			Genotypes for fat (%)		
Breeds	*N*	AA	GA	GG	AA	GA	GG
BR	18	-	2.27 ± 0.13^{b}	1.74 ± 0.18^{a}	-	6.12 ± 0.07^{a}	$6.42 \pm .09^{b}$
BG	17	2.89 ± 0.15^{a}	2.76 ± 0.15^{a}	2.65 ± 0.45^{a}	4.54 ± 0.08^{a}	4.84 ± 0.08^{b}	5.01 ± 0.24^{c}
HR	17	2.61 ± 0.22^{b}	2.51 ± 0.22^{b}	1.93 ± 0.15^{a}	3.43 ± 0.12^{a}	3.48 ± 0.12^{a}	3.57 ± 0.08^{a}
FG	18	2.9 ± 0.17^{c}	2.63 ± 0.22^{b}	2.42 ± 0.17^{a}	4.41 ± 0.09^{a}	4.67 ± 0.12^{b}	4.89 ± 0.09^{c}
BHC	17	8 ± 0.32^{c}	7.67 ± 0.18^{b}	4.62 ± 0.16^{a}	2.94 ± 0.17^{a}	3.21 ± 0.09^{b}	3.88 ± 0.08^{c}
		Genotypes for SNF (%)			Genotypes for protein (%)		
		AA	GA	GG	AA	GA	GG
BR	17	-	8.9 ± 0.05^{a}	9.14 ± 0.08^{b}	-	3.42 ± 0.04	3.82 ± 0.06
BG	16	8.87 ± 0.07^{b}	8.95 ± 0.06^{c}	8.67 ± 0.19^{a}	3.55 ± 0.05	3.65 ± 0.05	3.71 ± 0.15
HR	17	8.58 ± 0.09^{a}	8.87 ± 0.09^{b}	9.34 ± 0.06^{c}	3.36 ± 0.07	3.41 ± 0.07	3.46 ± 0.05
FG	18	7.83 ± 0.07^{a}	7.92 ± 0.09^{b}	8.14 ± 0.07^{c}	2.98 ± 0.06	3.06 ± 0.07	3.13 ± 0.06
BHC	24	8.12 ± 0.13^{a}	8.23 ± 0.08^{b}	8.27 ± 0.07^{c}	3.02 ± 0.11	3.02 ± 0.06	3.22 ± 0.05

N: number of cows; DMY: Daily Milk Yield. Predicted means and standard error of those means were derived from the GLM. ^{a,b,c} Values within a row with different superscripts differ significantly at p< 0.05 for DMY, fat(%) and SNF (%) traits, Whereas for protein(%) p>0.05.

The *PRL* allele substitution effect (represented as regression coefficient) was determined by regression of number of G allele copies (Fig 19). Substitution effect of G allele showed that one copy of this allele leads to significant (p<0.05) increase of fat content upto 0.52% in Boran * HF crosses, significant increase of SNF and protein content upto 0.39% and 0.40% in

Horro and Boran breeds, respectively. Whereas, substitution effect of one copy of G allele leads to significant decrease of daily milk yield up to 2.08L in Boran * HF crosses.

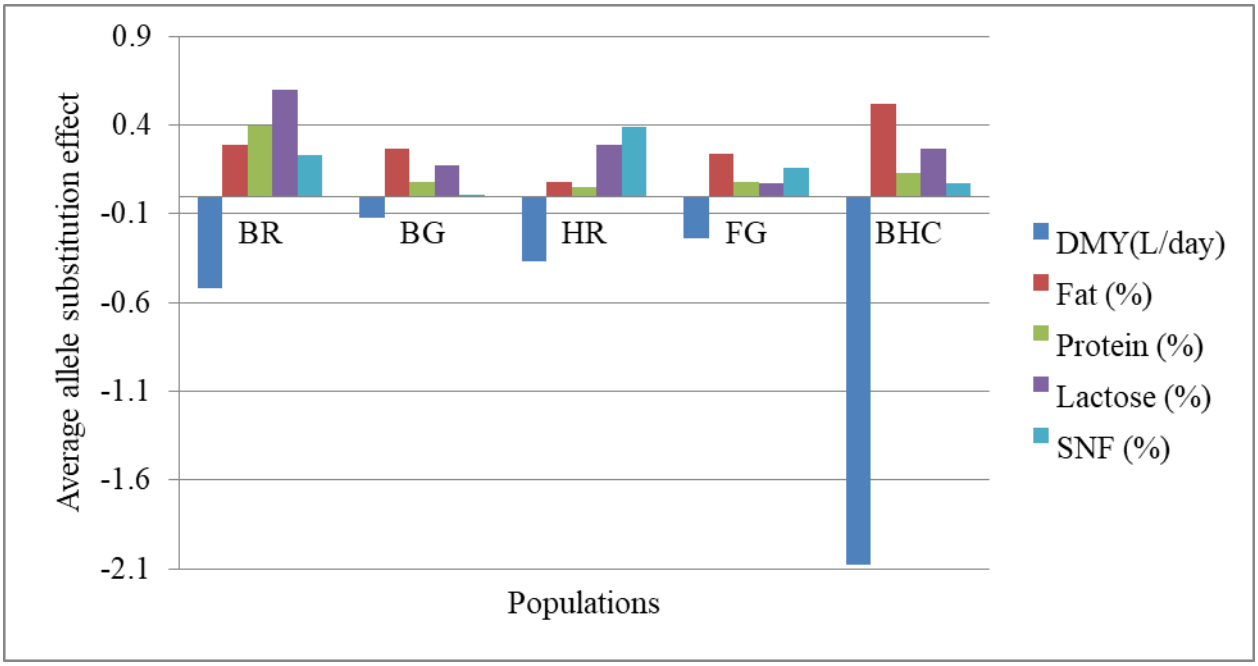


Figure 19. Estimates of the G allele average substitution effect

CHAPTER 5

5. DISCUSSION

5.1. Genetic variability of *DGAT1* gene linked to milk yield and composition traits in the studied Cattle populations

5.1.1. *DGAT1* SNP's Analysis and genetic diversity

The detection of sequence variations is the core for all genetic analysis that ultimately leads to variation discovery across candidate gene that may affect the trait of interest (Nickerson *et al*., 1997). Single nucleotide polymorphisms (SNPs) or one base changes including deletion, insertion and substitution, may play important roles in the regulation of genes transcription and amino acids sequences of mature proteins, which could be used for the association studies between candidate genes and complex traits in domestic animals (Yoon *et al*., 2005). The SNP's at position 10433-10434 were responsible for production of lysine variant of *DGAT1* gene. SNP's present at location other than AA/GC position in the same or different exon or intron region could influence the expression of *DGAT1* gene.

The reported haplotypes number, haplotype diversity and nucleotide diversity for *DGAT1* gene were 11, 0.615 and 0.0100, respectively. These values are higher than the values reported by Faraj and his colleagues (Faraj *et al*., 2020). Number of different haplotypes detected in each population reflects haplotype diversity. The combination of high haplotype diversity and low nucleotide diversity, suggested small differences between haplotypes and can be a signature of a rapid population expansion from a small effective population size (Brown, 2000). The unique breeding histories of some population may account for the high variability of genetic diversity.

Minor allele frequency (MAF) refers to the frequency of the second most common allele in a population, and it affects heritability and predictive ability (Zhu *et al*., 2017). In the present study, higher MAF recorded for indigenous cattle than Boran * HF crosses. The present MAF values were lower than the report of Edea et al. (2012) for Ethiopian cattle. This could be

attributed to the differences in genotyping platforms used and causal variants have lower MAF than SNPs in a panel (Yang *et al*., 2010).

The heterozygosity situation for *DGAT1* gene in the present study for H_e ranged from 0.120 - 0.256 were in line with the previously reported range of 0.02 to 0.50 for *Bos indicus* cattle (Krovvidi *et al*., 2021; Kaupe *et al*., 2004). Observed heterozygosity at *DGAT1* gene in this study was higher than the reports of Borgou (0.388) and White Fulani (0.155) cattle breeds of Benin (Houaga *et al*., 2017). The values of heterozygosity were lower when compared with previous study on Rathi, Sahiwal and Kankrej cattle breeds of India (Agrawal *et al*., 2018). Similarly higher values of heterozygosity in Holstein cattle breeds were reported by previous study where H_o ranged from 0.313 - 0.938 while H_e ranged from 0.264 - 0.498 (Asmarasari, 2013). The highest observed heterozygosity and haplotype diversity in Begait could indicates that *DGAT1* locus is more variable in Begait populations. These low estimates of heterozygotes imply that there could be high selection pressure, small population size, minimal or null immigration of new genetic materials into the population.

A marker with PIC>0.5 can be considered as highly informative, whereas, 0.5>PIC>0.25 recognized as reasonably informative and below 0.25 is measured as slightly informative (Botstein *et al*., 1980). In the present study breeds PIC values range between 0.101 and 0.200, indicating, the marker is slightly informative for the studied breeds. The results indicated that the sampled populations were slightly abundant in genetic diversity and had potential for further selection.

The F_{IS} coefficients are the classical Wright's F-statistic, which estimates the variation within populations. Specifically, it measures the reduction in heterozygosity in an individual caused by nonrandom mating within its subpopulation (Wright, 1951). The F_{IS} coefficient value was negative for the studied breeds ranged from -0.284 to -0.606. Within-populations inbreeding (F_{IS}) was not found to be significant ($p > 0.05$) which might be due to the large population from which the samples were drawn, and the fact that related individuals were purposely avoided. The present study is in agreement with the findings of Rincon et al.(2006) who also observed negative estimates of F_{IS} for *DGAT1* gene in Uruguayan Creole cattle population and for Borgou cattle of Benin (Houaga *et al*., 2017). Negative estimates of F_{IS} coefficient value

for Ethiopian cattle were also observed (Edea *et al*., 2012). The negative F_{IS} value within population suggests non-significant status of inbreeding in the population and reflects that samples were not from an inbred gene pool. The low and negative value of inbreeding coefficient of an individual relative to the sub-populations observed in the present study in all the studied breeds suggests the presence of genetic variation in the population to maintain the genetic diversity.

5.1.2. Neutrality test analysis

Neutrality tests of Tajima's *D* (Tajima, 1989) and Fu's *Fs* statistics (Fu, 1997) were carried out to assess signatures of recent historical demographic events. Tajima's *D* test is based on comparison of the allelic frequency of segregating nucleotide sites, while Fu's *Fs* test is based on the alleles or haplotypes distribution (Fu, 1997; Tajima, 1989). They estimate the deviation from neutrality, which is based on the expectation of a constant population size at mutation-drift equilibrium and negative values of both tests signifies an evidence for an excess number of alleles and could be expected from a recent population expansion or genetic hitchhiking and positive value signifies deficiency of alleles, as would be expected from a recent population bottleneck and/or balancing selection (Pichler, 2002; Fu, 1997; Tajima, 1989).

Overall Tajima's *D* (-0.147) and Fu's *Fs* (-1.534) tests statistics in all populations were negative and non-significant (p>0.10) indicating consistent with a populations in genetic equilibrium or in expansion (Ramos-Onsins and Rozas, 2002). However, the expected heterozygosity (*He*) estimation was not consistently higher than the observed heterozygosity (*Ho*) estimates (Annex Fig 7).

The two tests of neutrality were not significant for Boran * HF cross, Boran, Horro and Fogera cattle populations (Table 5). The results suggested that these populations are in genetic equilibrium or in expansion and Tajima's and Fu's neutrality tests were both significant for Begait population, indicating deficiency of alleles, as would be expected from a recent population bottleneck (Table 5). The Begait cattle have been reduced in population size; however, they maintained genetic diversity which is higher than to other studied breeds. The influence of factors which affect genetic diversity can complicate an attempt to interpret the

genetic diversity of any population in terms of population size as observed in Begat cattle populations (Amos, 1998).

5.1.3. Population Differentiation

Genetic divergence among populations of the same or different breeds is usually quantified by fixation indices or F statistics (Wright, 1951). The values of the F_{ST} fixation index lie between 0 (indicating the absence of genetic differences between populations) and 1 (indicates the fixation of specific alleles in the respective populations). The negative F_{ST} values recorded by Fogera vs. Horro with Boran as well as between Horro and Fogera indicated that genetic subdivision could not be established among these populations. Moreover, the genetic differentiation of all the breeds based on G_{ST} was small and supported by the previous studies (Agrawal *et al*., 2018; Edea *et al*., 2012).

Results of genetic distance and genetic differentiation estimates indicates, the populations under study are very closely related which was in agreement with previous SNP and microsatellite-based investigations (Mandefro *et al*., 2021; Edea *et al*., 2013; Zerabruk *et al*., 2012; Dadi *et al*., 2008). The moderate level of differentiation observed among sampled populations could be attributed to common ancestry and admixture among population (Bradley *et al*., 1998).

5.1.4. Evolutionary Analysis

It has been shown that a Ka/Ks ratio of or close to 1 indicates no strong selection pressure, a ratio larger than 1 indicates that the protein is subjected to positive selection, whereas less than 1 indicates the presence of purifying selection (Waterston *et al*., 2002). Positive selection is observed less often than purifying selection and most mammalian genes are under strong to moderate purifying selection (Waterston *et al*., 2002). In the present study the ratio of Ka/Ks was evaluated and the values recorded were less than one for all the breeds. Results revealed that cattle *DGAT1* gene have been subjected to purifying selection (Ka/Ks=0.00–0.708). The average Ka/Ks ratio for *DGAT1* gene was 0.564, indicating that the evolution of bovine *DGAT1* is largely shaped by strong purifying selection through the removal of alleles that are deleterious, resulting in stabilizing selection in the phenotypic outcomes. *DGAT1* genes were

under purifying selection in Ethiopian cattle. Purifying selection against newly arising deleterious mutations is essential to preserve biological function of *DGAT1* gene among cattle populations of Ethiopia.

A close genetic relationship between the sampled cattle populations from the inferred median joining-network tree was observed. The current result is consistent with the report of other researchers (Edea *et al*., 2012; Dadi *et al*., 2008;). The detection of consensus region among different breeds/species for *DGAT1* exon-8 region highlights the sequence fragments which have been subjected to least evolutionary pressure.

From evolution point of view, the low milk yield of sampled zebu populations observed could be due to the absence of the genotype AA, which is associated with high milk yield in *Bos taurus* cattle and their crosses. Prevalent of *DGAT1* K allele in the sampled zebu populations give insight into the evolution of *DGAT1* gene. The history of domestication of cattle falls out more than 10,000 years ago from *B.primigenius* in the Fertile Crescent. The present-day indicine cattle got separated from taurine cattle during the second domestication event from *Bos primigenius namadicus* that occurred in the Indus valley (Pitt *et al*. 2019). At the K232A locus of the *DGAT1* gene, the *DGAT1* K allele is ancestral to the *DGAT1* A allele (Grisart *et al*. 2002; Winter *et al*. 2002). Evolutionary divergence results revealed the presence of *DGAT1* K allele at K232A position in all species except in *B. taurus*, strengthening the hypothesis that the *DGAT1* K allele is the ancestral and existed before the divergence of bovine species. Further, the A variant was also observed to disrupt an exon splice enhancer that may likely be the cause of variation in lactation effects (Fink *et al*. 2020). These results together suggest that the drift driven mutations resulting in a non-synonymous K232A change in the protein aroused in taurine cattle post-domestication and a splice variation might be under strong selection leading to the more prevalence of the variant in taurine and their crosses.

5.2. Effect of breeds and *DGAT1 K232A* on milk yield and composition traits in the studied cattle populations

5.2.1. Effect of breed on milk yield and composition traits

The milk fat content of studied breeds was within the range for the milk composition standard requirement of cows (3.18-6.21%). Zebu cows can give milk containing upto 7% fat (O'Mahony, 1988) . Significant differences were observed for all milk components among the studied breeds. Similar study reported the effect of breed on milk composition in South African (Myburgh *et al*., 2012) and Benin cattle breeds (Houaga *et al*., 2017). Boran cows in Ethiopia presented higher fat content (6.21%) than the values of 2.68% for Boran in South Africa and Borgou and White Fulani cows (4.78 and 4.85%, respectively) in Benin (Houaga *et al*., 2017; Myburgh *et al*., 2012).

Thus, according to the present study Boran * HF cross cows are most favored for daily milk yield whereas Boran cows are most favored for their milk fat percent. The observed differences between breeds would therefore be due to their genetic background. Furthermore, the forage nutrient composition under natural grazing could be the reason for the variation in milk components among the cattle breeds. The investigated cattle breeds are reared in different agro-ecological zones with different floristic compositions. Cow's milk composition could be affected by forage species and variety, climate and stage of growth.

5.2.2. Allele and Genotype frequencies of *K232A* protein variant

The variants of the analyzed gene were found in all the five breeds and therefore it is the allelic distribution that characterizes the differences between breeds. The present study revealed that the lysine variant of the *DGAT1* gene, i.e., the K232 allele, was common in the zebu populations. This is in agreement with studies in other *Bos indicus* breeds where the frequency of the lysine variant K232 ranged from 0.80 to 1 (Four African Zebu (Banyo Gudali, Kenana, Butana and White Fulani) (Elzaki *et al*., 2022; Abu *et al*., 2015; Rahamattalla *et al*., 2015) and three Indian cattle (Sahiwal, Gir and Kankrej) (Gothwal *et al*.,2022; Patel and Chauhan, 2017; Ganguly *et al*., 2013) and was comparable to the frequencies of K232 in Borgou (0.77) and White Fulani (0.92) cattle of Benin (Houaga *et al*.,

2018). The higher K allele seen in those three sampled breeds might reflect their exceptional adaptive value.

In the present study, the observed frequencies of the K allele (0.50–0.97) were higher than those observed (0.210–0.380) for the native Turkish and European *Bos taurus* breeds (Kaupe *et al*., 2004), with the exception of the Modicana and Jersey breeds (Valenti *et al*., 2019). The finding in the present study indicated the tendency of fixation of *DGAT1* K allele in Horro cattle did not support its use as a reliable universal marker for milk yield and composition traits for the population. Fixation of the *DGAT1* K allele has been noticed in ongole breeds of India by Krovvidi et al., (2021), while Kaupe et al. (2004) found fixation of the *DGAT1* A allele in five *Bos taurus* breeds (Belgian Blue beef, Gelbvieh, Hereford, Pinzgaurer and Slavonian Syrmian). The high proportion of the lysine variant K232 observed in the current study might contribute to the higher fat content in native cattle in Ethiopia and other *Bos indicus* breeds compared to Holstein–Friesian cross cattle. The frequency of the lysine variant K232 in the investigated Boran * HF crosses was similar to studies in other Holstein–Friesian populations or their crosses where it ranged from 0.167 to 0.63 (Krovvidi *et al*., 2021, Li *et al*., 2020; Bhat *et al*., 2017; Komisarek and Kolenda, 2016; Ahani *et al*., 2015; Akyuz *et al*., 2015). The differences in the frequency of the lysine variant K232 among breeds might result from different sample size and the considerable Holstein contribution to the blood level of Boran * HF cross cattle.

Allele A is a rare allele detected in Boran, Begait, Fogera and Boran * HF crosses. A locus has been considered polymorphic if the most common allele has a frequency of ≤ 0.95 and therefore the less common allele (at a locus with two alleles) has a frequency of ≥ 0.05 (Hartl and Clark, 1997). Accordingly, in the present study *DGAT1* locus was not found to be polymorphic in Horro cattle populations may be that one animal among the sampled population had the A allele. The A allele frequency observed in the zebu populations was relatively comparable to the indigenous cattle of Sudan (0.037-0.15) (Abu *et al*., 2015). The A allele frequency observed in Boran * HF cross cattle was higher than the report of Li et al.(2021) for Kiwi cross cattle from New Zealand and relatively lower than the Butana-Holstein crossbred cattle of Sudan (Elzaki *et al*., 2022). To reiterate, the low A allele frequency observed in the zebu populations of Ethiopia might be attributed to *taurus* cattle

introgression. The KK and KA genotypic frequencies observed in the present study ranged from 0.50-0.94 and 0.03-0.50 for the zebu populations, respectively, which is comparable to most *Bos indicus* cattle (Elzaki *et al*., 2022; Houaga *et al*., 2017; Rahamattalla *et al*., 2015; Abu *et al*., 2015). An excess of heterozygotes over homozygotes in Boran * HF cross breeds indicates that heterozygotes are preferred which could be the effect of natural selection over the years of evolution and it would be effective to select for or against polymorphism.

5.2.3. Effects of *DGAT1 K232A* genotypes on daily milk yield and composition traits

The association study confirmed that the *DGAT1 K232A* marker had significant effects on daily milk yield, milk fat and lactose content in the investigated cattle. Higher average daily milk yield and low fat percentage is attributed to A allele inherited from *Bos taurus* cattle in the crossbred while differences among the local breeds attributed to their nutritional regimes. Similar effects of the KK genotype on milk composition traits and the higher effect of the AA and KA genotype on average milk yield were observed on Holstein, Holstein crosses, Jersey and its local Kashmiri cross, Holstein ×Jersey crosses, Gir and Kankrej cattle populations in different parts of the world (Li *et al*., 2021; Bhat *et al*., 2017; Patel andChauhan, 2017; Akyuz *et al*., 2015; Molee *et al*., 2012). In contrast, Manga and Ríha (2011) reported higher milk yield by allele *DGAT1* K (lysine variant) in Holstein cows.

Substitution effect of K allele showed that the one copy of this allele leads to significant (p<0.01) increase of fat and protein content and, decrease of milk yield, lactose and solid not fat content in all breeds considered. Comparable observation on average effect of allele substitution were reported by Grisart et al.(2002), who observed an allele substitution effect of -130 litre milk and +5.76kg milk fat in HF cattle and -110 litre milk and +3.30kg milk fat in jersey cattle on substitution of A allele with K allele in the population. Moreover, an increase of 548kg in milk yield with simultaneous decrease of 15.4kg milk fat was reported in HF animals homozygous for A allele of *DGAT1* gene (Hradecká *et al*., 2008).

Phenotypic correlation clearly indicated that the fat percentage had a moderate negative correlation (p<0.05) and a highly negative correlation (p<0.01) with daily milk yield in indigenous cows and Boran * HF cross cows, respectively. The lactose percentage in indigenous and Boran * HF cross cows had a high negative correlation with fat content. This

result is in agreement with previous reports by Fox (2009), who indicated the strong negative relationship between lactose and fat content in milk across species. Accordingly, the lysine variant K232 decreased lactose content in milk across breeds in the present study.

5.3. Sequence polymorphisms within exon-4 of the *PRL* gene and their effect on milk yield and composition traits in the studied cattle populations

5.3.1. Single nucleotide polymorphisms in *PRL* gene among the studied populations

PRL gene maintains lactation and primarily responsible for the synthesis of the main components of milk and is involved in each stage of milk protein genes expression (Nayeri and Stothard, 2016).

SNPs occurring within the *PRL* gene may influence the chemical composition of milk or at least be an effective DNA marker of a subregion of dairy cattle genome. Taking that into account, exon-4 of the *PRL* gene was screened to detect SNPs. Using direct sequencing methods, three SNPs found (Table 14). Two of them, i.e. g.8377G>C and g.8398A>G overlapped with mutations reported previously by Brym et al.(2005) and Sasavage et al. (1982). In several studies, polymorphic sites have been discovered in exon-4 of *PRL* gene's and flanking regions. Comparable number of SNPs was detected by Kaminski et al.(2005) and Brym et al.(2005) in agreement with the current study.

In the present analysis the less number of genotypes observed for the SNP g.8323T>A might be due to un-sampled populations and the missing genotype might be the unfavoured genotype. The remaining two SNPs exhibited all the three expected genotypes. SNP in positions g.8323T>A is breed specific detected in Begait and Fogera populations. Breed specific SNPs were only polymorphic within a breed, and one of the alleles is fixed in other breeds (Pant *et al*., 2012). Mutations in the four positions were detected by sequence comparison with the reference sequence AF426315 deposited in the GenBank (Brym *et al*., 2005).

The present study showed that SNP g.8398A>G, with allele frequencies of A (0.449) and G (0.551), whereas its genotypic variants GA (0.390) and GG (0.356) showed similar pattern.

Comparable allele frequencies in crossbred HF, Achai and Deoni breeds were reported by Shah et al.(2021), Ishaq et al.(2012) and Das et al.(2012), respectively. The studied breeds in the present study showed considerably less variation in gene frequency with preponderance of heterozygotes in the population (Table 14). The minor differences in allele frequencies among different breeds may have resulted from different histories of selection during course of evolution and utilization in different breeding programmes. However, the relatively higher percentage of heterozygotes in Boran breed could be due to over exploitation of the pure bred animals of this precious genetic resource. Similarly, higher percentages of heterozygotes were reported in Kashmir jersey and Deoni breeds by Shah et al.(2021) and Das et al.(2012), respectively. In contrast, Khaizaran and Al-razem(2014), Ishaq et al.(2012) and Boleckova et al.(2012) reported a low frequency of heterozygotes for Palestinian*HF, Sahiwal and Czech Fleckvieh breeds, respectively.

In contrast, the G allelic variant was observed frequently in nature in most European breeds of cattle. The observed allele frequency of G (0.551) is lower than to those reported previously by other researchers (Shah *et al*., 2021; Ozdemir, 2020; Thuy *et al*., 2018; Sonmez and Ozdemir, 2017; Khaizaran and Al-razem, 2014; Boleckova *et al*., 2012; Ishaq *et al*., 2012). This could be due to differences in the breeds studied. The breed specific variation in gene frequency was pronounced in exon-4 of *PRL* gene. Similarly, Brym et al. (2005) reported a quite variable nature of A and G allele frequencies in two different European breeds with frequencies of 0.113 and 0.706 for A allele of Black and White cows; and Jersey cattle, respectively. A lower frequency of the G allele was observed in Jersey breed (Dybus *et al*., 2005). The frequency of G allele varied from 0.29 in Jersey cows (Brym *et al*., 2005) to 0.94 in Palestinian *HF cross cows (Khaizaran and Al-razem, 2014) which could be explained by different history of breeds, long term geographical isolation and selection towards high fat and protein percents' of milk.

5.3.2. Genetic diversity in the studied populations

Haplotype based analysis of the *PRL* gene was carried out in this study. The result indicated that only three of them are common for all populations studied. In the shared haplotypes,

relatively highest haplotype frequencies were obtained in Boran and Boran * HF cross cattle population than the others.

The results of pairwise F_{ST} values indicated low genetic differentiation between the studied breeds. Considerably high similarities between the considered cattle breeds, attested by the F_{ST} values and low genetic distances, were suggestive of the origin of the population from a common ancestor. While, F_{ST} value between Fogera and Boran * HF cross (0.195) were moderate according to Weir (1984) and kalinowski (2002) which suggested the highest genetic distance (F_{ST}) to be higher than 0.25, moderate to be between 0.05 and 0.25 and the lowest estimate below 0.05. This could be due to less genetic material exchange (gene flow) between Fogera and Boran * HF cross.

Assuming that all populations have similar limitations, HWE could be compared. The results showed that all populations are in equilibrium for the locus analyzed. HWE test for exon-4 of *PRL* gene reflected similar pattern in different SNPs found (Table 14). *PRL* genotypes did not deviate from HWE (p>0.05) for SNPs g.8323T>A, g.8377 G>C and g.8398A>G. The studies in Iranian Holstein and Turkey East Anatolian Red cattle denoted maintenance of HWE (p>0.05) (Sonmez and Ozdemir, 2017; Mehmannavaz *et al*., 2009). Beside, Boleckova et al.(2012) observed distribution of genotypes according to HWE in Czech Fleckvieh cattle. On the other hand, Brym et al. (2005) observed non-significant deviation in gene and genotypic frequency of exon-4 of *PRL* gene in Black and White; and Jersey cattle. Beside, Das et al.(2012), report revealed genotypic frequencies of Deoni cattle breed deviated from HWE probabilities. The variability of results in different studies could be due to differences in breed (Brym *et al*., 2005). The results showed that *PRL* genotypes was in genetic equilibrium in the studied populations, which implied that these breeds had less pressure in long-term of the artificial selection and it had no influence on *PRL* gene.

However, different F_{IS} values (inbreeding coefficient) were observed in the three SNPs (Table 16). The F_{IS} value in the present study was low for SNP g.8323T>A and negative for SNPs g.8377G>C and g.8398A>G. Similarly, Sonmez and Ozdemir (2017) found positive F_{IS} value for East Anatolian Red cattle (0.072) in Turkey. Negative F_{IS} value suggests a heterozygosity in SNPs g.8377G>C and g.8398A>G. No evidence of inbreeding was found in the sampled

populations, based on the results obtained from analysis of averaged F_{IS} across subpopulations. F_{IS} does not directly detect inbreeding. However, an increase in the fraction of homozygotes (with a positive F_{IS}) would be expected.

Slightly informative PIC was detected for g.8323T>A, while reasonably informative PIC was noted for g.8377G>C and g.8398A>G SNPs. The results indicated that the sampled populations were abundant in genetic diversity and had strong potential for further selection (Table 16). SNPs g.8377G>C and g.8398A>G had closest MAF value, reasonably informative PIC value, and closest high heterozygosity (H_e = 0.485 and 0.494), indicating higher mutation frequencies of these SNPs (Table 14 and 16).

5.3.3. Phylogenetic relationships and Median joining Network

The construction of phylogenetic tree was on the basis of genetic distances, depicting phylogenetic relationships among the *PRL* nucleotide sequences of the investigated populations. The phylogenetic tree analysis evidenced that the cattle breeds considered in our study were relatively far from Buffalo, goat and sheep. At this juncture, the number of alleles identified at the *PRL* locus from different species demands special mention. The alleles identified in one species are more closely related to the alleles in a closely related species than to the other alleles in the same species. The species wise clustering might be due to species specific residues and such patterns of the sequences may be explained by gene conversion (Takahashi and Nei, 2000).

Horro and Fogera belong to Zenga—the result of crossing between Zebu and Sanga, while Boran and Begait find their place among the large East African Zebu. The relationship among the five cattle breeds via phylogenetic tree analysis revealed that Begait and Boran were more closely clustered compared to the Zenga types. Based on the analysis of the median-joining network tree, the haplotypes H-2 and H-3 with the highest number of individual sequences could be assumed to be the ancestral haplotypes for the *PRL* gene (Table 15). A greater in-depth evolutionary divergence study would surely assist in molecular breeding programs and facilitate genetic improvement of the indigenous cattle breeds.

5.3.4. *PRL* g.8398A>G genotypes effects on milk yield and composition traits

Daily milk yield, fat and SNF percentage present in milk varied significantly among detected genotypes within each breed (Table 17). The daily milk yield of the AA genotype ranged from 2.61 ± 0.22 L/day (HR) to 8 ± 0.32 L/day (BHC). Genotype GG was observed to be significantly better than AA genotype in all breeds considered for milk fat and SNF percentage, indicating the substitution effect of the mutant allele G.

The values of milk protein, lactose and ash were non-significantly varied between genotypes in all the studied breeds. The results suggest that exon-4 of the *PRL* gene imparts variance in average daily milk yield, milk fat and SNF percentage. An overall positive effect of the GG genotype irrespective of breed was observed on milk fat percentage which is in agreement with similar findings on Black and White cattle of Poland (Brym *et al*., 2005). As also observed in the present study, it was reported that animals having AA genotype yield more milk and protein (Alipanah et al., 2007) whereas more fat was observed in animals of GG and AG genotypes (Brym *et al*., 2005; Khatami *et al*., 2005). The effect of allele A on milk yield was indirectly revealed by Othman et al. (2011), who only observed G allele in buffalo population producing fat rich milk. In contrast to the current result, genotype GG was associated with highest milk yield whereas genotype AG was associated with highest protein percentage in Deoni cattle of India (Das *et al*., 2012). Dong et al. (2013) could not observe any association of different genotypes with percentage of milk fat and proteins.

The significant effect of *PRL* genotypes observed in the present study especially for daily milk yield, milk fat and SNF traits might be due to combined effect on numerous biological processes due to pleiotropic nature of *PRL* gene (Bole-Feysot *et al*., 1998). The non-observance of significant effect of protein and lactose with the different genotypic pattern of exon-4 of *PRL* gene could be due to interactions with different background genes in the five cattle population studied.

Quantitative effect of g.8398A>G genotypes showed that one copy of G allele leads to significant (p<0.05) increase of fat, SNF and protein content upto 0.52%, 0.39% and 0.40%, respectively. Whereas, significant decrease of daily milk yield up to 2.08L in sampled populations. Similarly Dybus et al. (2005) observed cows with the AA genotype produced milk with lower fat content (-0.19 and -0.18%) than AG and GG individuals, respectively in Polish black and white cattle.

6. CONCLUSIONS AND RECOMMENDATIONS

The role of *diacylglycerol O-acyl transferase* 1 and *prolactin* gene in determining milk quantity and quality is very important. Two proven genomic regions, viz., exon-8 of *DGAT1* and exon-4 of *PRL* genes were selected on the basis of their significant role in lactation. Sequencing of the selected gene products was carried out to detect the diversity and frequency of significant desirable variants of selected cattle populations in Ethiopia. The present study reported for the first time bovine *DGAT1* and *PRL* genetic diversity, allele and genotype frequency distribution and *DGAT1* K232A and *PRL* SNPs genotypes effects on milk yield and constituents in Ethiopian zebu and their crosses.

The overall diversity indices showed *DGAT1* locus was polymorphic in Boran, Begait, Fogera and Boran * HF cross populations. *DGAT1* gene was under purifying selection and the heterozygosity and negative inbreeding coefficient revealed genetic variation in the studied cattle populations. Allele K and genotype KK were found in higher proportion in the zebu populations. The association study indicated that the *DGAT1 K232A* marker had significant effects on daily milk yield, milk fat and lactose content in Boran, Begait, Fogera and Boran * HF cross populations.

Substantial genetic variation in the form of heterozygosis, PIC and inbreeding coefficient was detected for *PRL* exon-4 locus. Three SNPs were identified in the coding region and g.8323T>A, showed missense mutation. SNP g.8398A>G, with allelic frequencies of A (0.449) and G (0.551), only have significant effects ($p< 0.05$) on average daily milk yield, fat and solid not fat (SNF) percentage in all studied cattle populations.

The effects of *DGAT1* and *PRL* genotypes on milk yield and constituents illustrated worth of these genes for marker-assisted selection for milk traits. A critical analysis of the two important genes responsible for milk yield and milk composition traits in the present study concluded that the studied populations were naturally evolved and adapted for optimized milk production without comprise for production of milk fat which could be seen as survival advantage through production of energy rich fat required for neonatal survival during period of scarcity and drought.

Therefore, based on the present results the following recommendations were forwarded;

- The information from sequence polymorphism of *DGAT1* and *PRL* genes in cattle populations of Ethiopia could be used for dairy breeding to enhance the production potential and to accelerate the rate of genetic gain after validation in larger population.

- *DGAT1* K232A and *PRL* g.8398A>G found to influence bovine milk yield and composition traits in this study may be used as possible candidate marker for marker assisted selection programs in the cattle populations of Ethiopia, However, a robust candidate gene approach is required to develop a molecular marker score with respect to milk traits and to frame an integrated breeding policy for faster genetic improvement of dairy animals and the interaction effect between this genes and feed management and other environments should be studied.

- Results indicated within population's variation which could be resulted from randomly purchased populations from local markets, which suggest within breed improvement may be feasible. Therefore, strong animal regulatory policy and strategy are imperative parallel to designing effective informed breeding schemes.

- The results showed a decrease in population size of Begait and overexploitation of Boran cattle. Therefore, owing to such observations, we suggest that the two lowland zebu populations should be the prime targets for future breed conservation programs through the most economically viable in-situ preservation of the pure live animals within their production environment

- More mutations should be identified in the *DGAT1* and *PRL* genes in future for possible linkage between themselves and for their influence on milk traits across different promising cattle populations. However, sequencing the whole length of *DGAT1* and *PRL* genes and testing both the haplotype and measures of Linkage Disequilibrium with more sample size may help to arrive at firm conclusion.

- Functional validation and comparative expression profiling of *DGAT1* and *PRL* genes should be investigated in diverse physiological periods (dry, prenatal and milking) in mammary glands of promising cattle populations of Ethiopia.

7. REFERENCES

Abu, S. M., Said Ahmed, A., Ahmed, M. K. A., Reissmann, M., Brockmann, G. A., and Rahmatalla, S. A. (2015). Variants of the acyl-CoA:diacylglycerol acyltransferase 1 (*DGAT1*) Gene in Sudanese Dairy Cattle (Kenana and Butana). *Journal of Molecular Genetics*, *7*(1), 5–9. https://doi.org/10.36478/jmolgene.2015.5.9

Agrawal, V., Gahlot, G., Ashraf, M., Kumar, A., and Dhakad, G. (2018). Genetic Analysis of *DGAT1* Loci Related to Milk Production Traits in Native Sahiwal Cattle. *International Journal of Livestock Research*, *8*(5), 136. https://doi.org/10.5455/ijlr.20170928050917

Ahani, S., Mashhadi, M. H., Nassiri, M. R., Aminafshar, M., and Haddadi, M. (2015). Characterization of single nucleotide polymorphism in diacylglycerol acyltransferase (*DGAT1*) gene loci of Iranian Holstein cattle. *Research Opinions in Animal and Veterinary Sciences*, *5*(5), 231–236.

Akyuz, B., Agaoglu, O. K., Akçay, A., and Agaoglu, A. R. (2015). Effects of *DGAT1* and *GH1* polymorphism on milk yield in Holstein cows reared in Turkey. *Slovenian Veterinary Research*, *52*(4), 185–191.

Alipanah, M., Kalashnikova, L. A., and Rodionov, G. V. (2008). Kappa-casein and *PRL*-Rsa I genotypic frequencies in two Russian cattle breeds. *Archivos de Zootecnia*, *57*(218), 131–138.

Alipanah, M., Kalashnikova, L., and Rodionov, G. (2007). Association of genetic variants of the *prolactin* gene with milk production traits in Russian Red Pied cattle. *Iranian Journal of Biotechnology*, *5*(3), 158–161. https://doi.org/10.1017/s1752756200020597

Amos, W. (1998). Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1366). https://doi.org/10.1098/rstb.1998.0200

Anton, I., Kovács, K., Fésüs, L., Várhegyi, J., Lehel, L., Hajda, Z., Polgár, J. P., Szabó, F., and Zsolnai, A. (2008). Effect of *DGAT1* and *TG* gene polymorphisms on intramuscular fat and on milk production traits in different cattle breeds in Hungary. *Acta Veterinaria Hungarica*, *56*(2). https://doi.org/10.1556/AVet.56.2008.2.5

Asmarasari, S. (2013). *The relationship of Diacylglicerol acyltransferas (DGAT1) Gene Diversity to Friesian Holstein Dairy Cattle's Milk Production and Fatty Acid Profile* [Bogor Agricultural University, Bogor, Indonesia]. http://repository.ipb.ac.id/ handle/ 123456789/67075

Assefa, A., and Hailu, A. (2018). Ethiopian indigenous cattle breed's diversity, distribution, purpose of keeping, and their potential threats. *J. Biodiv. Innovativ. Assoc.*, *7*(5), 770–789.

Bandelt, H. J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036.37-48

Banos, G., Woolliams, J. A., Woodward, B. W., Forbes, A. B., and Coffey, M. P. (2008). Impact of single nucleotide polymorphisms in leptin, leptin receptor, growth hormone receptor, and diacylglycerol acyltransferase (*DGAT1*) gene loci on milk production, feed, and body energy traits of UK dairy cows. *Journal of Dairy Science*, *91*(8). https://doi.org/10.3168/jds.2007-0930

Barker, J. S. F. (1994). A global protocol for determining genetic distances among domestic livestock breeds. *5th World Congress on Genetics Applied to Livestock Production*, *21*(7-12 August), 501–508.

Bennewitz, J., Reinsch, N., Paul, S., Looft, C., Kaupe, B., Weimann, C., Erhardt, G., Thaller, G., Kühn, C., Schwerin, M., Thomsen, H., Reinhardt, F., Reents, R., and Kalm, E. (2004). The *DGAT1* K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. *Journal of Dairy Science*, *87*(2). https://doi.org/10.3168/jds.S0022-0302(04)73182-3

Bern, H. A., and Nicoll, C. S. (1968). The comparative endocrinology of *prolactin*. *Recent Progress in Hormone Research*, *24*, 681–720. https://doi.org/10.1016/b978-1-4831-9827-9.50019-8

Bernichtein, S., Touraine, P., and Goffin, V. (2010). New concepts in *prolactin* biology. *Journal of Endocrinology*, *206*(1), 1–11. https://doi.org/10.1677/JOE-10-0069

Berry, D. P., Howard, D., O'Boyle, P., Waters, S., Kearney, J. F., and McCabe, M. (2010). Associations between the K232A polymorphism in the diacylglycerol-O-transferase 1 (*DGAT1*) gene and performance in Irish Holstein-Friesian dairy cattle. *Irish Journal of Agricultural and Food Research*, *49*(1), 1–9.

Bhat, S. A., Ahmad, S. M., Ganai, N. A., Khan, S. M., Malik, A., Shah, R. A., and Iqbal, Z. (2017). Association of *DGAT1*, beta-casein and leptin gene polymorphism with milk quality and yield traits in Jersey and its cross with local Kashmiri cattle. *Journal of Entomology and Zoology Studies*, *5*(6), 557–561.

Bole-Feysot, C., Goffin, V., Edery, M., Binart, N., and Kelly, P. A. (1998). *prolactin* (*PRL*) and its receptor: Actions, signal transduction pathways and phenotypes observed in *PRL* receptor knockout mice. *Endocrine Reviews*, *19*(3), 225–268. https://doi.org/10.1210 /edrv.19.3.0334

Boleckova, J., Matejickova, J., Stipkova, M., Kyselova, J., and Barton, L. (2012). The association of five polymorphisms with milk production traits in Czech Fleckvieh cattle. *Czech Journal of Animal Science*, *57*(2), 45–53. https://doi.org/10.17221/5131-cjas

Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, *32*(3), 314–331.

Bradley D.G., MacHugh D.E., Cunningham P. and Loftus R.T. (1996). Mitochondrial diversity and the origins of African and European cattle. Proceedings of the National Academy of Sciences of the United States of America 93, 5131–5.

Bradley, D. G., Loftus, R. T., Cunningham, P., and Machugh, D. E. (1998). Genetics and domestic cattle origins. *Evolutionary Anthropology*,*6*(3). https://doi.org/ 10.1002/(SICI) 1520-6505(1998)6:3<79::AID-EVAN2>3.0.CO;2-R

Brown, J. H. (2000). Phylogeography: The History and Formation of Species . John C. Avise . *The Quarterly Review of Biology*, *75*(4). https://doi.org/10.1086/393655

Brym, P., Kamiński, S., and Wójcik, E. (2005). Nucleotide sequence polymorphism within exon 4 of the bovine *prolactin* gene and its associations with milk performance traits. *Journal of Applied Genetics*, *46*(2), 179–185.

Camper, S. A., Luck, D. N., Yao, Y., Woychik, R. P., Goodwin, R. G., Lyons, R. H., and Rottman, F. M. (1984). Characterization of the Bovine *prolactin* Gene. *DNA*, *3*(3), 237–249. https://doi.org/10.1089/dna.1.1984.3.237

Cao, X., Wang, Q., Yan, J. Bin, Yang, F. K., Huang, S. Z., and Zeng, Y. T. (2002). Molecular cloning and analysis of bovine *prolactin* full-long genomic as well as cDNA sequences. *Acta Genetica Sinica*, *29*(9), 768–773.

Caroli, A. M.Chessa,S. and Erhardt, G. J. (2009). Invited review: milk protein polymorphisms in cattle: effect on animal breeding and human nutrition. *Journal of Dairy Science*, 92(11), 5335– 52. doi: 10.3168/jds.2009-2461.

Cases, S., Smith, S. J., Zheng, Y. W., Myers, H. M., Lear, S. R., Sande, E., Novak, S., Collins, C., Welch, C. B., Lusis, A. J., Erickson, S. K., and Farese, R. V. (1998). Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(22). https://doi.org/10.1073/pnas.95.22.13018

Collier, R. J., McNamara, J. P., Wallace, C. R., and Dehoff, M. H. (1984). A review of endocrine regulation of metabolism during lactation. *Journal of Animal Science*, *59*(2), 498–510. https://doi.org/10.2527/jas1984.592498x

Collins, D. W., and Jukes, T. H. (1994). Rates of transition and transversion in coding sequences since the human- Rodent divergence. *Genomics*, *20*(3). https://doi.org/ 10.1006/ geno. 1994.1192

CSA (2021).Agricultural Sample Survey 2020/21 (2013 E.C). Report on Livestock and Livestock Characteristics.Volume II, Statistical Bulletin 589, Central Statistical Authority, Addis Ababa, 13.

https://www.scirp.org/(S(czeh2tfqw2orz553k1w0r45))/reference/referencespapers.aspx?referenceid=3214558

Curnow, R. N., and Wright, S. (1979). Evolution and the Genetics of Populations, Volume 4: Variability within and among Natural Populations. *Biometrics*, *35*(1). https://doi.org /10.2307/2529965

Dadi, H., Tibbo, M., Takahashi, Y., Nomura, K., Hanada, H., and Amano, T. (2008). Microsatellite analysis reveals high genetic diversity but low genetic structure in Ethiopian indigenous cattle populations. *Animal Genetics*, *39*(4). https://doi.org/10.1111/ j.1365-2052.2008.01748.x

Dahl, G. E. (2008). Effects of short day photoperiod on *prolactin* signaling in dry cows: a common mechanism among tissues and environments? *Journal of Animal Science*, *86*(13 Suppl), 10–14. https://doi.org/https://doi.org/10.2527/jas.2007-0311

Das, N. D., Hatkar, D. N., Hari, V. G. S., Srinivas, B. V, Kaliaperumal, R., Reddy, O. A., and Krishnamurthy, L. (2012). Genetic polymorphisms of exons 3 and 4 of *prolactin* (*PRL*) gene in Deoni cattle breed and their association with milk production trait. *International Journal of Livestock Research*, *2*(3), 120–126.

Dokso, A., Ivanković, A., Zečević, E., and Brka, M. (2015). Effect of *DGAT1* gene variants on milk quantity and quality in Holstein, Simmental and Brown Swiss cattle breeds in Croatia. *Mljekarstvo*, *65*(4). https://doi.org/10.15567/mljekarstvo.2015.0403

Dong, C. H., Song, X. M., Zhang, L., Jiang, J. F., Zhou, J. P., and Jiang, Y. Q. (2013). New insights into the *prolactin*-RsaI (*PRL*-RsaI) locus in Chinese Holstein cows and its effect on milk performance traits. *Genetics and Molecular Research*, *12*(4), 5766–5773. https://doi.org/10.4238/2013.November.22.3

Duchemin, S., Bovenhuis, H., Stoop, W. M., Bouwman, A. C., van Arendonk, J. A. M., and Visker, M. H. P. W. (2013). Genetic correlation between composition of bovine milk fat in winter and summer, and *DGAT1* and *SCD1* by season interactions. *Journal of Dairy Science*, *96*(1). https://doi.org/10.3168/jds.2012-5454

Dybus, A. (2002). Associations of growth hormone (GH) and *prolactin* (*PRL*) genes polymorphisms with milk production traits in Polish Black-and-White cattle. *Animal Science Papers and Reports*, *20*(4), 203–212.

Dybus, A., Grzesiak, W., Kamieniecki, H., Szatkowska, I., Sobek, Z., Błaszczyk, P., Czerniawska-Piątkowska, E., Zych, S., and Muszyńska, M. (2005). Association of genetic variants of bovine *prolactin* with milk production traits of Black-and-White and Jersey cattle. *Archives Animal Breeding*, *48*(2), 149–156. https://doi.org/10.5194/aab-48-149-2005

Edea, Z., Dadi, H., Kim, S. W., Dessie, T., and Kim, K.-S. (2012). Comparison of SNP Variation and Distribution in Indigenous Ethiopian and Korean Cattle (Hanwoo) Populations. *Genomics andInformatics*, *10*(3). https://doi.org/10.5808/gi.2012.10.3.200-5

Edea, Z., Dadi, H., Kim, S. W., Dessie, T., Lee, T., Kim, H., Kim, J. J., and Kim, K. S. (2013). Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean hanwoo breeds using SNP markers. *Frontiers in Genetics*, *4*(35), 1–9. https://doi.org/10.3389/fgene.2013.00035

Eggen, A. (2012). The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers*, *2*(1), 10–15.

Elliott RB, Harris D P, Hill JP, Bibby NJ, Wasmuth HE. (1999). Type I (insulin-dependent) diabetes mellitus and cow milk: casein variant consumption. *Diabetologia*, 42(3), 292-6. doi: 10.1007/s001250051153

El-Magd, Mohammed Abu, Abo-Al-Ela, H. G., El-Nahas, A., Saleh, A. A., and Mansour, A. A. (2014). Effects of a novel SNP of IGF2R gene on growth traits and expression rate of IGF2R and IGF2 genes in gluteus medius muscle of Egyptian buffalo. *Gene*, *540*(2). https://doi.org/10.1016/j.gene.2014.02.059

Elzaki, S., Korkuć, P., Arends, D., Reissmann, M., and Brockmann, G. A. (2022). Effects of *DGAT1* on milk performance in Sudanese Butana × Holstein crossbred cattle. *Tropical Animal Health and Production*, *54*(2).142-7. https://doi.org/10.1007/s11250-022-03141-7

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

FAO. (2013). *Milk and dairy products in human nutrition* (E. Muehlhoff, A. Bennett, andD. McMahon (eds.)). Food and Agriculture Organization of the United Nation.

Faraj, S. H., Ayied, A. Y., and Seger, D. K. (2020). *DGAT1* gene polymorphism and its relationships with cattle milk yield and chemical composition. *Periodico Tche Quimica*, *17*(35). https://doi.org/10.52571/ptq.v17.n35.2020.16_faraj_pgs_174_180.pdf

Farnir, F., Grisart, B., Coppieters, W., Riquet, J., Berzi, P., Cambisano, N., Karim, L., Mni, M., Simon, P., and Wagenaar, D. (2002). Simultaneous mining of linkage and LD to fine map QTL in outbred half-sib pedigrees: Revisiting the location of a QTL with major effect on milk production on bovine chromosome 14. *Genetics*, *161*(1), 275–287.

Freeman, M.E., L.B. Kanyicska, A. Lerant, and G. Nagy (2000). Prolactin: Structure, Function, and Regulation of Secretion. Physiological Reviews, 80:1523-1631.

Fink, T., Lopdell, T.J., Tiplady, K., Handley, R., Johnson, T.J.J.,Spelman, R.J., Davis, S.R., Snell, R.G. and Littlejohn, M.D.,2020. A new mechanism for a familiar mutation – bovine *DGAT1* K232A modulates gene expression through multi-junction exon splice enhancement, BMC Genomics, 21, 591

Fox, P. F., and McSweeney, P. L. H. (Eds.). (2009). Lactose: Chemistry and properties. In *Advanced dairy chemistry volume 3: lactose, water, salts and minor constituents* (Vol. 3, pp. 1–15). Springer. https://doi.org/10.1007/978-0-387-84865-5_1

Frankham, R., Ballou, J. D., Briscoe, D. A., and McInnes, K. H. (2002). *Introduction to Conservation Genetics*. Cambridge University Press. https://doi.org/10.1017/ cbo97805 11808999

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*(2).https://doi.org/10.1093/ genetics/147.2.915

Furbass, R., Winter, A., Fries, R., and Kuhn, C. (2006). Alleles of the bovine *DGAT1* variable number of tandem repeat associated with a milk fat QTL at chromosome 14 can stimulate gene expression. *Physiological Genomics*, *25*(1), 116–120.

Ganguly, I., Kumar, S., Gaur, G. K., Singh, U., Kumar, A., Kumar, S., and Mann, S. (2013). *DGAT1* polymorphism K232A in Sahiwal (Indian zebu) and frieswal (Holstein Friesian x Sahiwal crossbred) cattle. *Indian Journal of Animal Research*, *47*(4). 360-363

Gautier, M., Capitan, A., Fritz, S., Eggen, A., Boichard, D., and Druet, T. (2007). Characterization of the *DGAT1 K232A* and variable number of tandem repeat polymorphisms in French dairy cattle. *Journal of Dairy Science*, *90*(6), 2980–2988. https://doi.org/10.3168/jds.2006-707

Gebreyohannes, G., Koonawootrittriron, S., and Mauricio, A. (n.d.). Elzo and Thanathip Suwanasopee.(2013). Variance components and genetic parameters for milk production and lactation pattern in an Ethiopian multi-bred dairy cattle population. *Asian Australas Journal of Animal Science*, *26*(9), 1237–1246.

Geldermann, H. (1990). Genome analysis in domestic animals. VCH, weinheim, *Genetics*. 134: 943-951

Ghasemi, N., Zadehrahmani, M., and Rahimi, G. (2009). Associations between *prolactin* gene polymorphism and milk production in montebeliard cows. *International Journal of Genetics and Molecular Biology*, *1*(3), 48–51. http://www.academicjournals.org/IJGMB

Gholizadeh, M., Mianji, G.R., and Zadeh, H.S. (2008). Potential use of molecular markers in the genetic improvement of livestock. *Asian J. Anim. Vet Adv***.** 3(3)**:** 120-128

Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, *10*(6), 381–391. https://doi.org/10.1038/nrg2575

Gojam, Y., Tadesse, M., Efffa, K., and Hunde, D. (2016). Performance of crossbred dairy cows suitable for smallholder production systems at Holetta Agricultural Research Centre. *Ethiopian Journal of Agricultural Sciences*, *27*(1), 121–131.

Gothwal, A., Magotra, A., Bangar, Y. C., Malik, B. S., Yadav, A. S., and Garg, A. R. (2022). Candidate *K232A* mutation of *DGAT1* gene associated with production and reproduction traits in Indian Dairy cattle. *Animal Biotechnology*, (33), 1–9. https://doi.org/doi: 10.1080/10495398.2022.2109041

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M., and Snell, R. (2002). Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Research*, *12*(2), 222–231. https://doi.org/10.1101/gr.224202

Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J. J., Kvasz, A., Mni, M., Simon, P., Frère, J. M., Coppieters, W., and Georges, M. (2004). Genetic and functional confirmation of the causality of the *DGAT1* K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(8), 2398–2403. https://doi.org/10.1073/pnas.0308518100

Guo, S. W., and Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics*, *48*(2). https://doi.org/10.2307/2532296

Gurcan, E. K., Cobanoglu, O., Kul, E., Abaci, H. S., and Cankaya, S. (2019). Determination of the genetic polymorphism for *DGAT1* gene in Holstein, Jersey and native cattle breeds of Turkey. *Indian Journal of Animal Research*, *53*(2). https://doi.org/10.18805/ijar.B-873

Haile, A., Joshi, B. K., Ayalew, W., Tegegne, A., and Singh, A. (2011). Genetic evaluation of Ethiopian Boran cattle and their crosses with Holstein Friesian for growth performance in central Ethiopia. *Journal of Animal Breeding and Genetics*, *128*(2). https://doi.org/ 10.1111/j.1439-0388.2010.00882.x

Hale ML, Burg TM, Steeves TE .(2012). Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. PLoS ONE 7(9): e45170. doi:10.1371/journal.pone.0045170

Hallerman, E. ., Theilmann, J. L., Beckmann, J. ., Soller, M., and Womack, J. E. (1988). Mapping of bovine *prolactin* and rhodopsin genes in hybrid somatic cells. *Animal Genetics*, *19*(2). https://doi.org/10.1111/j.1365-2052.1988.tb00798.x

Hansen, C., Shrestha, J. N. B., Parker, R. J., Crow, G. H., and Derr, J. N. (2003). Genetic base and inbreeding of Canadienne, Brown Swiss, Holstein and Jersey cattle in Canada. *Animal Genetic Resources Information*, *33*. https://doi.org/10.1017/s1014233900001590

Hartl, D.L., and Clark, A.G. (1997). Principles of Population Genetics, 3rd edition. Sunderland USA, Sinauer Associates Inc. 542 p.

Haug, A., Hostmark, A. T., and Harstad, O. M. (2007). *Bovine milk in human nutrition – a review*. *16*, 1–16. https://doi.org/10.1186/1476-511X-6-25

Hayes, B. and Goddard, M.E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution.* 33: 209– 229

He, F., Sun, D., Yu, Y., Wang, Y., and Zhang, Y. (2006). Association between SNPs within *prolactin* gene and milk performance traits in Holstein Dairy Cattle. *Asian-Australasian Journal of Animal Sciences*, *19*(10). https://doi.org/10.5713/ajas.2006.1384

Hori-Oshima, S., and Barreras-Serrano, A. (2003). Relationships between *DGAT1* and *PIT-1* genes polymorphism and milk yield in Holstein cattle. *Proceedings-American Society of Animal Science Western Section, 54*, 129–131.

Hosseinpour Mashhadi, M., Nassiri, M. R., Mahmoudi, M., Rastin, M., Kashan, N. E. J., Vaez Torshizi, R., Tabasi, N., and Nooraee, S. E. (2012). Polymorphism and sequencing of *DGAT1* gene in Iranian Holstein bulls. *Iranian Journal of Applied Animal Science, 2*(1), 63–67.

Houaga, I., Muigai, A. W. T., Kyallo, M., Githae, D., Youssao, I. A. K., and Stomeo, F. (2017). Effect of breed and Diacylglycerol acyltransferase 1 gene polymorphism on milk production traits in Beninese White Fulani and Borgou cows. *Global Journal of Animal Breeding and Genetics , 5*(5), 403–412. http://www.globalscienceresearchjournals.org/

Houaga, I., Muigai, A. W. T., Ng'ang'a, F. M., Ibeagha-Awemu, E. M., Kyallo, M., Youssao, I. A. K., and Stomeo, F. (2018). Milk fatty acid variability and association with polymorphisms in *SCD1* and *DGAT1* genes in White Fulani and Borgou cattle breeds. *Molecular Biology Reports, 45*(6), 1849–1862.

Hradecká, E., Cítek, J., Panicke, L., Řehout, V., and Hanusová, L. (2008). The relation of GH1, GHR and *DGAT1* polymorphisms with estimated breeding values for milk production traits of German Holstein sires. *Czech Journal of Animal Science*, *53*(6), 238–245. https://doi.org/ 10.17221/362-cjas

Hu, X., Lü, A., Chen, H., Gao, X., Xu, H., Zhang, C., Fang, X., and Lei, C. (2009). Preliminary evidence for association of *prolactin* and prolactin receptor genes with milk production traits in chinese holsteins. *Journal of Applied Animal Research*, *36*(2), 213–217. https:// doi.org/10.1080/09712119.2009.9707062

Ishaq, R., Suleman, M., Naeem, M., Yousaf, M., Shah, A., and Ghafoor, A. (2012). *prolactin* gene polymorphism in Nili-Ravi buffaloes in relation to Sahiwal and Achai Cattle. *International Journal of Dairy Technology*, *65*(1), 1–5. https://doi.org/10.1111/j.1471-0307.2012.00875.x

Jiang, Z., Wang, H., Michal, J. J., Zhou, X., Liu, B., Woods, L. C. S., and Fuchs, R. A. (2016). Genome wide sampling sequencing for SNP genotyping: Methods, challenges and future development.*International Journal of Biological Sciences, 12*(1), 100–108. https:// doi.org/10.7150/ijbs.13498

Kalinowski, S. T. (2002). How many alleles per locus should be used to estimate genetic distances? *Heredity (Edinb)*, *88*(1), 62–65. https://doi.org/10.1038/sj.hdy.6800009

Kaminski, S., Ahman, A., Rusc, A., Wojcik, E., and Malewski, T. (2005). MilkProtChip - A microarray of SNPs in candidate genes associated with milk protein biosynthesis. Development and validation. *Journal of Applied Genetics*, *46*(1), 179–185.

Kaupe, B., Winter, A., Fries, R., and Erhardt, G. (2004). *DGAT1* polymorphism in *Bos indicus* and *Bos taurus* cattle breeds. *Journal of Dairy Research*, *71*(2), 182–187. https://doi.org /10.1017/S0022029904000032

Khaizaran, Z. A., and Al-razem, F. (2014). *Analysis of selected milk traits in Palestinian Holstein- Friesian cattle in relation to genetic polymorphism*. *8*(5), 74–85. https://doi.org /10.5897/JCAB2014.0409

Khatami, S. R., Lazebny, O. E., Maksimenko, V. F., and Sulimova, G. E. (2005). Association of DNA polymorphisms of the growth hormone and *prolactin* genes with milk productivity in Yaroslavl and Black-and-White cattle. *Russian Journal of Genetics*, *41*(2), 229–236. https://doi.org/10.1007/s11177-005-0040-x

Khatib, H., Huang, W., Mikheil, D., Schutzkus, V., and Monson, R. L. (2009). Effects of signal transducer and activator of transcription (STAT) genes STAT1 and STAT3 genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. *Journal of Dairy Science*, *92*(12). https://doi.org/10.3168/jds.2009-2439

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2). https://doi.org/10.1007/BF01731581

Komisarek, Jolanta, and Kolenda, M. (2016). The effect of *DGAT1* polymorphism on milk production traits in dairycows depending on environmental temperature. *Turkish Journal of Veterinary andAnimal Sciences*, *40*(2), 251–254.

Koopaei, H. K., Mohammad Abadi, M. R., Mahyari, S. A., Koshkoiyeh, A. E., Tarang, A. R., and Potki, P. (2012). Effect of *DGAT1* variants on milk composition traits in Iranian Holstein cattle population. *Animal Science Papers and Reports*, *30*(3).

Korwin-Kossakowska,A., Wyszyńska-Koko, J., andSender, G. (2006). A novel polymorphism of the porcine *prolactin* gene (*PRL*). *Neuroendocrinology Letters*, *27*(1–2).

Krovvidi, S., Sudhakar, K., Panneerselvam, S., Thiruvenkadan, A. K., Abraham, J., and Vinodkumar, G. (2013). Factors effecting milk composition of crossbred dairy cattle in Southern India. *International Journal of Food, Agriculture and Veterinary Sciences*, *3*(1),

229–232. https://doi.org/10.1007/s11250-021-02560-2

Krovvidi, S., Thiruvenkadan, A. K., Murali, N., Saravanan, R., Vinoo, R., and Metta, M. (2021). Evaluation of non-synonym mutation in *DGAT1 K232A* as a marker for milk production traits in Ongole cattle and Murrah buffalo from Southern India. *Tropical Animal Health and Production*, *53*(118), 1–7.

Kuhn, C., Thaller, G., Winter, A., Bininda-Emonds, O. R. P., Kaupe, B., Erhardt, G., Bennewitz, J., Schwerin, M., and Fries, R. (2004). Evidence for multiple alleles at the *DGAT1* locus better explains a quantitative trait locus with major effect on milk fat content in cattle. *Genetics*, *167*(4). https://doi.org/10.1534/genetics.103.022749

Laval G.,SanCristobal M., and Chevalet C. (2002). Measuring genetic distances between breeds: use of some distances in various shortterm evolution models. Genetics Selection Evolution, 34: 481–507.

Laurie, A. D., and G eorge, P. M. (2009). Evaluation of high-resolution melting analysis for screening the LDL receptor gene. *Clinical Biochemistry*, *42*(6). https://doi.org/10.1016/ j.clinbiochem.2008.11.015

Lemecha H., Mulatu W., Hussein I. et al. (2006). Response of four indigenous cattle breeds to natural tsetse and trypanosomosis challenge in the Ghibe Valley of Ethiopia. Veterinary Parasitology 141, 165–76.

Leskova, L., Bauer, M., Chrenek, P., Lacková, Z., Soročinová, J., Petrovič, V., and Kováč, G. (2013). Detection of *DGAT1* gene polymorphism and its effect on selected biochemical indicators in dairy cows after calving. *Acta Veterinaria Brno*, *82*(3). https://doi.org /10.2754/avb201382030265

Li, F., Cai, C., Qu, K., Liu, J., Jia, Y., Hanif, Q., Chen, N., Zhang, J., Chen, H., Huang, B., and Lei, C. (2020). *DGAT1* K232A polymorphism is associated with milk production traits in Chinese cattle. *Animal Biotechnology*.(32),1-5. https://doi.org/10.1080 /10495398.

Li, Y., Zhou, H., Cheng, L., Edwards, G. R., and Hickford, J. G. H. (2021). Effect of *DGAT1* variant (K232A) on milk traits and milk fat composition in outdoor pasture-grazed dairy cattle. *New Zealand Journal of Agricultural Research*, *64*(1). https://doi.org/10.1080/ 00288233.2019.1589537

Librado, P., and Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11). https://doi.org/10.1093/bioinformatics/ btp 187

Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, *21*(9), 2128–2129.

Mandefro, A., Sisay, T., Edea, Z., Uzzaman, M. R., Kim, K. S., andDadi, H. (2021). Genetic assessment of BoLA-DRB3 polymorphisms by comparing Bangladesh, Ethiopian, and Korean cattle. *Journal of Animal Science and Technology*, *63*(2). https://doi.org/ 10.5187/JAST.2021.E37

Manga, I., andŘiha, H. (2011). The *DGAT1* gene *K232A* mutation is associated with milk fat content, milk yield and milk somatic cell count in cattle (Short Communication). *Archives Animal Breeding*, *54*(3), 257–263. https://doi.org/10.5194/aab-54-257-2011

Marshall, T. C., Slate, J., Kruuk, L. E. B., and Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, *7*(5). https://doi.org/10.1046/j.1365-294x.1998.00374.x

Martínez-Arias, R., Calafell, F., Mateu, E., Comas, D., Andrés, A., and Bertranpetit, J. (2001). Sequence variability of a human pseudogene. *Genome Research*, *11*(6), 1071–1085.

Matukumalli, L. K., Grefenstette, J. J., Hyten, D. L., Choi, I. Y., Cregan, P. B., and Van Tassell, C. P. (2006). Application of machine learning in SNP discovery. *BMC Bioinformatics*, *7*. https://doi.org/10.1186/1471-2105-7-4

Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press. https://doi.org/ 10.4159/harvard.9780674865327

McDonald, J. H. (2009). *Handbook of biological statistics* (2nd ed.). sparky house publishing Baltimore, MD.

Mehmannavaz, Y., Amirinia, C., Bonyadi, M., andTorshizi, R. V. (2009). Effects of bovine *prolactin* gene polymorphism within exon 4 on milk related traits and genetic trends in Iranian Holstein bulls. *African Journal of Biotechnology*, *8*(19).

Mirkena T., Duguma G., Haile A. et al. (2010) Genetics of adaptation in domestic farm animals: a review. Livestock Science 132, 1–12

Mishra BP, Mukesh M, Prakash B, Monika S, Kapila R, Kishore A, et al. (2009). Status of milk protein, b-casein variants among Indian milch animals. *Indian J Anim Sci*; 79:72-5.

Mohammed, S. A., Rahamtalla, S. A., Ahmed, S. S., Elhafiz, A., Dousa, B. M., Elamin, K. M., and Ahmed, M. K. A. (2015). *DGAT1* gene in dairy cattle. *Global Journal of Animal Scientific Research*, *3*(1), 191–198.

Mpofu, N., Ojango, J. M. K., and Andersson-Eklund, L. (2011). Quantitative methods to improve the understanding and utilisation of animal genetic resources. *AGTR Version 2 Training Module*.

Myburgh, J., Osthoff, G., Hugo, A., de Wit, M., Nel, K., and Fourie, D. (2012). Comparison of the milk composition of free-ranging indigenous African cattle breeds. *South African Journal of Animal Sciences*, *42*(1),1-13. https://doi.org/10.4314/sajas.v42i1.1

Naqvi, A. N. (2007). Application of Molecular Genetic Technologies in Livestock Production: Potentials for Developing Countries. Advances in Biological Research, 1 (3-4): 72-84.

Nasiri, H., Forouzandeh, M., Rasaee, M. J., and Rahbarizadeh, F. (2005). Modified salting-out method: High-yield, high-quality genomic DNA extraction from whole blood using laundry detergent. *Journal of Clinical Laboratory Analysis*, *19*(6), 229–232. https://doi.org/10.1002/jcla.20083

Nayeri, S., and Stothard, P. (2016). Tissues, Metabolic Pathways and Genes of Key Importance in Lactating Dairy Cattle. *Springer Science Reviews*, *4*(2), 49–77. https://doi.org/10.1007/ s40362-016-0040-3

Nei M (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Science*. 70: 3321-3323.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, *89*(3). https://doi.org/10.1093/genetics/89.3.583

Nei, Masatoshi, and Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford University Press.

Niguse, A., and Aleme, A. (2015). Modeling the impact of climate change on production of Sesame in Western zone of Tigray, Northern Ethiopia. *Journal of Climatology and Weather Forecasting*, 1–7.

Nowacka-Woszuk, J., Noskowiak, A., Strabel, T., Jankowski, T., and Świtoński, M. (2008). An effect of the *DGAT1* gene polymorphism on breeding value of Polish Holstein-Friesian sires. *Animal Science Papers and Reports*, *26*(1).

O'Mahony, F. (1988). *Rural dairy technology: Experiences in Ethiopia* (No. 4; Dairy Technology Unit). ILRI.

OECD-FAO Agricultural Outlook 2022-2031. (2022). OECD. https://doi.org/10.1787/ f1b0b29c-en

Ogorevc, J., Kunej, T., Razpet, A., and Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics*, *40*(6). https://doi.org /10.1111/j.1365-2052.2009.01921.x

Oikonomou, G., Angelopoulou, K., Arsenos, G., Zygoyiannis, D., and Banos, G. (2009). The effects of polymorphisms in the *DGAT1*, leptin and growth hormone receptor gene loci on body energy, blood metabolic and reproductive traits of Holstein cows. *Animal Genetics*, *40*(1), 10–17.

Othman, O. E., Zayed, F. A., El Gawead, A. A., and El-Rahman, M. R. A. (2011). Genetic polymorphism of three genes associated with milk trait in Egyptian buffalo. *Journal of Genetic Engineering and Biotechnology*, *9*(2), 97–102. https://doi.org/10.1016/ j.jgeb. 2011.09.002

Ozdemir, M. (2020). A Prl/RsaI Polymorphism in Exon 3 and 4 of *prolactin* Gene in Dairy Cattle. *Pakistan Journal of Zoology*, *52*(1), 393–396. https://doi.org/https://dx.doi.org /10.17582/journal.pjz/2020.52.1.sc7

Pant, S. D., Schenkel, F. S., Verschoor, C. P., and Karrow, N. A. (2012). Use of breed-specific single nucleotide polymorphisms to discriminate between Holstein and Jersey dairy cattle breeds. *Animal Biotechnology*, *23*(1), 1–10.

Patel, J., and Chauhan, J. (2017). Evaluation of *DGAT1*-exon 8 K232A substitution in Gir and Kankrej (*Bos indicus*), Indian origin cattle and its association with milk production traits. *Genetika*, *49*(2), 627–634.

Pichler, F. B. (2002). *Genetic assessment of population boundaries and gene exchange in Hector's dolphin: Vol. DOC Scienc*. Department of Conservation, Wellington, New Zealand.

Pitt, D.,Sevane,N., Nicolazzi, E.L.,MacHugh, D.E.,Park,S.D.E., Colli,L.,Martinez, R.,Bruford, M.W. and Orozco-terWengel,P., 2019. Domestication of cattle: Two or three events? Evolutionary Applications, 12, 123–136

Rahmatalla, Reibmann, M., Mueller, U., and Brockmann, G. (2015). Identification of genetic variants influencing milk production traits in Sudanese dairy cattle. *Research Journal of Animal Sciences*, *9*(2), 12–22.

Ramkaran, A. G., Gera, S., Preeti, A. M., and Kumar, S. (2019). *Diacylglycerol O-acyltransferase 1* (*DGAT1*) polymorphism in Hardhenu cross bred cattle (Holstein Friesian x Sahiwal x Hariana breed). *Journal of Entomology and Zoology Studies*, *7*(1), 1170–1173.

Ramos-Onsins, S. E., and Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, *19*(12). https://doi.org/10.1093/ oxfordjournals.molbev.a004034

Rege, J. E. O., and Tawah, C. L. (1999). The state of African cattle genetic resources II. Geographical distribution, characteristics and uses of present-day breeds and strains. *Animal Genetic Resources Information*, *26*. https://doi.org/10.1017/s1014233900001152

Rincón, G., Armstrong, E., and Postiglioni, A. (2006). Analysis of the population structure of Uruguayan Creole cattle as inferred from milk major gene polymorphisms. *Genetics and Molecular Biology*, *29*(3). https://doi.org/10.1590/S1415-47572006000300016

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology (Clifton, N.J.),132*, 365–386. https://doi.org/ 10.1385/ 1-59259-192-2:365

Rychtarova, J., Sztankóová, Z., Kyselová, J., Zink, V., Štípková, M., Vacek, M., andŠtolc, L. (2014). Effect of *DGAT1*, *BTN1A1*, *OLR1*, and *STAT1* genes on milk production and reproduction traits in the czech fleckvieh breed. *Czech Journal of Animal Science*, *59*(2). https://doi.org/10.17221/7228-cjas

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4). https://doi.org/10.1093 /oxfordjournals.molbev.a040454

Sanders, K., Bennewitz, J., Reinsch, N., Thaller, G., Prinzenberg, E. M., Kühn, C., and Kalm, E. (2006). Characterization of the *DGAT1* mutations and the *CSN1S1* promoter in the German Angeln dairy cattle population. *Journal of Dairy Science*, *89*(8), 3164–3174. https://doi.org/10.3168/jds.S0022-0302(06)72590-5

Sasavage, N. L., Nilson, J. H., Horowitz, S., and Rottman, F. M. (1982). Nucleotide sequence of bovine *prolactin* messenger RNA. Evidence for sequence polymorphism. *Journal of Biological Chemistry*, *257*(2), 1–10. https://doi.org/10.1016/s0021-9258(19)68247-5

Schoen, D. J., and Brown, A. H. D. (1991). Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(10). https://doi.org/10.1073/pnas.88.10.4494

Shah, R. M., Ganai, N. A., Sheikh, F. D., Shanaz, S., Khan, H. M., Alam, S., Khan, N. N., Sheikh, T., Bukhari, S., Hamadani, A., and Rather, M. A. (2021). Exon IV *prolactin* (*PRL*) gene polymorphism and its association with milk production traits in dairy cattle

of Kashmir , India. *Journal of Entomology and Zoology Studies*, *9*(2), 521–524.

Shete, S., Tiwari, H., and Elston, R. C. (2000). On estimating the heterozygosity and polymorphism information content value. *Theoretical Population Biology*, *57*(3). https://doi.org/10.1006/tpbi.2000.1452

Shorten, P. R., Pleasants, T. B., and Upreti, G. C. (2004). A mathematical model for mammary fatty acid synthesis and triglyceride assembly: The role of stearoyl CoA desaturase (SCD). *Journal of Dairy Research*, *71*(4). https://doi.org/10.1017/S0022029904000354

Singh, M., and Ludri, R. S. (1999). Plasma *prolactin*, Blood Metabolites and Yield and Composition of Milk during Early Lactation in Goats Following Administration of Bromocryptine. *Asian-Australasian Journal of Animal Sciences*, *12*(4), 585–589. https://doi.org/10.5713/ajas.1999.585

Sinha, Y. N. (1995). Structural variants of *prolactin*: occurrence and physiological significance. *Endocrine Reviews*, *16*(3), 354–369.

Smith, S. J., Cases, S., Jensen, D. R., Chen, H. C., Sande, E., Tow, B., Sanan, D. A., Raber, J., Eckel, R. H., and Farese, R. V. (2000). Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking *DGAT*. *Nature Genetics*, *25*(1). https://doi.org /10.1038/75651

Sonmez, Z., and Ozdemir, M. (2017). *prolactin*-RsaI gene polymorphism in East Anatolian Red cattle in Turkey. *South African Journal of Animal Science*, *47*(2), 124–129.

Spelman, R. J., Ford, C. A., McElhinney, P., Gregory, G. C., and Snell, R. G. (2002). Characterization of the *DGAT1* gene in the New Zealand dairy population. *Journal of Dairy Science*, *85*(12), 3514–3517. https://doi.org/10.3168/jds.S0022-0302(02)74440-8

Spiess, E. B. (1977). *Genes in populations* (Second). A Wiley-lnterscience.

Strzałkowska, N., Siadkowska, E., Słoniewski, K., Krzyżewski, J., and Zwierzchowski, L. (2005). Effect of the *DGAT1* gene polymorphism on milk production traits in Black-and-White (Friesian) cows*. *Animal Science Papers and Reports*, *23*(3).

Suchocki, T., Komisarek, J., and Szyda, J. (2010). Testing candidate gene effects on milk production traits in dairy cattle under various parameterizations and modes of inheritance. *Journal of Dairy Science*, *93*(6). https://doi.org/10.3168/jds.2009-2550

Sumantri, C., and Jakaria, J. (2015). The Genetic Variability of *prolactin* and Signal Transducers and Activators of Transcription 5A (STAT5A) Genes in Bali Cattle. *Media Peternakan*, *38*(1), 1–11. https://doi.org/10.5398/medpet.2015.38.1.1

Szyda, J., and Komisarek, J. (2007). Statistical modeling of candidate gene effects on milk production traits in dairy cattle. *Journal of Dairy Science*, *90*(6). https://doi.org /10.3168/jds. 2006-724

Szyda, J., Komisarek, J., and Antkowiak, I. (2014). Modelling effects of candidate genes on complex traits as variables over time. *Animal Genetics*, *45*(3). https://doi.org /10.1111/age.12144

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3). https://doi.org/10.1093/genetics/123.3.585

Takahashi, K., and Nei, M. (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution*, *17*(8). https://doi.org/10.1093/oxfordjournals.molbev.a026408

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3). https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura, Koichiro, Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Teweldemedhn, M., and Selam, M. (2020). Characterization of Begait cattle using morphometric and qualitative traits in Western Zone of Tigray, Ethiopia. *International Journal of Livestock Production*, *11*(1). https://doi.org/10.5897/ijlp2019.0637

Thaller, G., Kramer, W., Winter, A., Kaupe, B., Erhardt, G., and Fries, R. (2003). Effects of *DGAT1* variants on milk production traits in German cattle breeds. *Journal of Animal Science*, *81*(8), 1911–1918. https://doi.org/10.2527/2003.8181911x

Thuy, N. T. D., Thu, N. T., and Cuong, N. H. (2018). Polymorphism of *PIT-1* and *prolactin* Genes and Their Effects on Milk Yield in Holstein Frisian Dairy Cows Bred in Vietnam. *Animal Genetics*, *54*(3), 346–352. https://doi.org/10.1134/S1022795418030146

Valenti, B., Criscione, A., Moltisanti, V., Bordonaro, S., De Angelis, A., Marletta, D., Di Paola, F., and Avondo, M. (2019). Genetic polymorphisms at candidate genes affecting fat content and fatty acid composition in Modicana cows: Effects on milk production traits in different feeding systems. *Animal*, *13*(6), 1332–1340.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915). https://doi.org/10.1038/nature01262

Weir, B. S., and Basten, C. J. (1990). Sampling strategies for distances between DNA sequences *Biometrics*, *46*(3).

Weir, B. S., and Cockerham,C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6).https://doi.org/10.1111/j.1558-5646.1984.tb0 5657. x

Weiss, S. B., and Kennedy, E. P. (1956). The Enzymatic Synthesis of Triglycerides. *Journal of the American Chemical Society*, *78*(14). https://doi.org/10.1021/ja01595a088

Weller, J. I. (2009). *Quantitative trait loci analysis in animals*. CABI.

Wheeler, D., and Baghwat, M. (2007). BLAST QuickStart: example-driven web-based BLAST tutorial. In N.H.Bergman (Ed.), *Methods in Molecular Biology™* (Comparativ, (395)149–176).Humana Press. https://doi.org/https://doi.org/10.1007/978-1-59745-514-5_9

Widmer, A., and Lexer, C. (2001). Glacial refugia: Sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology and Evolution*, *16*(6), 267–269. https://doi.org/10.1016/S0169-5347(01)02163-2

Winter, A., Krämer, W., Werner, F. A. O., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J. E., Thaller, G., and Fries, R. (2002). Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:Diacylglycerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk fat content. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14). https://doi.org/10.1073/pnas.142293799

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*(4). https://doi.org/10.2307/2407273

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., and Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7). https://doi.org/10.1038/ng.608

Yaro, M., Munyard, K. A., Stear, M. J., and Groth, D. M. (2017). Molecular identification of livestock breeds: A tool for modern conservation biology. *Biological Reviews*, *92*(2). https://doi.org/10.1111/brv.12265

Yen, C.-L. E., Stone, S. J., Koliwad, S., Harris, C., and Farese, R. V. (2008). Thematic Review Series: Glycerolipids. *DGAT* enzymes and triacylglycerol biosynthesis. *Journal of Lipid Research*, *49*(11), 489-501. https://doi.org/10.1194/jlr.r800018-jlr200

Zerabruk, M., Li, M. H., Kantanen, J., Olsaker, I., Ibeagha-Awemu, E. M., Erhardt, G., and Vangen, O. (2012). Genetic diversity and admixture of indigenous cattle from North Ethiopia: Implications of historical introgressions in the gateway region to Africa. *Animal Genetics*, *43*(3). https://doi.org/10.1111/j.1365-2052.2011.02245.x

Zerabruk, M., Vangen, O., and Haile, M. (2007). The status of cattle genetic resources in North Ethiopia: On-farm characterization of six major cattle breeds. *Animal Genetic Resources Information*, (*40*),15-32. https://doi.org/10.1017/s1014233900002169

Zhou, X., Rao, N. P., Cole, S. W., Mok, S. C., Chen, Z., and Wong, D. T. (2005). Progress in concurrent analysis of loss of heterozygosity and comparative genomic hybridization utilizing high density single nucleotide polymorphism arrays. *Cancer Genetics and Cytogenetics*, *159*(1). https://doi.org/10.1016/j.cancergencyto.2004.09.014

Zhu, B., Zhang, J. jing, Niu, H., Guan, L., Guo, P., Xu, L. yang, Chen, Y., Zhang, L. pei, Gao, H. jiang, Gao, X., and LI, J. ya. (2017). Effects of marker density and minor allele frequency on genomic prediction for growth traits in Chinese Simmental beef cattle. *Journal of Integrative Agriculture*, *16*(4).https://doi.org/10.1016/S2095-3119(16)61474-0

8. APPENDICES

Supplementary Table 1.Haplotypes frequencies of *DGAT1* gene

Haplotype:	Boran * HF cross(23)	Boran(17)	Begait(16)	Fogera(17)	Horro(16)
Hap_1	0.087	-	0.438	0.0588	0.0625
Hap_2	0.478	0.706	0.375	0.706	0.75
Hap_3	0.0435	-	-	-	-
Hap_4	0.391	0.118	-	-	0.0625
Hap_5	-	0.0588	-	0.0588	-
Hap_6	-	0.0588	0.0625	0.0588	0.0625
Hap_7	-	0.0588	-	-	-
Hap_8	-	-	0.0625	-	-
Hap_9	-	-	0.0625	-	0.0625
Hap_10	-	-	-	0.0588	-
Hap_11	-	-	-	0.0588	-

Supplementary Table 2. Evolutionary divergence among breed/species for *DGAT1* gene

Breed/Species	BTHA	BTHM	BBM	BBK	CDAM	CDKM	CHM	CHL	OAE	BHK	BHA	BRK	BRA	BGK	BGA	FGK	FGA	HRK
*Bos_taurus*_Holstein_ AJ318490(BTHA)																		
*Bos_taurus*_Holstein_ MF069174(BTHM)	0.0147																	
*Bubalus Bubalis*_ MF069172(BBM)	0.0098	0.0247																
Bubalus Bubalis KX965992(BBK)	0.0098	0.0147	0.0197															
*Camelus dromedarius*_ Allele A _ MF069170(CDAM)	0.0147	0.0297	0.0049	0.0247														
*Camelus dromedarius*_Allele K_MF069171(CDKM)	0.0197	0.0348	0.0098	0.0297														
*Capra hircus*_AlleleK_ MF069173(CHM)	0.0279	0.0336	0.0394	0.0166	0.0451	0.051												
*Capra hircus*_Allele_K_ LT221856(CHL)	0.0098	0.0247	-	0.0197	0.0049	0.0098	0.0394											
*Ovis aries*_EU178818 (OAE)	0.0222	0.0279	0.0336	0.011	0.0393	0.0451	0.0055	0.0336										
Boran * HF cross_K(BHK)	0.0049	0.0098	0.0148	0.0049	0.0198	0.0248	0.0222	0.0148	0.0166									
Boran * HF cross _A(BHA)	0.0147	-	0.0247	0.0147	0.0297	0.0348	0.0336	0.0247	0.0279	0.0098								
Boran_Allele_K(BRK)	0.0049	0.0098	0.0147	0.0049	0.0197	0.0247	0.0222	0.0147	0.0166	-	0.0098							
Boran_Allele_A(BRA)	0.0148	-	0.0248	0.0148	0.0299	0.035	0.0336	0.0248	0.0279	0.0098	-	0.0098						
Begait_Allele_K(BGK)	0.0298	0.0348	0.04	0.0297	0.0451	0.0503	0.051	0.04	0.0451	0.0248	0.0348	0.0247	0.035					
Begait_Allele_A(BGA)	0.0348	0.0197	0.0452	0.0348	0.0503	0.0556	0.0569	0.0452	0.051	0.0299	0.0197	0.0298	0.0198	0.0147				
Fogera_Allele_K(FGK)	0.0049	0.0098	0.0147	0.0049	0.0197	0.0247	0.0222	0.0147	0.0166	-	0.0098	-	0.0098	0.0247	0.0298			
Fogera_Allele_A(FGA)	0.04	0.0248	0.0504	0.0399	0.0556	0.0608	0.0629	0.0504	0.0569	0.035	0.0248	0.0348	0.0249	0.0197	0.0049	0.0348		
Horro_Allele_K(HRK)	0.0049	0.0098	0.0147	0.0049	0.0197	0.0247	0.0222	0.0147	0.0166	-	0.0098	-	0.0098	0.0247	0.0298	-	0.0348	
Horro_Allele_A(HRA)	0.0147	-	0.0247	0.0147	0.0297	0.0348	0.0336	0.0247	0.0279	0.0098	-	0.0098	-	0.0348	0.0197	0.0098	0.0248	0.0098

Mean genetic distance=0.02±0.01

Supplementary Table 3. Evolutionary divergence among breed/species for *PRL* gene

Breed/Species	BTAF	BTAH	BTJF	BBEU	CHEU	OAKC	BHA	BHG	BGA	BGG	BRA	BRG	FGA	FGG	HRA	HRG
*Bos_taurus*_AF426315(BTAF)																
*Bos_taurus*_AH013356(BTAH)	-															
*Bos_taurus*_JF826522(BTJF)	0.0178	0.0178														
Bubalus_Bubalis EU340420(BBEU)	0.1216	0.1216	0.0118													
*Capra_hircus*_EU256165(CHEU)	0.0267	0.0267	0.0238	0.0552												
*Ovis_aries*_KC764410(OAKC)	0.0398	0.0398	0.0178	0.1102	0.0213											
Boran Holstein_Allele_A(BHA)	0.0097	0.0097	0.0118	0.1157	0.0267	0.0296										
Boran Holstein_Allele_G(BHG)	0.0146	0.0146	0.0059	0.1099	0.0213	0.0245	0.0048									
Begait_Allele_A(BGA)	0.0146	0.0146	0.0118	0.1099	0.0267	0.0245	0.0048	0.0097								
Begait_Allele_G(BGG)	0.0195	0.0195	-	0.1155	0.0266	0.0295	0.0097	0.0048	0.0146							
Boran_Allele_A(BRA)	0.0146	0.0146	0.0118	0.1099	0.0267	0.0245	0.0048	0.0097	-	0.0146						
Boran_Allele_G(BRG)	0.0195	0.0195	-	0.1155	0.0266	0.0295	0.0097	0.0048	0.0146	-	0.0146					
Fogera_Allele_A(FGA)	0.0146	0.0146	0.0118	0.1099	0.0267	0.0245	0.0048	0.0097	-	0.0146	-	0.0146				
Fogera_Allele_G(FGG)	0.0245	0.0245	-	0.1212	0.0321	0.0345	0.0146	0.0097	0.0195	0.0048	0.0195	0.0048	0.0195			
Horro_Allele_A(HRA)	0.0146	0.0146	0.0118	0.1099	0.0267	0.0245	0.0048	0.0097	-	0.0146	-	0.0146	-	0.0195		
Horro_Allele_G(HRG)	0.0196	0.0196	0.0059	0.1042	0.0213	0.0195	0.0096	0.0048	0.0048	0.0097	0.0048	0.0097	0.0048	0.0146	0.0048	

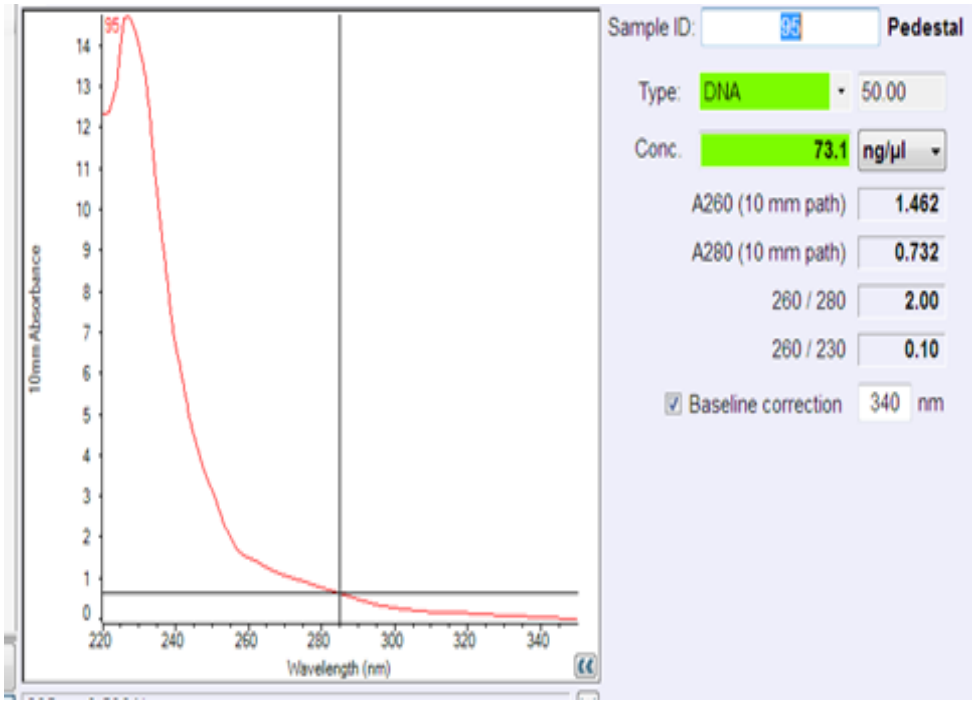
Mean genetic distance=0.03±0.01

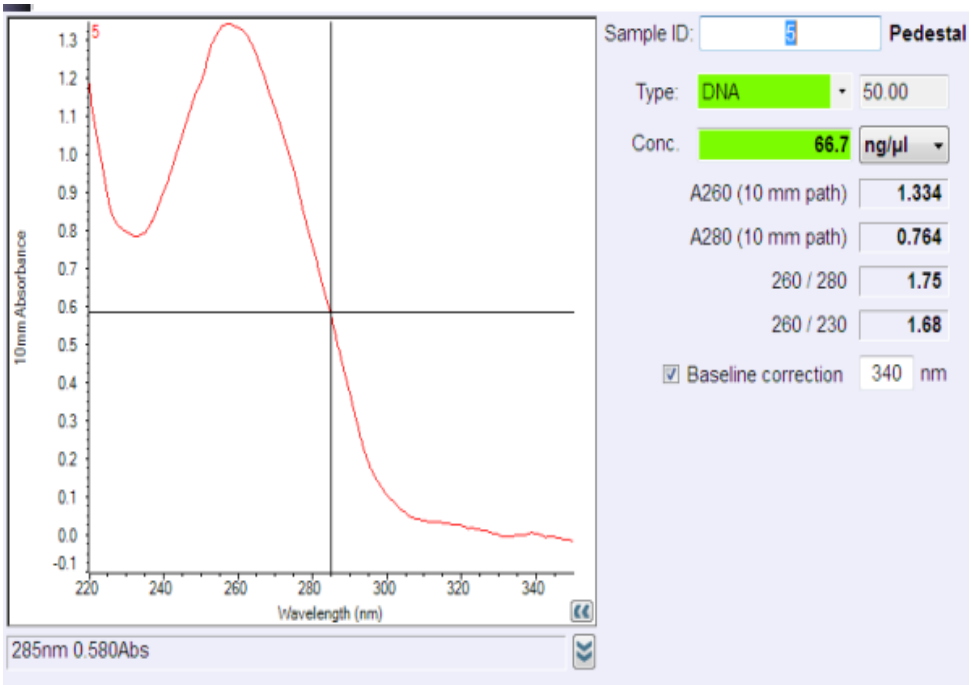
Supplementary Table 4. Homology analysis of studied *DGAT1* base on BLAST

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Bos taurus dgat gene for Acyl-CoA:1,2-diacylglycerol O-transferase and partial hsf1 gene for Heat Shock Factor Protein 1	*Bos taurus*	353	353	100%	5.00E-93	97.58	14117	AJ318490.1
Bos indicus breed Murrah *DGAT1* gene, exons 7 through 9 and partial cds	*Bos indicus*	353	353	100%	5.00E-93	97.58	321	DQ435290.1
Bos indicus breed Karan Fries *DGAT1* gene, exons 7 through 9 and partial cds	*Bos indicus*	353	353	100%	5.00E-93	97.58	321	DQ435286.1
Bos indicus x *Bos taurus* isolate HFX6 *diacylglycerol O-acyltransferase 1* (*DGAT1*) gene, partial cds	*Bos indicus x Bos taurus*	342	342	100%	1.00E-89	96.62	411	KX965994.1
Capra hircus *dgat1* gene for *diacylglycerol O-acyltransferase 1*	*Capra hircus*	342	342	100%	1.00E-89	96.62	37251	LT221856.1
Ovis aries diacylglycerol O-acyltransferase (*DGAT1*) gene, exons 8, 9 and partial cds	*Ovis aries*	342	342	100%	1.00E-89	96.62	285	FJ415875.1
Bos indicus voucher S6 breed Sokoto Gudali *diacylglycerol O-acyltransferase 1* (*DGAT1*) gene, partial cds	*Bos indicus*	340	340	100%	4.00E-89	96.14	370	MW999686.1
Bos indicus voucher S13 breed Sokoto Gudali *diacylglycerol O-acyltransferase 1* (*DGAT1*) gene, partial cds	*Bos indicus*	339	339	100%	1.00E-88	96.12	366	MW999684.1
Camelus dromedarius breed Bikaneri diacylglycerol O-acyltransferase (DGAT) gene, DGAT-B allele, partial cds	*Camelus dromedaries*	337	337	100%	5.00E-88	96.14	411	MF069171.1

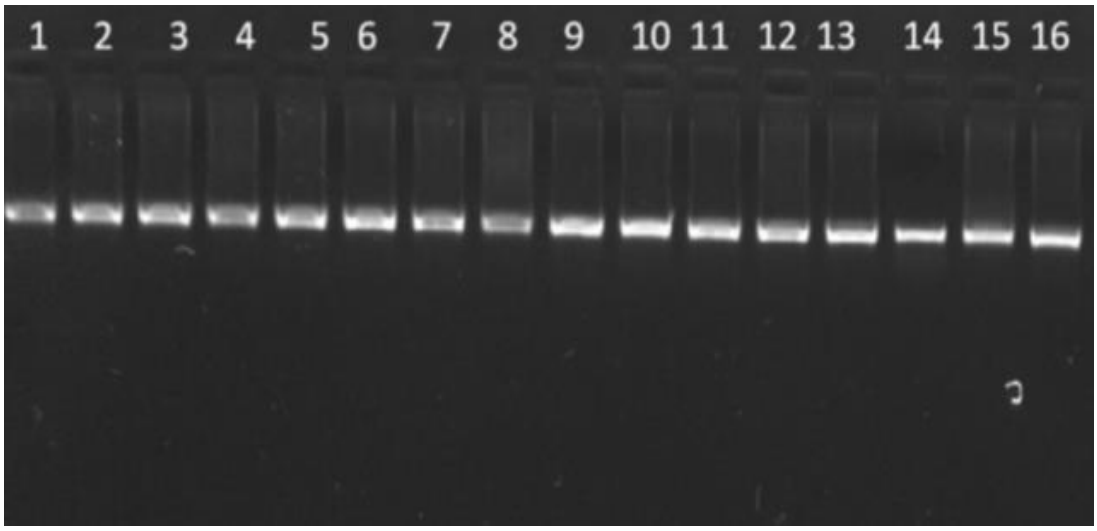
Supplementary Table 5. Homology analysis of studied *PRL* base on BLAST

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Bos taurus prolactin precursor gene, complete cds	*Bos taurus*	366	366	100%	6.00E-97	98.1	9388	AF426315.1
Bos indicus breed Sahiwal pop variant Rsa A *PRL* gene, exon-4 and partial cds	*Bos indicus*	372	372	100%	1.00E-98	98.57	294	KY777612.1
Capra hircus haplotype C *PRL* gene, exon-4 and partial cds	Capra hircus	324	324	91%	4.00E-84	96.89	461	EU256167.1
Bubalus bubalis *PRL* mRNA, partial cds	Bubalus bubalis	322	322	87%	1.00E-83	98.36	636	EU340420.1
Bos taurus PRL, Mrna	*Bos taurus*	322	322	87%	1.00E-83	98.36	907	NM173953.2
Bos indicus breed Gaolao Cattle *PRL* gene, *PRL*-T allele, exon-4 and partial cds	*Bos indicus*	346	346	91%	8.00E-91	98.96	220	MH791171.1
Bos indicus breed Gaolao Cattle *PRL* gene, *PRL*-A allele, exon-4 and partial cds	*Bos indicus*	339	339	94%	1.00E-88	97.47	198	MH791172.1
Ovis aries HMJ4 *PRL* gene, partial sequence	Ovis aries	217	217	58%	7.00E-52	98.37	261	LC718488.1
Bubalus bubalis haplotype A *PRL* gene, exon-4 and partial cds	Bubalus bubalis	274	274	73%	4.00E-69	98.7	154	MZ302991.1

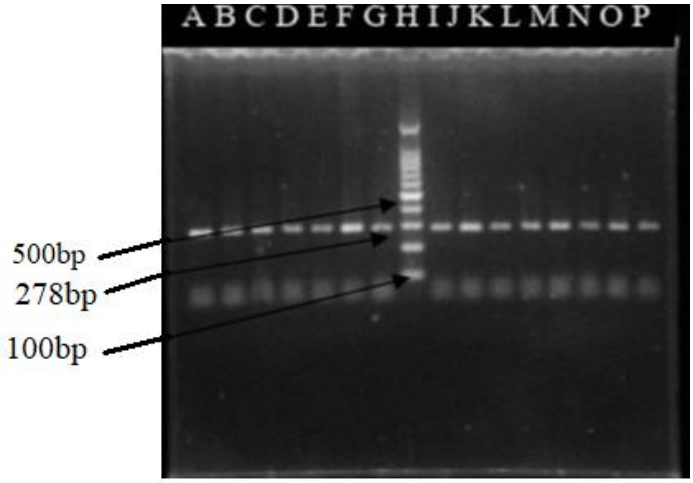




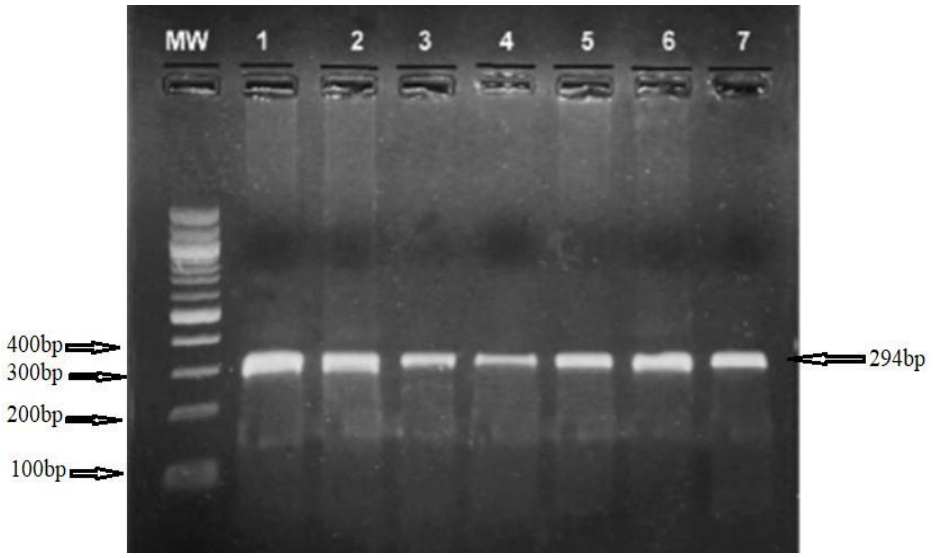
Supplementary Figure 1. Genomic DNA quality and concentration check of studied breeds using NANODROP spectrophotometry



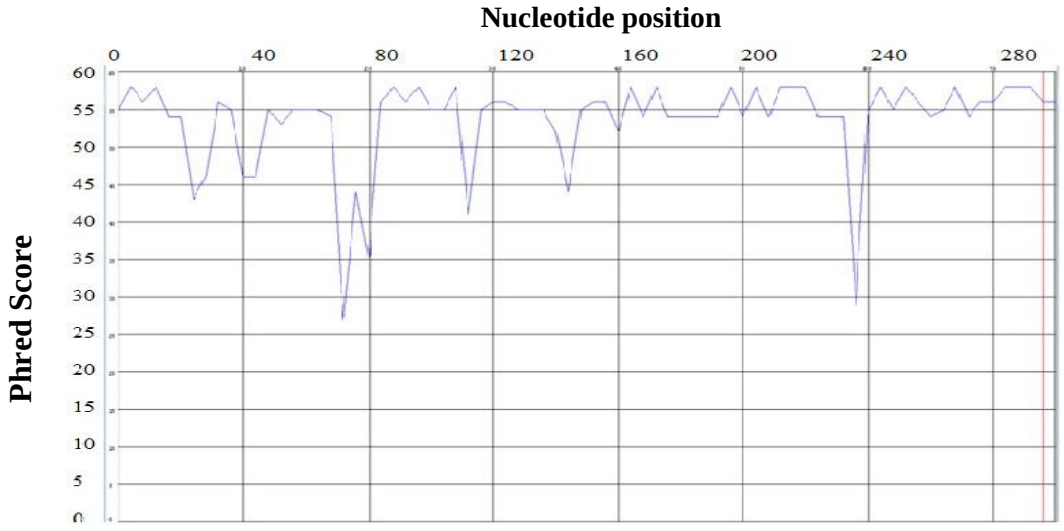
Supplementary Figure 2. Lane 1-16 Gel electrophoresis results of genomic DNA of sampled populations



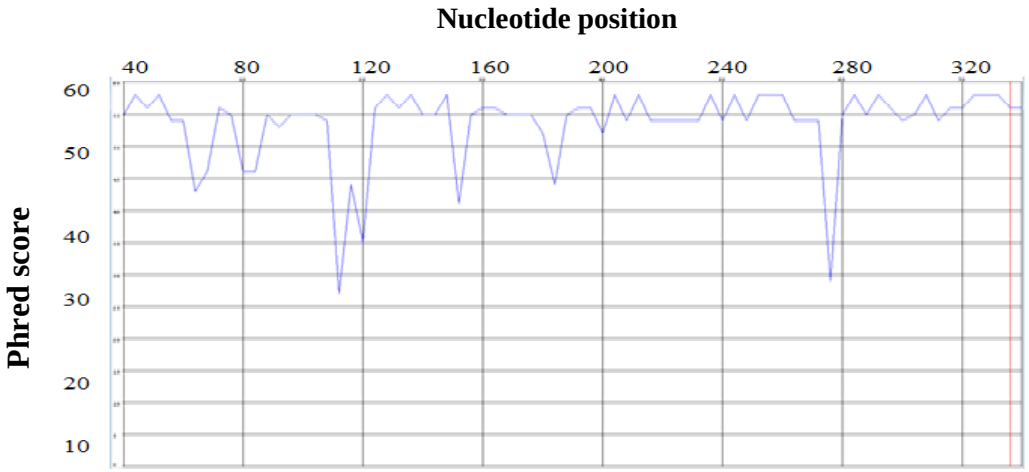
Supplementary Figure 3. Lane A-G &I-P PCR amplicons of 278bp of *DGAT1*.H: is 100bp DNA ladder



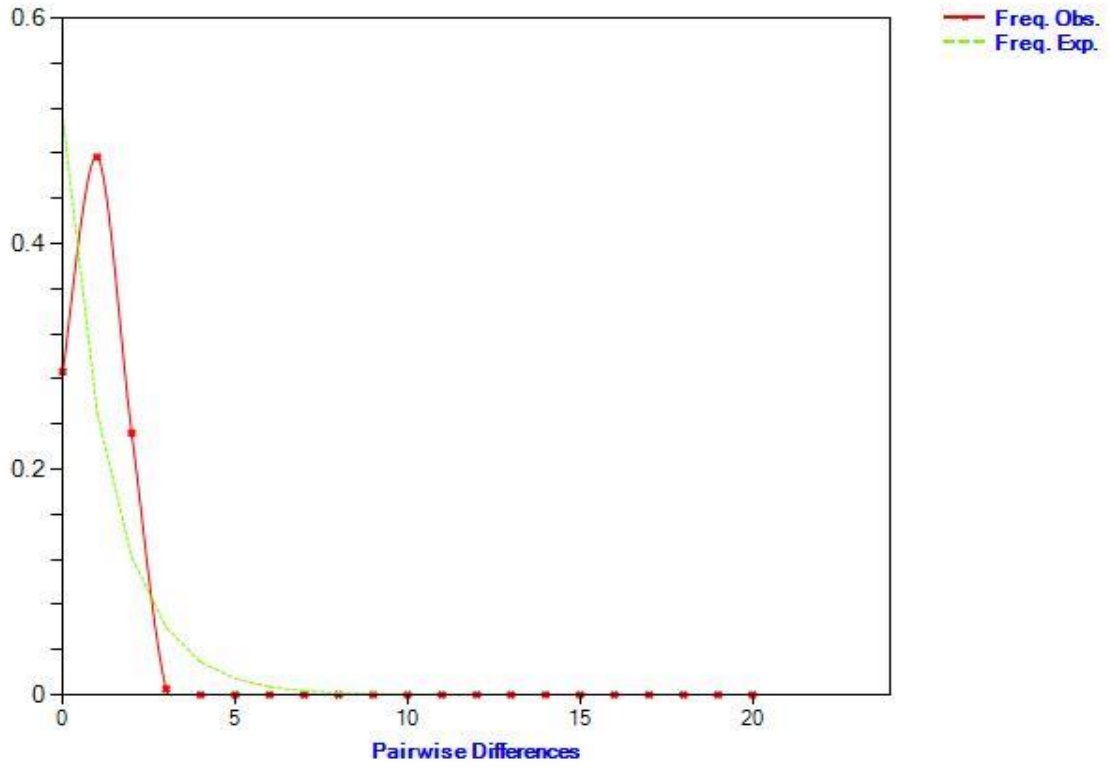
Supplementary Figure 4. Lane 1-7 PCR amplicons of 294bp of *PRL*.MW: is 100bp ladder



Supplementary Figure 5. Quality scores of Boran cattle *DGAT1* gene contig



Supplementary Figure 6. Quality scores of Boran cattle *PRL* gene contig



Supplementary Figure 7. Trend of observed and expected haplotype heterozygosities