



OPEN ACCESS

EDITED BY

Ntanganedzeni Mapholi,
University of South Africa, South Africa

REVIEWED BY

Guangxin E.,
Southwest University, China
Paolo Ajmone Marsan,
Catholic University of the Sacred Heart,
Italy

*CORRESPONDENCE

Tsadkan Zegeye,
✉ tsadkan96@gmail.com
Olivier Hanotte,
✉ o.hanotte@cgiar.org

RECEIVED 21 September 2022

ACCEPTED 14 July 2023

PUBLISHED 28 July 2023

CITATION

Zegeye T, Belay G, Vallejo-Trujillo A,
Han J and Hanotte O (2023), Genome-
wide diversity and admixture of five
indigenous cattle populations from the
Tigray region of northern Ethiopia.
Front. Genet. 14:1050365.
doi: 10.3389/fgene.2023.1050365

COPYRIGHT

© 2023 Zegeye, Belay, Vallejo-Trujillo,
Han and Hanotte. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Genome-wide diversity and admixture of five indigenous cattle populations from the Tigray region of northern Ethiopia

Tsadkan Zegeye^{1,2,3*}, Gurja Belay², Adriana Vallejo-Trujillo⁴,
Jianlin Han^{3,5} and Olivier Hanotte^{3,4,6*}

¹Mekelle Agricultural Research Center, Tigray Agricultural Research Institute, Mekelle, Ethiopia, ²Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa, Ethiopia, ³Live Gene—CTLGH, International Livestock Research Institute (ILRI), Addis Ababa, Ethiopia, ⁴Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, ⁵CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China, ⁶Cells, Organism and Molecular Genetics, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

The Tigray region, where we found around eight per cent of the indigenous cattle population of Ethiopia, is considered as the historic centre of the country, with the ancient pre-Aksumite and Aksumite civilisations in contact with the civilisations of the Fertile Crescent and the Indian subcontinent. Here, we used whole genome sequencing data to characterise the genomic diversity, relatedness, and admixture of five cattle populations (Abergelle, Arado, Begait, Erob, and Raya) indigenous to the Tigray region of Ethiopia. We detected 28 to 29 million SNPs and 2.7 to 2.9 million indels in each population, of which 7% of SNPs and 34% of indels were novel. Functional annotation of the variants showed around 0.01% SNPs and 0.22%–0.27% indels in coding regions. Enrichment analysis of genes overlapping missense private SNPs revealed 20 significant GO terms and KEGG pathways that were shared by or specific to breeds. They included important genes associated with morphology (*SCN4A*, *TAS1R2* and *KCNG4*), milk yield (*GABRG1*), meat quality (*MMRN2*, *VWC2*), feed efficiency (*PCDH8* and *SLC26A3*), immune response (*LAMC1*, *PCDH18*, *CELSR1*, *TLR6* and *ITGA5*), heat resistance (*NPFFR1* and *HTR7*) and genes belonging to the olfactory gene family, which may be related to adaptation to harsh environments. Tigray indigenous cattle are very diverse. Their genome-wide average nucleotide diversity ranged from 0.0035 to 0.0036. The number of heterozygous SNPs was about 0.6–0.7 times higher than homozygous ones. The within-breed average number of ROHs ranged from 777.82 to 1000.45, with the average sum of the length of ROHs ranging from 122.01 Mbp to 163.88 Mbp. The genomic inbreeding coefficients differed among animals and breeds, reaching up to 10% in some Begait and Raya animals. Tigray indigenous cattle shared a common ancestry with Asian indicine (85.6%–88.7%) and African taurine (11.3%–14.1%) cattle, with very small, if any, European taurine introgression. This study identified high within-breed genetic diversity representing an opportunity for breeding improvement programs and, also, significant novel variants that could increase the number of known cattle variants, an important contribution to the knowledge of domestic cattle genetic diversity.

KEYWORDS

Abergelle, Arado, Begait, Erob, private missense variants, Raya, runs of homozygosity, Tigray cattle

Introduction

From the main domestication centres, cattle dispersed worldwide through trading and human migration routes (Hanotte et al., 2002; Beja-Pereira et al., 2006; Freeman et al., 2006; Ajmone-Marsan et al., 2010). The long process of their origin, domestication, and natural and artificial selection led to diversified phenotypic attributes related to their history, agro-ecologies and production systems (Ajmone-Marsan et al., 2010; Purfield et al., 2012; FAO, 2015). About 1019 local cattle breeds have been recognised worldwide (FAO, 2015). However, the growing demand for animal-based food products is resulting in the extensive introduction of a few specialised high-yielding milk and beef breeds (Ajmone-Marsan et al., 2010; FAO, 2015; Kukučková et al., 2017) with crossbreeding or replacement of the indigenous genotypes. It may trigger a sharp decline in the population size of local breeds (Medugorac et al., 2009) and erosion of their genetic makeup. It has been estimated that around 50% of the global cattle breeds' diversity remains unknown (FAO, 2015). Characterising the diversity of indigenous breeds is important for understanding their adaptive traits and for targeted conservation strategies (FAO, 2015; Mwai et al., 2015; Addo et al., 2019; Eusebi et al., 2020).

Ethiopia is a major entry gate for cattle into the African continent (Hanotte et al., 2002; Li et al., 2007; Edea et al., 2015). It is the home of Africa's largest cattle population and ranks the fifth worldwide (Mwai et al., 2015; CSA, 2018). It has 60.39 million heads of cattle, of which 98.24% are indigenous to the country (CSA, 2018) and managed by smallholder farmers (Rowlands et al., 2006; EBI, 2016). The indigenous cattle of Ethiopia produce, reproduce, and survive with little veterinarian intervention and limited feed resources, including in extreme temperatures (hot or cold) and diverse agro-ecologies ranging from low altitude (<500 m above sea level (m.a.s.l.) to high altitude mountainous areas (>3000 m.a.s.l.) (EBI, 2016; Bekuma and Hirpha, 2018).

The region of Tigray in the North of Ethiopia is an ancient centre of civilisations (e.g., ancient pre-Aksumite Kingdom of Da'amat and Aksumite Kingdom of Axum) which were in trading contacts with the ancient civilisations of the Fertile Crescent and the Indus Valley (Finneran, 2007; Pagani et al., 2012). Accordingly, it had an important role in the introduction of livestock into the Horn of Africa (Woldekiros and D'Andrea, 2017).

Tigray is the fourth most cattle-populated Ethiopian region, with about 8% of the country's cattle genetic resource (CSA, 2018). Previous studies have characterised some of these populations using low-density molecular markers such as microsatellites, Y-chromosome markers or SNPs arrays (Li et al., 2007; Zerabruk et al., 2007; Dadi et al., 2008; Zerabruk et al., 2011; Edea et al., 2015). Using five Y chromosome markers, Li et al. (2007) identified indicine but no taurine Y chromosome in the Tigray cattle with the exception of an Arado bull. Dadi et al. (2008) characterised the genetic diversity of Raya (Tigray cattle) and other cattle from different parts of Ethiopia using 30 microsatellite loci. Zerabruk et al. (2007) reported the genetic diversity of the five recognised Tigray cattle populations (Abergelle, Arado, Begait, Erob and Raya) using 20 autosomal microsatellite markers and observed that the Begait cattle had the highest within-population diversity among the examined ones. Using the same set of 20 autosomal microsatellite

markers, Zerabruk et al. (2011) characterised the admixture composition of the Tigray cattle and reported a small proportion of European taurine background in some animals. Edea et al. (2015) genotyped three Tigray cattle populations (Arado, Begait and Raya) and four other Ethiopian cattle populations using the GeneSeek Genomic Profiler HD Bead Chip SNP array and found high genetic differentiation and unique admixture patterns in the Begait cattle.

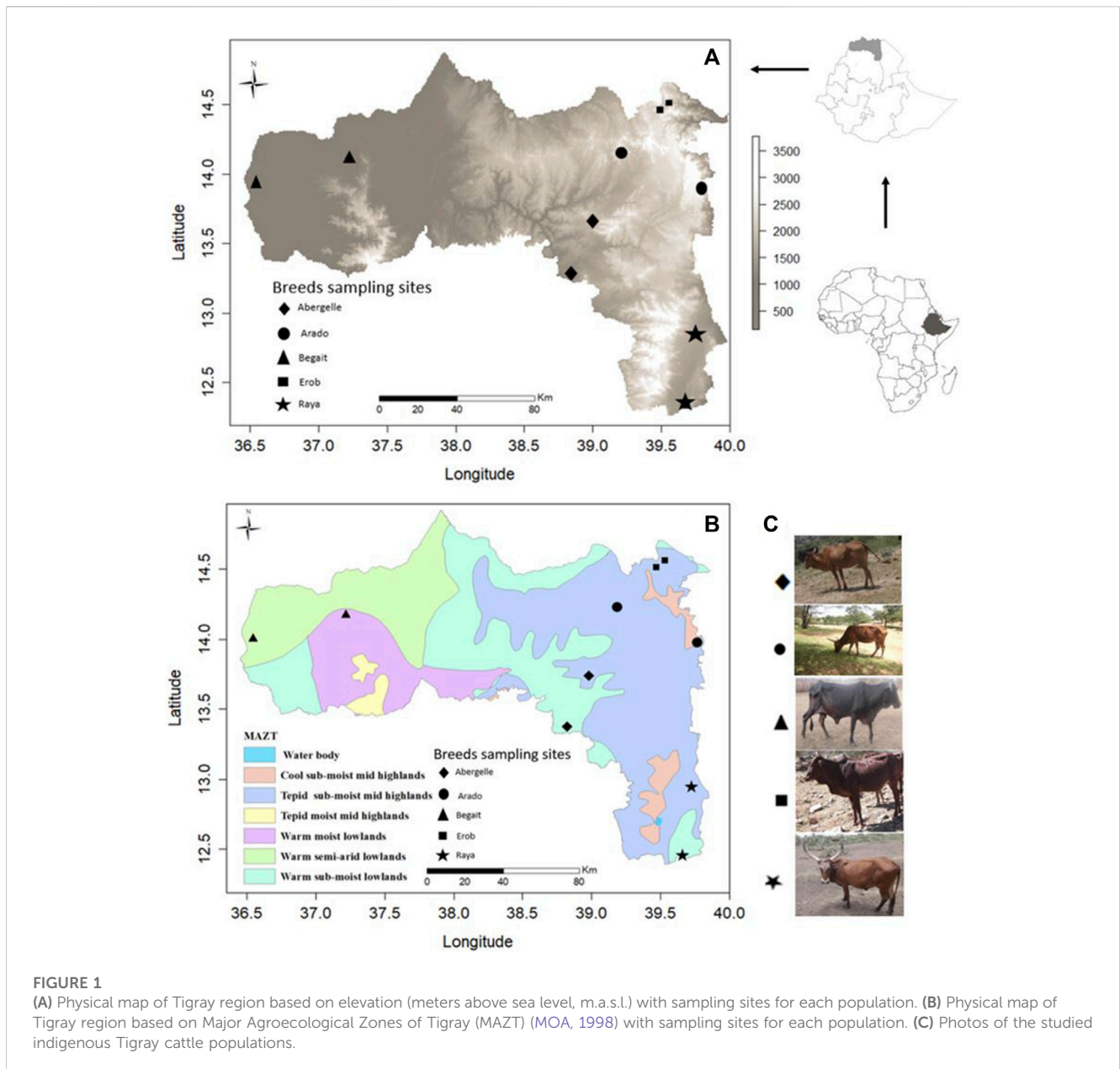
Whole-genome sequence analyses are now the method of choice for genome diversity characterisation (Shendure and Ji, 2008; Stafuzza et al., 2017). At the opposite of microsatellite and SNPs arrays, they provide a complete representation of the diversity of a genome and an entry point to the identification of candidate causative variants associated with Mendelian and quantitative traits (Jiang et al., 2014; Das et al., 2015). Compared to SNPs arrays, often selected for polymorphisms in a reduced number of breeds, they are less prone to ascertainment biases. However, it should be noted that polymorphism detection relies on sequence alignment against a single genome of reference, which will still introduce biases in the identification of SNPs following the genome of reference chosen. Recently, a few studies (Kim et al., 2020; Jang et al., 2022; Terefe et al., 2022; Terefe et al., 2023) have reported the whole-genome characterisation of Ethiopian indigenous cattle, making it the African country with the largest number of cattle genome sequences available. However, they are still several main gaps in our knowledge with cattle populations from some geographic areas and cattle populations living in extreme environments yet to be characterised at the whole genome level (Edea et al., 2013; EBI, 2016). For instance, all the previous whole-genome based characterization studies on Ethiopian cattle populations (Kim et al., 2020; Jang et al., 2022; Terefe et al., 2022) did not include any Tigray cattle, with the exception of one Tigray cattle population (Begait) (Terefe et al., 2023).

We reported previously a multivariate morphological description of the Tigray cattle populations, using 21 qualitative traits and 21 body measurements (Zegeye et al., 2021). For the five Tigray indigenous cattle populations (Arado, Begait, Abergelle, Erob, and Raya cattle), four distinct clusters were identified with the Abergelle and Erob grouped together (Zegeye et al., 2021). Here, we characterised the same five populations using autosomal SNPs and insertion/deletion (indels) variants to assess their genetic diversity, differentiation, relatedness and admixture. We aimed to examine their genetic uniqueness and to pave the way for further analysis to identify genomic regions and, ultimately, the genetic control of their morphological and adaptative traits.

Materials and methods

Sample collection

Fifty-four whole blood samples were collected from five indigenous cattle populations (11 Abergelle, 11 Arado, 11 Begait, 10 Erob, and 11 Raya cattle) in the Tigray region of Northern Ethiopia (Figures 1A–C). The sampling area and morphological descriptions of the populations were reported previously (Zegeye et al., 2021). The whole blood was collected from the jugular vein of each animal by venipuncture with a 10 mL (millilitre) vacutainer



blood collection tube containing ethylenediaminetetraacetic acid (EDTA) as an anticoagulant. The blood was gently mixed with the EDTA and placed into an icebox containing ice. It was brought to the International Livestock Research Institute molecular laboratory facility (ILRI - Addis Ababa), where it was stored at -21°C (degree centigrade) until the extraction of the genomic DNA (gDNA).

Genomic DNA extraction and quality checking

The gDNA was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's standard procedure. The extracted gDNA samples were evaluated for

their concentration and quality using a Nanodrop spectrophotometer (DeNovix-DS-11+spectrophotometer, USA) and 1% agarose gel electrophoresis. A minimum of $5\ \mu\text{g}$ of high quality gDNA with a concentration >50 nanogram per microliter ($\text{ng}/\mu\text{L}$) (Supplementary Table S1) was used for whole-genome sequencing.

Library construction and sequencing

The gDNA samples were sent to the ILRI-CAAS Joint Laboratory of Livestock and Forage Genetic Resources in Beijing, P.R. China, which supervised the genome sequencing. Following the manufacturer's specifications, a paired-end DNA library was constructed for each of the 54 samples. The gDNA was sequenced on an Illumina HiSeq X10 platform.

Short read mapping and variant calling

The sequence reads were checked for quality using FastQC version 0.11.5. Following quality checking, paired-end reads (FASTAQ format) were aligned against the cattle reference genome assembly (ARS_UCD1.2, *Bos taurus*, Hereford breed) using the BWA version 0.7.17 (Li and Durbin, 2009). The mapped reads were sorted using samtools version 1.8 (Li et al., 2009) and converted to BAM formats using PICARD tools version 2.18.2. Duplicated reads were marked and removed using PICARD's MarkDuplicates command. Moreover, the percentages of reads mapped to the reference genome were computed from `dedup_recal.bam` file using the Genome Analysis Toolkit's (GATK, version 3.8-1-0-gf15c1c3ef) `DepthOfCoverage -ct 5 -ct 10 -ct 20 -ct 40`.

The base quality score recalibration (BQSR) was performed using the GATK's BaseRecalibrator and the uniquely mapped reads for variant calling were selected using the GATK's HaplotypeCaller (McKenna et al., 2010). The genomic variants (GVCF files) generated from each sample were jointly analysed using the GATK's GenotypeGVCFtool. Called variants (SNPs and indels) were separately subjected to variant filtration (GATK hard filter) setting `MQ > 40`, `QD > 2.0`, `ReadPosRankSum > 8.0`, `MappingQualityRankSum > 12.5` and `HaplotypeScore > 13` for SNPs and `FS > 200.0`, `QD < 2.0`, `ReadPosRankSum < -20.0` and `QUAL < 20` for indels. Only bi-allelic variants that meet the specified filtering criteria were selected for further analysis.

Variant statistics and annotation

To compute the variant statistics (e.g., total number of SNPs, total number of indels, indel length, and nucleotide substitution), we used the VCF-stats command of VCFtools/0.1.14/Perl. The number of transition and transversion, average ratios of transitions-to-transversions (Ti/Tv), and distribution of SNPs and indels at different allelic frequencies were analysed using stats command and plot-VCF-stats of BCFTools/1.8 (Li et al., 2009). Moreover, SNPs and indels density across chromosomes were computed for each population using VCFtools version 0.1.15 and then averaged using R version 3.6.1 (R Development Core Team, 2019). We searched and compared our SNPs against the dbSNP *ver150* (<https://genome.ucsc.edu/cgi-bin/hgGateway>, last accessed in July 2021). Finally, the variants (SNPs and indels) were classified according to their potential functions using the Ensemble Variant Effect Predictor tool (VEP, (<https://www.ensembl.org/info/docs/tools/vep/index.html>)), and the genes overlapping private missense variants were functionally annotated by DAVID version 6.8 (<https://david.ncifcrf.gov/home.jsp>). Significant Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were selected based on different criteria including $p < 0.05$, Bonferroni < 0.05 , FDR < 0.05 and fold enrichment > 1 .

Genome-wide nucleotide diversity and heterozygosity

The genome-wide nucleotide diversity (π) was analysed for each population using VCFtools version 0.1.15 in 20 kb windows

with a 10 kb sliding step (with the `--window-pi 20000 --window-pi-step 10000` option) (Danecek et al., 2011). The numbers of non-reference heterozygous and homozygous variants (SNPs and indels) were analysed using the VCF-stats command of VCFtools/0.1.14/Perl. Further, the observed heterozygosity (H_o) was calculated following the command `"--het"` in PLINK version 1.9 (Purcell et al., 2007).

Runs of homozygosity and genomic inbreeding

The runs of homozygosity (ROH) were detected using PLINK version 1.9 (Purcell et al., 2007) by setting a sliding window of 50 SNPs (`--homozyg-window-snp 50`), one possible heterozygous genotype (`--homozyg-window-het 1`), two missing genotypes (`--homozyg-window-missing 2`), a minimum SNP density of 1 SNP every 50 kb (`--homozyg-density 50`), a minimum number of 100 SNPs (`--homozyg-snp100`), a minimum length of 100 kb (`--homozyg-kb 100`), a maximum gap of 1 Mb between consecutive homozygous SNPs (`--homozyg-gap 1000`) and the presence of the SNP in at least five homozygous reads (`--homozyg-window-threshold 0.05`).

We calculated the genomic inbreeding value for each cattle by dividing the sum of ROHs length with the total length of the genome ($F_{ROH} = L_{ROH}/LAUTO$), following McQuillan et al. (2008), Zhang et al. (2015a), Addo et al. (2019) and Guo J. et al. (2019). For this analysis, we considered a total genome length of 2,715,853,792 bp (2.72 Gb) (ARS_UCD1.2). An alternate inbreeding coefficient (F_{HOM}) was also calculated for each animal using the `"--het"` command in PLINK version 1.9 (Purcell et al., 2007) following Addo et al. (2019).

Genetic relationship and differentiation

Publicly available genome sequences of 15 cattle breeds (Supplementary Table S2) from six reference groups were added to the dataset for genetic relationship and differentiation analyses. These included African sanga (Afar and Ankole, crosses between African zebu and longhorn humpless taurine), African zenga (Fogera and Horro, crosses between African zebu and sanga), African zebu (Ethiopian Boran and Kenana), African taurine (Muturu and N'Dama), European taurine (Angus and Holstein), Asian zebu (Bhagnari, Cholistani, Dhanni, Sahiwal and Tharparkar). The VCFs of all the reference populations were generated from their raw sequence reads by applying the same procedures mentioned above, and subsequently merged with the five Tigray cattle populations. The merged dataset included 164 cattle genomes and 42,766,398 raw SNPs. It was pruned using PLINK version 1.9 (Purcell et al., 2007) by setting different filtering and quality control thresholds, such as `"--mind 0.25 --geno 0.1 --maf 0.05 --indep-pairwise 50 10 0.5 --set-missing-var-ids C@P"`. Where `--mind 0.25` = individual sample to be removed following 25% or more missing genotype data, `--geno 0.1` = variants to be removed due to 10% of missing genotype data, `--maf 0.05` = variants to be removed due to minor allele frequency less than 0.05, `--indep-pairwise 50 10`

TABLE 1 Variant statistics within cattle populations from Tigray region, Ethiopia.

| Variables | Cattle breeds | | | | |
|----------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | Abergelle | Arado | Begait | Erob | Raya |
| No. of samples | 11 | 11 | 11 | 10 | 11 |
| SNPs | | | | | |
| Novel (%) | 2135111 (7.22) | 2138760 (7.2) | 2091767 (7.23) | 1999834 (7.13) | 2161976 (7.36) |
| Known | 27428853 | 27574364 | 26840780 | 26045211 | 27231606 |
| Total | 29563964 | 29713124 | 28932547 | 28045045 | 29393582 |
| Indels | | | | | |
| Novel (%) | 985169 (34.11) | 991807 (34.13) | 968237 (34.22) | 938479 (34.06) | 1003423 (34.6) |
| Known | 1902736 | 1914050 | 1861562 | 1817017 | 1896421 |
| Total | 2887905 | 2905857 | 2829799 | 2755496 | 2899844 |

No. = number, this abbreviation works for all No. in the tables.

0.5 = SNPs with pairwise r^2 values higher than 0.5 in sliding windows of 50 SNPs moving stepwise with ten SNPs at a time across the genome and set-missing-var-ids C@P = missing IDs set. After applying the quality control and filtering thresholds, the pruned final data set including 3,695,054 SNPs and 164 animals was converted to plink. fam, plink. bin, and plink. bed file using the flag “--make-bed” in PLINK version 1.9 (Purcell et al., 2007).

Principal component analysis

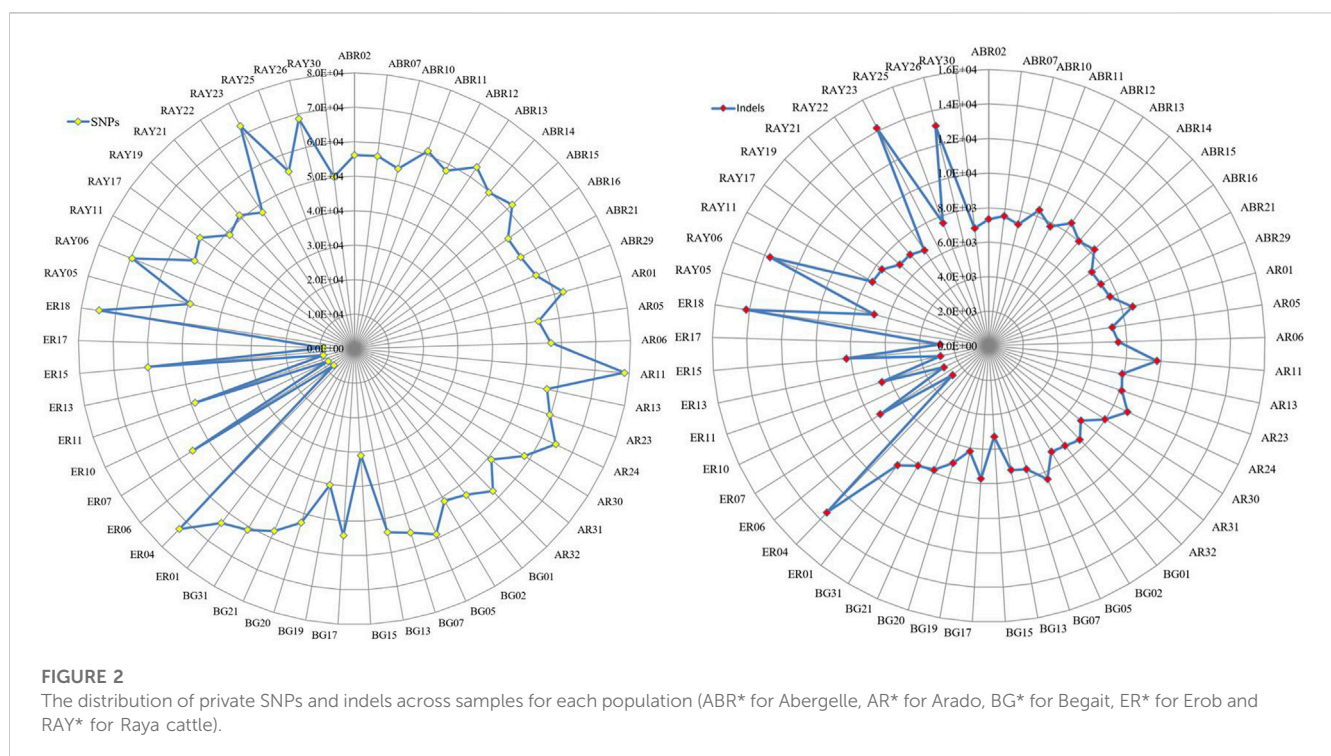
The LD-pruned dataset consisting of 3,695,054 SNPs and 164 individuals was used for principal component analysis (PCA). To calculate `pca.eigenval` and `pca.eigenvec`, we used the flag “`plink-pca`” with a default parameter, for the first 20 principal components (PCs). Then, the proportions of variances explained by the eigenvector were computed by dividing each `eigenval` by the total sum of all `eigenvals` (1–20) and expressing it as a percentage. Finally, the two first PCs were plotted against each other using the `ggplot2` package in R version 3.6.1 (R Development Core Team, 2019) to illustrate the population clustering.

Genetic admixture analysis

Using the same LD-pruned dataset (3,695,054 SNPs), the ADMIXTURE version 1.3.0 software (Alexander et al., 2009) was used to determine the optimal number of clusters (K) and to describe individual ancestry. A cross-validation procedure was performed using the program’s flag “`-cv`” for $K = 1$ to $K = 10$. The K with the lowest cross-validation error was taken as the recommended number of clusters for the dataset. The cross-validation error value for each K (1–10) and the cluster assignments were plotted using R version 3.6.1 (R Development Core Team, 2019).

Genetic differentiation

The genetic distance (F_{ST}) between pairs of populations (Weir and Clark Cockerham, 1984) was analysed using VCFtools version 0.1.15 in 100 kb windows with a 50 kb sliding step (with the `--window-pi 100000 --window-pi-step 50000` option) (Purcell et al., 2007). The pairwise weighted F_{ST} -based heat map with a dendrogram was plotted in R version 3.6.1 (R Development Core



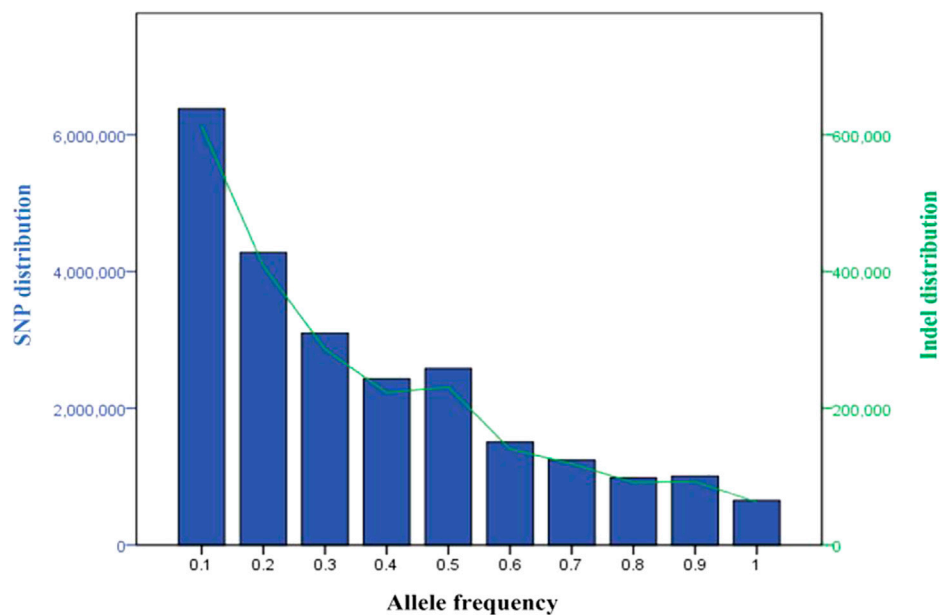


FIGURE 3

Distribution of the variants based on allele frequency (blue bars represent SNPs and the green line represents indels).

Team, 2019). Next, a Neighbor-Net tree based on pairwise F_{ST} values was constructed using the Neighbor-Net algorithm (Bryant and Moulton, 2004) implemented in SplitsTree5 V 5.0.0” (Huson and Bryant, 2006) and plotted in R version 3.6.1 (R Development Core Team, 2019).

Results

Intra-population genetic diversity in the Tigray cattle

Mapping and variant detection

The number of paired-sequence reads for each animal ranged from 200,684,387 to 289,752,799, with a mapping rate of 99.61%–99.79% to the reference genome (ARS_UCD1.2). The average sequencing depth among populations ranged from 10.13 X (Erob cattle) to 10.64 X (Begait cattle). Furthermore, over 88% of the bases were covered with at least five reads, and 39%–42% were covered with at least ten reads (Supplementary Table S3).

Variant calling and filtration combining the five Tigray cattle populations resulted in the detection of around 36 million (M) SNPs ($n = 36,003,573$) and 3.7 M indels ($n = 3,703,659$) (Supplementary Table S4B). The number of SNPs detected per individual sample ranged from 12 M to 13 M (Supplementary Table S4A). The number of SNPs at population level ranged from 28 M to 29 M, of which 7% were novel (Table 1). A total of 2,113,093 (7.15%) SNPs were shared among Abergelle cattle, 2,062,642 (6.94%) among Arado cattle, 2,182,704 (7.54%) among Begait cattle, 2,724,442 (9.71%) among Erob cattle, and 2,161,735 (7.35%) among Raya cattle (Supplementary Table S4B). Around 674,019 (1.87%) SNPs were shared across the five Tigray cattle populations (Supplementary Table S4B).

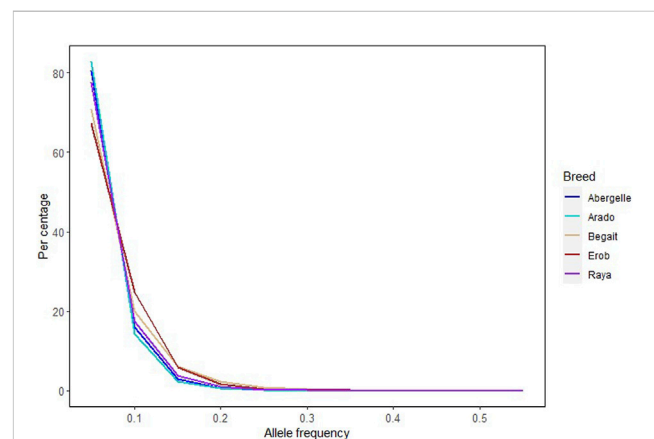


FIGURE 4

Allele frequency of private SNPs, where each coloured line represents a cattle population.

We detected around 1.1 M–1.2 M indels in each individual cattle (Supplementary Table S4A), while the number of indels in each population ranged from 2,755,496 (Erob) to 2,905,857 (Arado). Of these, around 34% were novel (Table 1). Within a population, 177,353 (6.14%), 173,842 (5.98%), 183,302 (6.48%), 228,704 (8.3%) and 181,610 (6.26%) indels were common to all samples in Abergelle, Arado, Begait, Erob, and Raya cattle populations, respectively. Around 1.43% (52,992) of indels were shared across all the five Tigray cattle populations (Supplementary Table S4B).

Except for Erob cattle, the number of private SNPs across individual samples ranged from 32,245 to 81,933, and the number of private indels ranged from 5,276 to 14,182 (Figure 2). Among Erob cattle, four samples (ER06, ER17, ER13 and ER10) had

TABLE 2 Population level summary of annotation of SNPs in the Tigray cattle.

| All consequences | Abergelle | Arado | Begait | Erob | Raya |
|------------------------------------|-----------|----------|----------|----------|----------|
| No. of samples | 11 | 11 | 11 | 10 | 11 |
| Bi-allelic variants processed | 29563964 | 29713124 | 28932547 | 28045045 | 29393582 |
| Splice donor variant | 707 | 684 | 693 | 672 | 690 |
| Splice acceptor variant | 439 | 452 | 442 | 429 | 435 |
| Stop gained | 1682 | 1748 | 1706 | 1648 | 1765 |
| Stop lost | 245 | 238 | 237 | 239 | 243 |
| Start lost | 397 | 406 | 382 | 373 | 413 |
| Missense variant | 149334 | 150764 | 146883 | 141733 | 149448 |
| Splice region variant | 40755 | 41135 | 39987 | 38861 | 40791 |
| Synonymous variant | 228481 | 229776 | 224200 | 217399 | 228249 |
| Stop retained variant | 169 | 183 | 169 | 160 | 180 |
| Coding sequence variant | 1 | 1 | 1 | 1 | 1 |
| Mature miRNA variant | 123 | 125 | 121 | 120 | 125 |
| 5_prime UTR variant | 49794 | 49766 | 48751 | 46913 | 50016 |
| 3_prime UTR variant | 137652 | 138671 | 135235 | 131341 | 137754 |
| Non-coding transcript exon variant | 31400 | 31309 | 30515 | 29622 | 31084 |
| Intron variant | 22387867 | 22473356 | 21940886 | 21255569 | 22194958 |
| Non-coding transcript variant | 502772 | 505660 | 496259 | 476848 | 500512 |
| Upstream gene variant | 2074721 | 2090124 | 2042625 | 1979084 | 2075740 |
| Downstream gene variant | 2111713 | 2121229 | 2070242 | 2007161 | 2104907 |
| Intergenic variant | 17584768 | 17682416 | 17197428 | 16677542 | 17502131 |

fewer private variants (9,929 to 11,526 SNPs and 2,675 to 2,870 indels) compared to the remaining Tigray cattle samples. At the population level, we detected 571,535, 634,275, 583,831, 433,013 and 569,013 private SNPs for Abergelle, Arado, Begait, Erob and Raya cattle populations, respectively.

Density of the variants and their allele frequencies

The density of genome-wide SNPs ranged from 11.27 ± 7.69 to 11.94 ± 7.88 SNPs/kb and of indels from 1.08 ± 1.34 to 1.17 ± 1.41 indels/kb across the five Tigray cattle populations (Supplementary Tables S5, S6). Chromosomes 23, 27 and 28 had the highest density of variants (13–14 SNPs/kb and 1.3 to 1.4 indel/kb), while chromosomes 19, 13, 3 and 11 had the lowest ones (10–11 SNPs/kb and <1.1 indels/kb). The chromosome-wise distributions of variants (SNPs and indels) were proportional to the length of the chromosomes (Supplementary Tables S7, S8). As expected, large chromosomes had more variants than small ones (Supplementary Figures S1A–D). However, the density of variants (SNPs/kb or indel/kb) was higher on small chromosomes than large ones.

Across the five Tigray indigenous cattle populations, the average alternate (non-reference) allele frequencies of SNPs and indels were 0.32 and 0.28 to 0.3, respectively. The proportion of SNPs with mean alternate allele frequency (AAF) < 0.5 ranged from 77% to 79% and the proportion of SNPs with mean AAF > 0.9 was around 4%. The

proportion of indels with mean AAF < 0.5 ranged from 78% to 80% (Supplementary Table S9). However, most of the variants (SNPs and indels) had frequencies of 10% or less (Figure 3). Allele frequencies of private SNPs ranged from 0.05 to 0.55, of which 67% (Erob cattle) to 83% (Arado cattle) of these SNPs had an allele frequency of 0.05 (Figure 4).

Nucleotide substitutions and indel length

The Ti/Tv ratio was around 2.35 (Supplementary Table S10). It supported a high sequencing accuracy for all samples. Across all the samples, the highest number of nucleotide substitutions were recorded for the bases Cytosine to Thymine (C > T) and the bases Guanine to Adenine (G > A) while the least number of nucleotide substitutions for the bases Adenine to Thymine (A > T) and Thymine to Adenine (T > A) (Supplementary Table S10 and Supplementary Figure S2A). The number of insertions was about 0.33–0.35 times higher than deletions. Furthermore, the length of indels ranged from –28 bp (deletion, Abergelle) to +23 bp (insertion, Begait) (Supplementary Table S11). Almost 50% of the indels had a length of 1 bp, while the majority of indels were less than 5 bp (87.12% in Abergelle, 87.08% in Arado, 87.11% in Begait, 87.18% in Erob and 87.11% in Raya cattle). Only around 13% of the total indels had lengths greater than or equal to 6 bp (Supplementary Table S11 and Supplementary Figure S2B).

TABLE 3 Population level summary of annotation of indels in the Tigray cattle.

| All consequences | Abergelle | Arado | Begait | Erob | Raya |
|------------------------------------|-----------|---------|---------|---------|---------|
| No. of samples | 11 | 11 | 11 | 10 | 11 |
| Bi-allelic variants processed | 2887905 | 2905857 | 2829799 | 2755496 | 2899844 |
| Transcript ablation | 4 | 5 | 5 | 5 | 5 |
| Splice donor variant | 248 | 240 | 236 | 206 | 227 |
| Splice acceptor variant | 232 | 234 | 215 | 208 | 221 |
| Stop gained | 57 | 61 | 64 | 62 | 61 |
| Frameshift variant | 4788 | 4875 | 4688 | 4737 | 4977 |
| Stop lost | 28 | 28 | 28 | 24 | 25 |
| Start lost | 42 | 39 | 41 | 34 | 38 |
| Inframe insertion | 825 | 834 | 815 | 799 | 853 |
| Inframe deletion | 1565 | 1578 | 1561 | 1540 | 1781 |
| Protein altering variant | 39 | 41 | 46 | 40 | 44 |
| Splice region variant | 3797 | 3913 | 3805 | 3716 | 3952 |
| Stop retained variant | 17 | 17 | 21 | 22 | 20 |
| Start retained variant | 13 | 14 | 13 | 13 | 12 |
| Coding sequence variant | 202 | 204 | 206 | 191 | 206 |
| Mature miRNA variant | 12 | 16 | 13 | 16 | 17 |
| 5 prime UTR variant | 4855 | 4835 | 4842 | 4647 | 4985 |
| 3 prime UTR variant | 16000 | 16025 | 15703 | 15327 | 16165 |
| Non-coding transcript exon variant | 2439 | 2440 | 2409 | 2321 | 2441 |
| Intron variant | 2237586 | 2246670 | 2192419 | 2135981 | 2239005 |
| Non-coding transcript variant | 47834 | 48290 | 47388 | 45827 | 48017 |
| Upstream gene variant | 214934 | 216603 | 211424 | 205980 | 217097 |
| Downstream gene variant | 230871 | 232624 | 227186 | 220614 | 232751 |
| Intergenic variant | 1642232 | 1653813 | 1607488 | 1565117 | 1648903 |

Functional distribution of variants (SNPs and indels)

The annotation of the SNPs showed that around 59.5% of them were in the intergenic regions. Around 76% of annotated SNPs were in introns, 7% in upstream of genes, 7.2% in downstream of genes, 0.5% in 3' untranslated region (UTR), 0.2% in 5' UTR and 0.11% in non-coding transcript exon. The number of SNPs in the coding regions (stop gain, stop lost and stop retained, start lost, missense and synonymous SNPs, and coding sequences) was approximately 0.01% in all populations (380,309 in Abergelle, 383,116 in Arado, 373,578 in Begait, 361,55 in Erob and 380,299 in Raya cattle) (Table 2) (Supplementary Figure S3), of which around 17%–18% had deleterious effects (Supplementary Figure S4).

Around 57%, 77%, 7.4%, 8%, 0.6%, 0.2% and 0.1% of the indels were in intergenic regions, introns, upstream of genes, downstream of genes, 3' UTR, 5' UTR and non-coding transcript exons, respectively. The total numbers of indels located within the coding regions (stop gain, stop lost, stop retained, start lost, start

retained, frameshifts, inframe insertions, inframe deletions, protein-altering variants and coding sequences) ranged from 0.22% (Raya cattle, 6,414) to 0.27% (Erob cattle, 7,462) (Table 3). Among the indels located in the coding regions, 63.20%, 63.39%, 62.65%, 63.48% and 77.6% resulted in codon frameshifts (codon alteration), of which 0.81%, 0.84%, 0.98%, 0.84% and 0.88% may affect protein functions in Abergelle, Arado, Begait, Erob, and Raya cattle populations, respectively (Supplementary Figure S5).

Enrichment analysis of the genes overlapping private SNPs

A separate analysis of the private SNPs for each population showed 1,455, 1,809, 1,470, 1,203, and 1,701 private missense SNPs for Abergelle, Arado, Begait, Erob and Raya cattle populations, respectively; of which 97.9% (Erob) to 98.9% (Raya) were in coding regions while the remaining ones overlapped with splice regions. Of the missense SNPs in the coding regions, 33.1%, 34.4%, 52.7%, 64.8% and 58.1% had deleterious effects in Abergelle, Arado, Begait, Erob and Raya cattle populations, respectively (Figure 5).

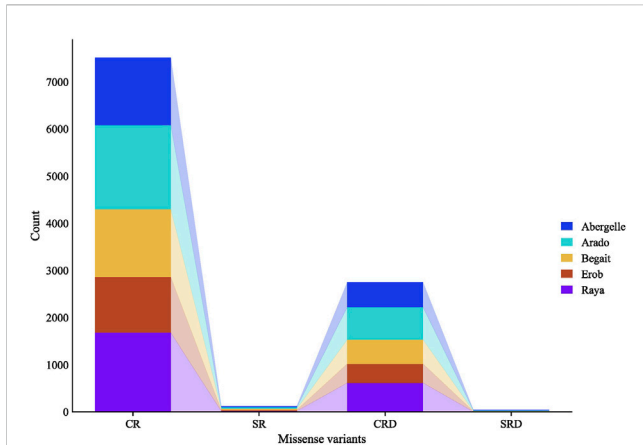


FIGURE 5
Private SNPs with missense effects overlapping coding and splicing regions, where, CR represents the number of missense SNPs overlapping coding regions, SR represents the number of missense SNPs overlapping splice regions, CRD represents the number of missense SNPs overlapping coding regions with deleterious effects, and SRD represents the number of missense SNPs overlapping splice regions with deleterious effects.

Functional enrichment analysis of the genes overlapping private SNPs in coding regions identified 16, 12, 16, 6 and 10 significant ($p < 0.05$, Bonferroni < 0.05 , FDR < 0.05 and fold enrichment > 1) GO terms of biological process (BP), cellular component (CC), molecular function (MF) and KEGG pathways in Abergelle, Arado, Begait, Erob and Raya cattle populations, respectively (Figure 6). Out of the enriched terms, the top three most significant ones ($p = 1.6 \times 10^{-5}$ to 6.8×10^{-67} , Bonferroni = 1.65×10^{-2} to 7.5×10^{-64} , FDR = 8.3×10^{-3} to 7.5×10^{-64} and fold enrichment = 2.3–4.3) were olfactory receptor

activity (GO:0004984), olfactory transduction (bta04740) and odorant binding (GO:0005549). These were common to all five Tigray cattle populations.

Within populations, the enrichment analysis further identified many population specific genes associated with the aforementioned three most significant terms. Around 11% (Erob cattle) to 17% (Abergelle cattle) of the genes were associated with olfactory receptor activity, 13% (Erob cattle) to 18% (Abergelle cattle) with olfactory transduction, and 3% (Erob cattle) to 6% (Abergelle cattle) with odorant binding (Additional file, Sheet 1).

Eleven genes, including *OR4F73*, *OR1L21*, *OR5AN1*, *OR9S29*, *OR9M1D*, *OR2H20*, *OR4X16*, *OR5AK29*, *OR6C4*, *OR8B1AU* and *OR9S40*, were commonly enriched in the three shared GO terms. Six GO terms (GO:0004871~signal transducer activity, GO:0004872~receptor activity, GO:0004888~transmembrane signaling receptor activity, GO:0038023~signaling receptor activity, GO:0060089~molecular transducer activity, GO:0099600~transmembrane receptor activity) with related molecular functions were enriched in four populations (Abergelle, Arado, Begait and Raya cattle). Interestingly, two significant GO terms of the cellular component associated with an integral component of the plasma membrane (GO:0005887) and the intrinsic component of the plasma membrane (GO:0031226) were only enriched in Begait cattle. Two GO terms of the cellular component (GO:0005578) related with proteinaceous extracellular matrix and extracellular matrix component (GO:0044420) were significantly enriched only in the Erob cattle (Additional file, Sheet 1).

Nucleotide diversity and heterozygosity

The average genome-wide nucleotide diversity (π) ranged from $3.5 \pm 1.77 \times 10^{-3}$ (Raya) to $3.57 \pm 1.76 \times 10^{-3}$ (Arado) (Figure 7 and Supplementary Table S12). Besides, the average non-reference heterozygous variants (SNPs and indels) were around 0.6 to 0.7 times higher than the corresponding homozygous variants

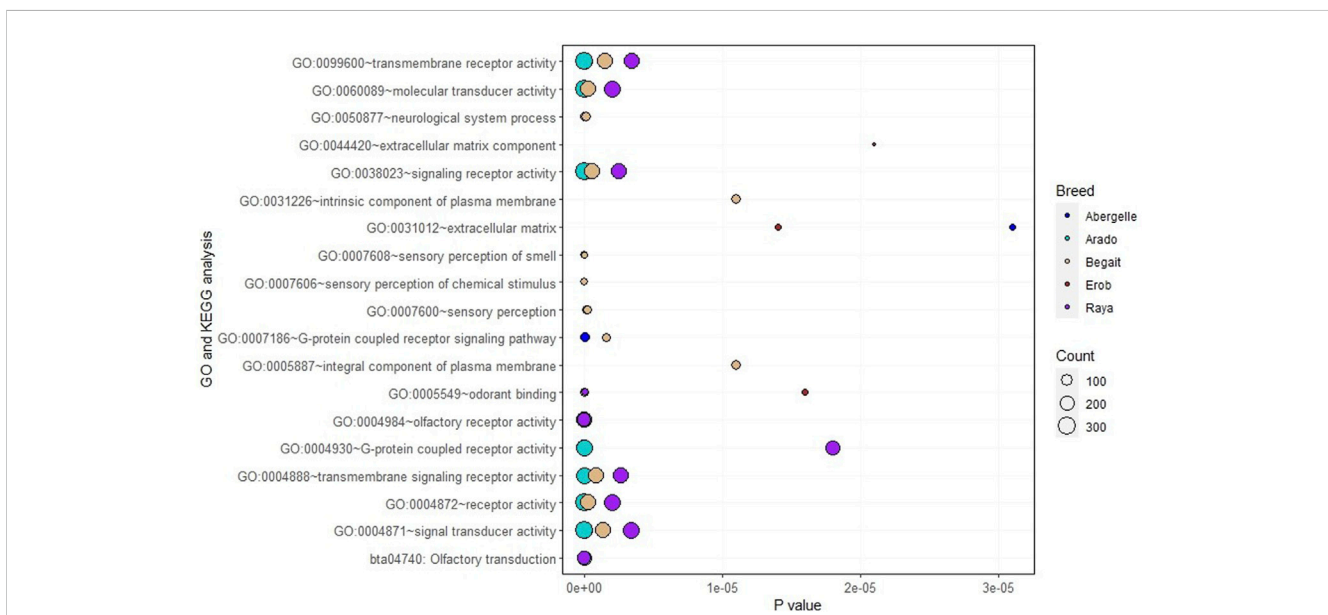


FIGURE 6
Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms in the Tigray cattle, where the size of the circles represents how large the number of genes represented in a specific GO or KEGG pathway term and the level of significance, while each coloured circle represents a cattle population.

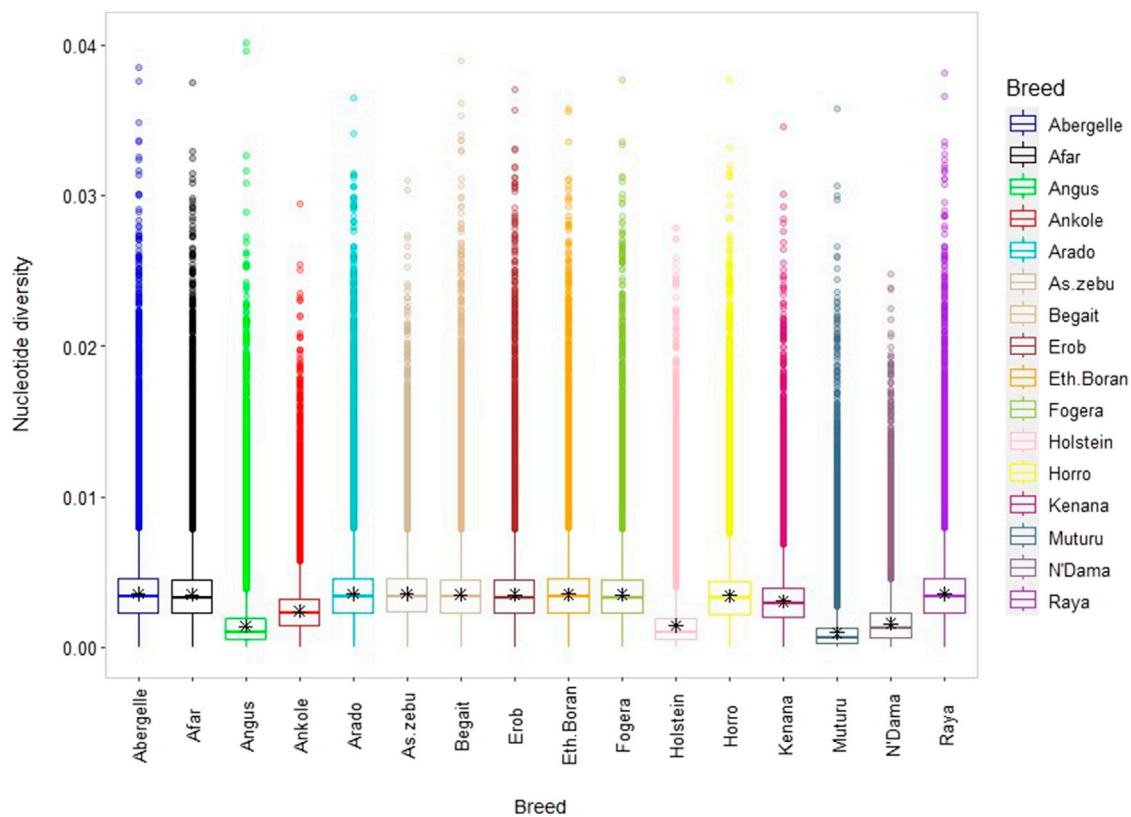


FIGURE 7
Box plot for nucleotide diversity, where each coloured box plot represents a cattle population.

(Supplementary Table S12). At the individual level, the ratio of heterozygous to homozygous SNPs ranged from 1.39 to 1.89, and of the indels from 1.39 to 1.84 (Supplementary Figures S6A–E). The mean observed heterozygosity (H_o) was the highest in Arado cattle (0.302 ± 0.010) but the lowest in Raya cattle (0.278 ± 0.016) (Supplementary Table S12).

Runs of homozygosity and genomic inbreeding Abundance and length of ROH in the Tigray cattle compared to major reference cattle groups

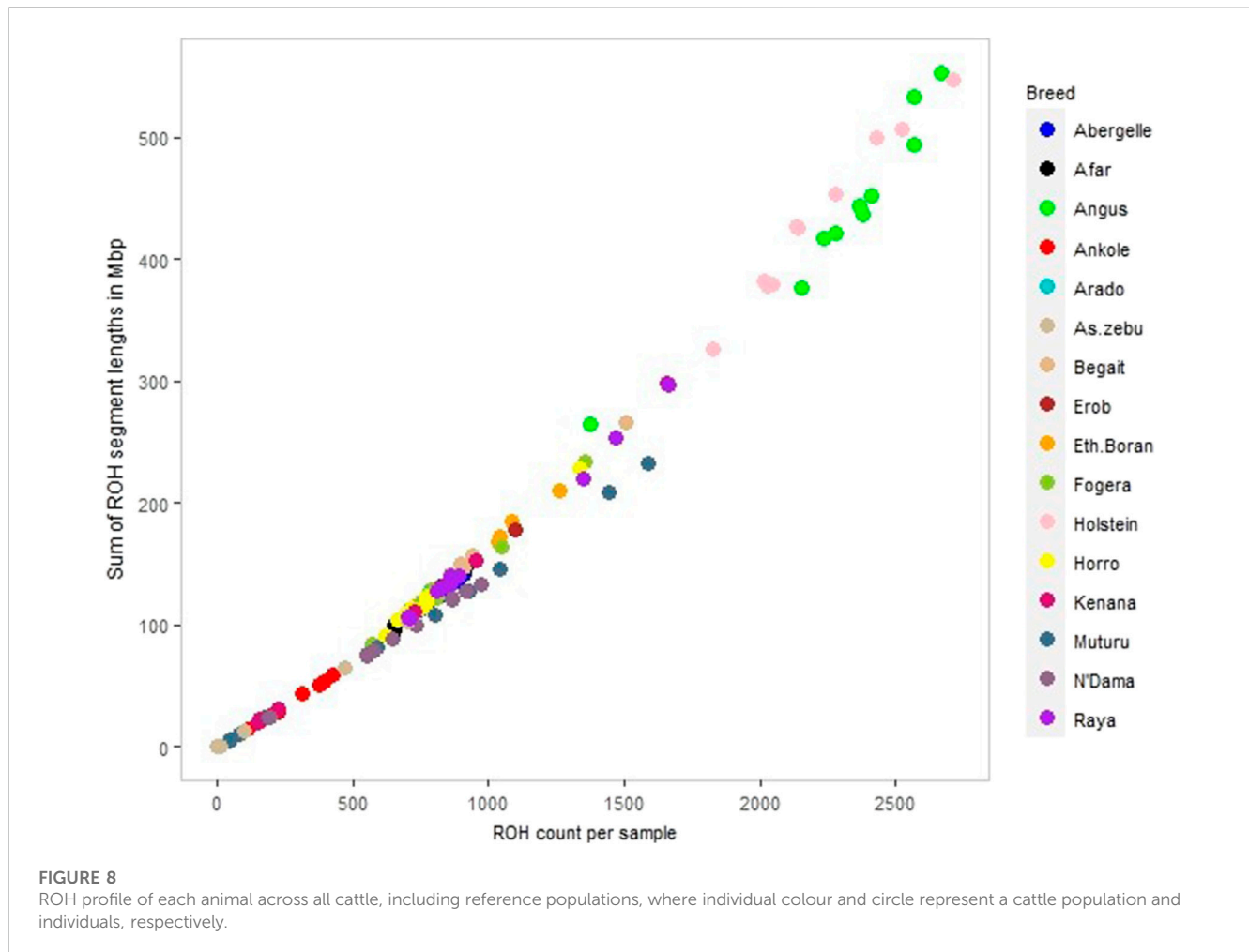
We calculated ROH for the five Tigray cattle populations and the reference breeds. The average number and length of ROH segments varied considerably within and among breeds (Figure 8; Table 4 and Supplementary Table S13). The within breed average number of ROH for the Tigray cattle ranged from 777.82 (Arado cattle) to 1000.45 (Raya cattle), and the within breed average sum of the length of ROH ranges from 122.01 megabase pairs (Mbp) (Arado cattle) to 163.88 Mbp (Raya cattle). The average number of ROHs and the average sum of the length of ROH in the Tigray cattle were higher than the ones recorded in Asian zebu, African zebu from Sudan (Kenana), African taurine (Muturu and N'Dama) and African sanga from Uganda (Ankole). But they were much lower compared to European taurine cattle (Holstein and Angus). However, in general, all the Tigray cattle had close ROH profiles with the other cattle populations originating from Ethiopia (Afar, Eth. Boran, Fogera and Horro) (Table 4).

Distribution of ROH based on segment length categories

The number of ROH across length categories (0.1–0.25 Mbp, > 0.25–0.5 Mbp, > 0.5–1 Mbp and > 1 Mbp) varied among breeds. ROH in the length category of 0.1–0.25 Mbp accounted for 90%–92% of the total ROH. For the length categories >0.25–0.5 Mbp and > 0.5–1 Mbp, ROH frequencies were about 7%–10% and 0.2%–0.4%, respectively (Supplementary Tables S14, S15). ROH > 1 Mbp were only found in Begait and Raya cattle (BG15, RAY11, RAY22 and RAY26) (Supplementary Table S13). In the length categories 0.1–0.25 Mbp, the Tigray cattle had more ROH when compared to Holstein and Angus, but they were less than the African sanga from Uganda (Ankole), African taurine (Muturu and N'Dama) and the Asian zebu. Above the 0.25 Mbp length category, the Tigray cattle had more ROH when compared to Ankole, Muturu, N'Dama and Asian zebu, but they were less than Angus and Holstein (Supplementary Tables S14, S15).

Genomic positions under runs of homozygosity and inbreeding across the Tigray cattle

The chromosome-wise distributions of the number and length of ROH and the incidence of SNPs on ROH were different across the five Tigray cattle populations, except for the Abergelle and Erob cattle (Figures 9A–C and Supplementary Tables S16–S18). At the population level, the mean genomic inbreeding coefficient was the smallest in Arado ($F_{ROH} = 0.047 \pm 0.004$ and $F_{HOM} = 0.043 \pm 0.035$) but the highest in Raya cattle ($F_{ROH} = 0.064 \pm 0.025$ and $F_{HOM} = 0.07 \pm 0.054$)



(Supplementary Table S19). One Begait cattle (BG15 with $F_{ROH} = 0.103$ and $F_{HOM} = 0.107$) and three Raya cattle (RAY11 with $F_{ROH} = 0.114$ and $F_{HOM} = 0.12$, RAY22 with $F_{ROH} = 0.099$ and $F_{HOM} = 0.168$, and RAY26 with $F_{ROH} = 0.088$ and $F_{HOM} = 0.111$) showed strong genomic inbreeding values based on both methods (Figure 9D and Supplementary Table S20).

Relationship and inter-population genetic differentiation

Principal component analysis

The PCA showed the presence of six potential clusters of populations (Figure 10). PC1 and PC2 explained 32.27% and 12.33% of the total variation, respectively (Figure 10 and Supplementary Figure S7). PC1 separated Ankole and taurine cattle (African and European) from Asian zebu, Kenana and all the cattle from Ethiopia, including the Tigray cattle (Abergelle, Arado, Begait, Erob and Raya). PC2 divided the European taurine cattle and Asian zebu from the Ankole and African taurine cattle (Figure 10). Combining PC1 and PC2 illustrated that the Tigray cattle populations (Abergelle, Arado, Begait, Erob and Raya) were close to the Asian zebu and the African sanga (Ankole).

PC1 and PC2 for the five Tigray cattle populations and other Ethiopian cattle representing three cattle groups of African zebu (Ethiopian Boran), African sanga (Afar) and African zenga (Fogera and Horro) jointly accounted for 15.34% of the total variation, of which the Begait and Erob cattle were separated from the other populations (Figure 11A). The PC1 (6.09%) of the five Tigray cattle populations alone separated Begait cattle from the other four Tigray cattle populations, while PC2 (5.23%) divided Raya cattle from the other four Tigray cattle populations (Figure 11B).

Genetic admixture and population genetic differentiation

As indicated by the lowest cross-validation error (0.51) (Supplementary Figure S8), the admixture analysis suggested three ancestral sources. At $K = 3$, the taurine ancestry for the Tigray cattle was shown to be mainly shared with the African taurine, except for some individuals in Arado ($n = 3$) and Begait ($n = 2$) cattle having 0.1%–1.8% of European taurine ancestry (Figure 12). In each population, the African taurine ancestry ranged from 11.3% (Erob cattle) to 14.1% (Begait cattle) and, accordingly, the indicine ancestry from 85.6% (Arado cattle) to 88.7% (Erob cattle) (Supplementary Figure S9). More interestingly, as the number of potential ancestries increased, the Tigray cattle

TABLE 4 Number of animals with and without ROH, breed-wise total and average number of ROH and average sum of the length of ROH for the Tigray cattle populations compared to reference cattle populations.

| Breed | No. w/o ROH ^a | No. w/ ROH ^b | Total no. of ROH ^c | Avg. no. of ROH segments (min-max) ^d | Avg. ROH segment lengths in Mbp (min-max) ^e |
|------------|--------------------------|-------------------------|-------------------------------|---|--|
| Abergelle | 0 | 11 | 9131 | 830.09 (748–907) | 130.13 (112.92–148.83) |
| Arado | 0 | 11 | 8556 | 777.82 (706–861) | 122.01 (108.02–135.65) |
| Begait | 0 | 11 | 10003 | 909.36 (767–1504) | 148.9 (121.14–266.71) |
| Erob | 0 | 10 | 8371 | 837.1 (727–1099) | 131.92 (110.40–177.46) |
| Raya | 0 | 11 | 11005 | 1000.45 (703–1662) | 163.88 (105.87–297.29) |
| Eth.Boran | 0 | 10 | 8937 | 893.7 (734–1259) | 143.94 (113.6–210.41) |
| Kenana | 0 | 10 | 4624 | 462.4 (150–1657) | 73.79 (20.24–298.06) |
| Fogera | 0 | 9 | 7817 | 868.56 (571–1353) | 124.5 (84.47–234.48) |
| Horro | 0 | 11 | 8508 | 773.45 (618–1334) | 121.74 (91.91–228.27) |
| Ankole | 0 | 10 | 2299 | 209 (6–423) | 31.16 (0.78–58.99) |
| Afar | 0 | 10 | 7441 | 744.1 (649–918) | 115.21 (95.06–146.82) |
| Holstein | 0 | 10 | 22124 | 2212.4 (1827–2710) | 432.61 (325.89–547.33) |
| Angus | 0 | 10 | 22999 | 2299.9 (1374–2670) | 439.31 (264.38–553.70) |
| Muturu | 0 | 10 | 6755 | 675.5 (40–1590) | 94.88 (4.79–232.72) |
| N'Dama | 0 | 10 | 6870 | 687 (194–969) | 94.49 (23.67–133.10) |
| Asian zebu | 1 | 9 | 1292 | 143.56 (0–704) | 20.39 (113.57–103.09) |

^aNumber of animals without ROH.

^bNumber of animals with ROH.

^cTotal number of ROH across each population.

^dAverage number of ROH segments (Minimum to maximum).

^eAverage ROH segment length in megabyte (Minimum to maximum).

local ancestry appeared. At $K = 7$ and $K = 10$, both Erob and Begait cattle showed some unique local ancestries (Supplementary Figure S10).

F_{ST} for the Tigray cattle populations ranged from 0.07 to 0.08 with Asian Zebu, 0.084 to 0.108 with Ankole (African sanga originating from Uganda), 0.236 to 0.264 with N'Dama (African taurine), 0.328 to 0.36 with Muturu (African taurine) and 0.300 to 0.335 with the European taurine cattle (Angus and Holstein) (Supplementary Table S21). Within Ethiopian cattle, we observed two groups among the Tigray cattle populations for the F_{ST}

estimates with other Ethiopian cattle populations, with higher genetic differentiation ($F_{ST} > 0.02$) for Begait and Erob cattle from Ethiopian Boran, Fogera and Horro cattle than for Abergelle, Arado and Raya cattle (F_{ST} around 0.01) (Figure 13 and Supplementary Table S21).

Overall, the heat map and dendrogram (Figure 13 and Supplementary Table S21) generated from pairwise weighted F_{ST} values among the Tigray cattle populations and the Tigray cattle populations against other cattle breeds (African sanga, African zenga, African zebu, Asian zebu, African and European taurine cattle) showed two main genetic clades: One comprising the taurine group with two sub-clusters African and European) and another including the Asian zebu and other non-taurine African origin breeds (including the Tigray cattle populations). This was consistent with the PCA and admixture analysis results. Further, the Neighbor-Net tree based on the pairwise F_{ST} values (Figure 14) supported the admixture, the heat map and the dendrogram results.

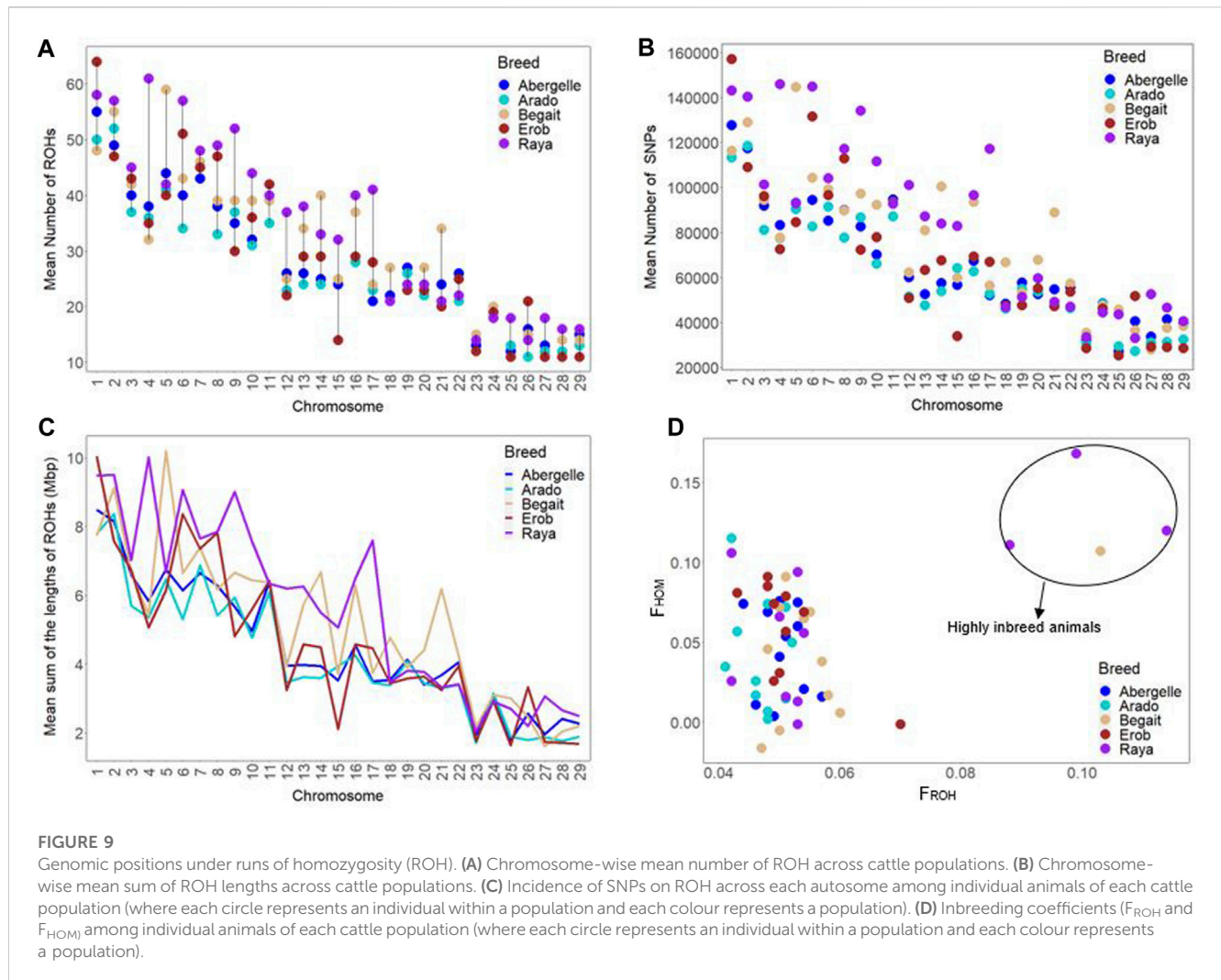
Discussion

We report the first whole-genome-sequence-based characterisation of the genetic diversity, relatedness and admixture of cattle populations indigenous to the Ethiopia's Tigray region. We used abundance, distribution and functional description of SNPs and indels, genome-wide nucleotide diversity (π), heterozygosity (Ho), runs of homozygosity (ROH) and genomic inbreeding coefficient to evaluate the intra-population genetic diversity. The pairwise population differentiation (F_{ST}) and relationship based on PCA and admixture analysis were employed to assess the inter-population differentiation and relationship among the Tigray cattle and between the Tigray cattle and other reference cattle groups (African sanga, African zenga, African and Asian zebu, and African and European taurine cattle).

We compared our findings with previous studies using the same ARS-UCD1.2 as reference genome (Rosen et al., 2020). For cattle populations as in our study, predominantly of indicine ancestry, this would have likely inflated the number of detected SNPs. Aligning our reads to an indicine reference genome would minimize subspecies ascertainment of SNPs biases (Low et al., 2020).

The alignment rates of the Tigray cattle sequence reads were similar to those of other African (Kim et al., 2020), Chinese (Jiaxian Red, Wenshan, Wannan and Leiqiong) (Zhang et al., 2019; Xia et al., 2021) and European (Angus and Holstein) (Kim et al., 2020) cattle breeds, suggesting the overall similarity in autosomal structures across cattle populations in the world, despite their distinct evolutionary histories.

The high variation in number of genetic variants (SNPs and indels) within and across the Tigray cattle populations illustrated their rich genetic diversity. Also, we found a substantial number of novel variants (SNPs and indels) in the Tigray cattle (Table 1), indicating their importance as a reservoir of genetic diversity previously uncharacterized. Interestingly, we found many new indels (around 34%) compared to novel SNPs. However, it should be emphasized that indels have been given so far less attention in cattle genomic analyses (Stafuzza et al., 2017), despite being part of the important drivers of phenotypic and



genetic diversity (Iqbal et al., 2019). Most of the whole genome analyses on African cattle breeds were based on SNPs (Kim et al., 2017; Tijjani et al., 2019; Kim et al., 2020; Jang et al., 2022; Mauki et al., 2022; Terefe et al., 2022; Terefe et al., 2023), while our study is the first to report indels for Ethiopian cattle.

High genetic “functional” variability in the Tigray cattle

We observed private variants in each Tigray cattle population. Though they only represented around 1.5% (Erob cattle) to 2.1% (Arado cattle) of the total variants in respective populations, they might serve as important diagnostic markers. A small proportion of these SNPs (around 0.23%–0.32%) were missense variants, of which the majority were located in coding regions (97.9%–98.9%), while a few (1.1%–2.1%) were in splice regions. Amongst these SNPs, one-third to two-thirds had a predicted deleterious effect.

Further analysis of all private missense variants identified several GO terms and KEGG pathways shared by different populations or to be population specific. The two most significant GO terms of the molecular function (the olfactory receptor activity and odorant binding) and one most significant KEGG pathway (olfactory

transduction) present in the five Tigray cattle populations were associated with olfaction or odour recognition. An efficient olfactory reception is an important fitness mechanisms essential for adaptation, including food and water search behaviour and reproduction (Kour et al., 2022). Odour recognition influences food intake identification and preference (Soria-Gómez et al., 2014).

The GO terms of the cellular component related to the integral component of the plasma membrane (GO:0005887) and its subtype intrinsic component of the plasma membrane (GO:0031226) were only found in Begait cattle. In these GO terms, several important genes were found to be associated with morphology, production, reproduction, feed efficiency, immune response and environmental adaptation. For example, *SCN4A* (Cai et al., 2019) and *TAS1R2* (Zhang et al., 2012) were reported to be associated with body height in cattle. *KCNG4* was found to be related to morphometric traits like rump height, body length and chest depth in goats (Easa et al., 2022). *FLT4* was identified to be relevant to proliferation and growth in cattle (Keogh et al., 2019). *GABRG1* was implicated in milk yield (Pedrosa et al., 2021). Other genes in Begait cattle included *PCDH8* (Tausat et al., 2020) and *SLC26A3* (Kern et al., 2016) associated with feed efficiency in cattle, *PCDH18* related to the immune system and adipogenesis (de Lima et al., 2020). *DUOX2* is important for thyroid hormones production and in innate immunity (Maruo et al.,

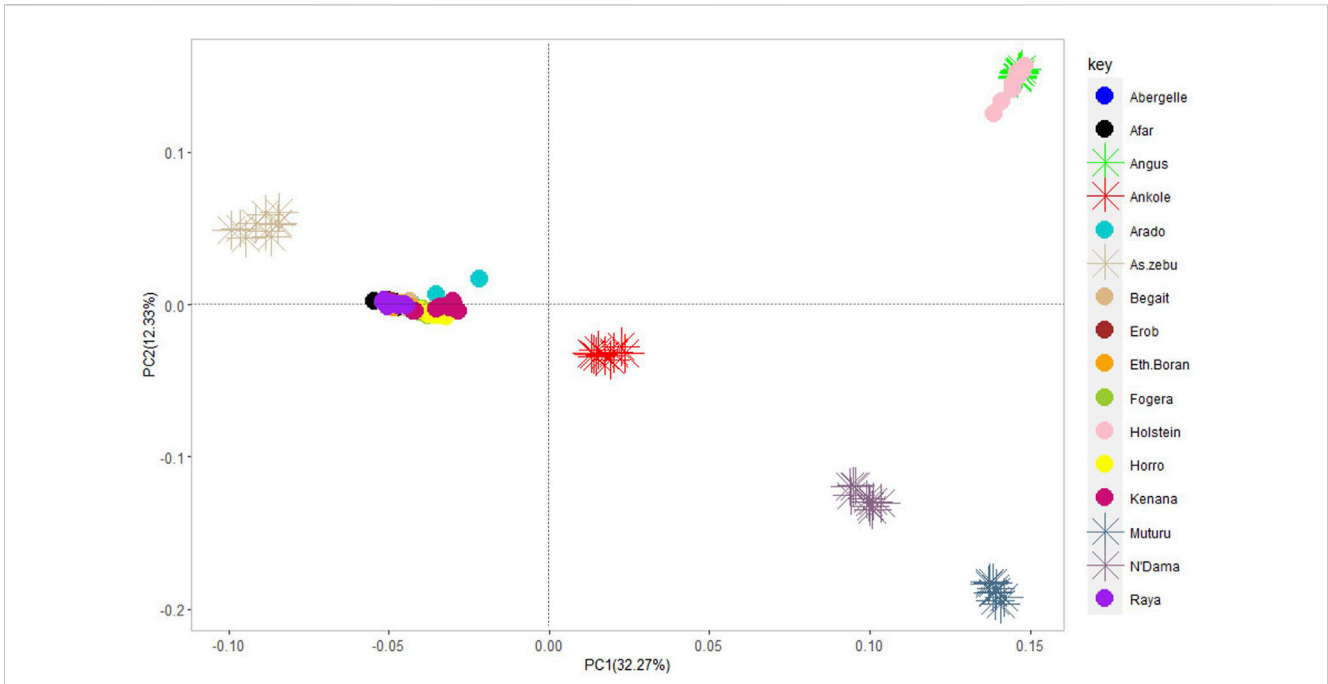


FIGURE 10
Principal component analysis plot (PC1 and PC2) of the five Tigray cattle populations and reference cattle groups. African sanga (Afar and Ankole, crosses between African zebu and longhorn humpless taurine), African zenga (Fogera and Horro, crosses between African zebu and sanga), African zebu (Ethiopian Boran and Kenana), African taurine (Muturu and N'Dama), European taurine (Angus and Holstein) and Asian zebu (Bhagnari, Cholistani, Dhanni, Sahiwal and Tharparkar).

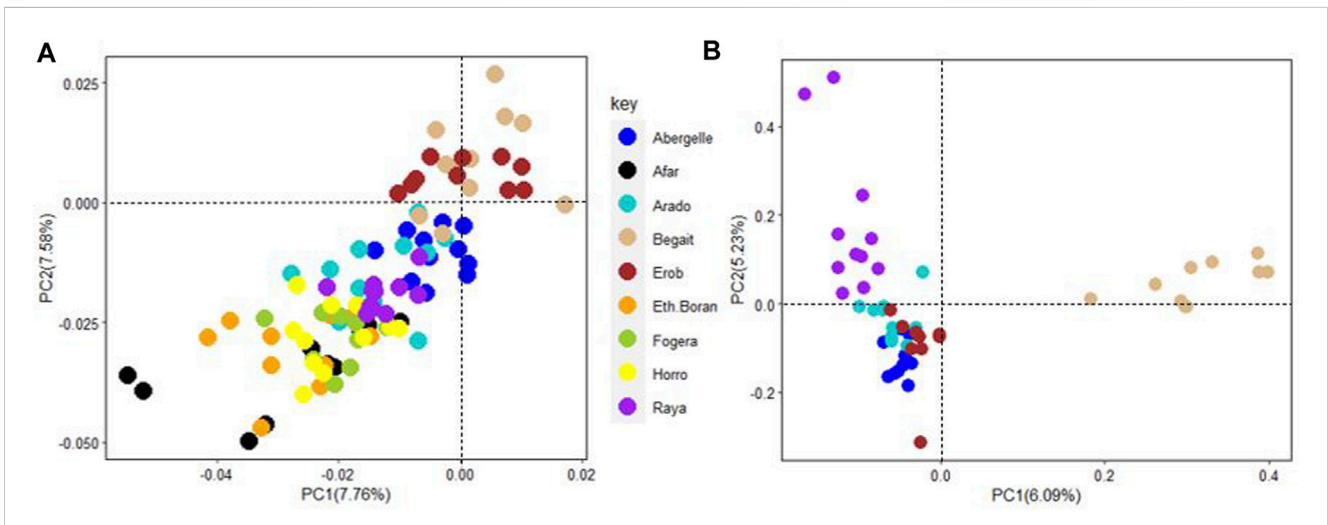


FIGURE 11
(A) Principal component analysis plot (PC1 and PC2) of the five Tigray cattle populations and other Ethiopian cattle representing three cattle groups of African zebu (Ethiopian Boran), African sanga (Afar) and African zenga (Fogera and Horro). (B) Principal component analysis plot (PC1 and PC2) for the five Tigray cattle populations alone (Abergelle, Arado, Begait, Erob and Raya).

2016), *Mfsd2b* important in S1P transport activity (Kobayashi et al., 2018) essential for various cellular functions (Spiegel and Milstien, 2011; Cyster and Schwab, 2012), *ITGA5* involved in different inflammation and immune response functions such as PI3K–Akt signaling pathway, bacterial invasion of epithelial cells, phagosome and human papillomavirus infection (Wang et al., 2021). *NPPFR1* (Moulédous et al., 2010) and *HTR7* (Hedlund et al., 2003) are

important in body temperature regulation. Last but not least, *Kciv2* was reported to be associated with visual adaptation in a changing lighting condition environment (Hölter et al., 2012). Two significant GO terms of the cellular component (GO: 0005578~proteinaceous extracellular matrix and GO: 0044420~extracellular matrix component) were explicitly enriched in Erob cattle. Genes such as multimerin 2 (*MMRN2*),

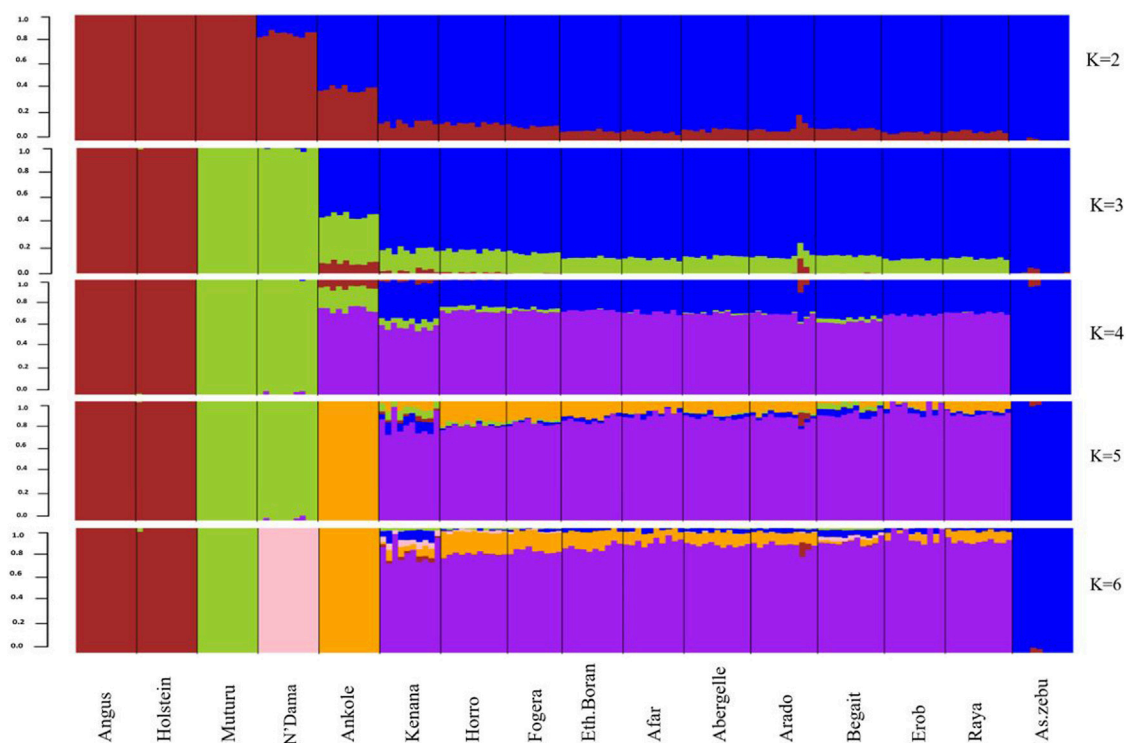


FIGURE 12

Admixture at $K = 2$ to $K = 6$ (the black lines separate the populations labelled below the figure).

von Willebrand factor C domain containing 2 (*VWC2*) and laminin subunit gamma 1 (*LAMC1*) were important in these GO terms. The *MMRN2* is associated with a meat quality trait called meat juiciness (Leal-Gutiérrez et al., 2019). Similarly, *VWC2* was considered as a candidate gene for intramuscular fat content, one of the most important meat quality traits in beef cattle (Halli et al., 2022). *VWC2* was also reported to be associated to feed efficiency in pigs (Wang et al., 2015). *LAMC1* was involved in different inflammation and immune response pathways, including prion diseases (bovine spongiform encephalopathy), amoebiasis and toxoplasmosis in cattle. Moreover, *LAMC1* was also shown to be relevant to temperature range in cattle (Flori et al., 2019).

Six molecular function GO terms relevant to intra- or extracellular activity were significantly enriched in several Tigray cattle populations (Abergelle, Arado, Begait and Raya), in which a few genes such as cadherin EGF LAG seven-pass G-type receptor 1 (*CELSR1*), gamma-aminobutyric acid type A receptor rho3 subunit (*GABRR3*), plexin A2 (*PLXNA2*) and toll-like receptor 6 (*TLR6*) were identified in Abergelle, Arado and Raya cattle, while the gene macrophage stimulating 1 receptor (*MST1R*) was overrepresented in the six GO terms (in Abergelle, Arado, Begait and Raya). *CELSR1* (Guo Y. et al., 2019) was found to be overexpressed following *in vitro* treatment of lipopolysaccharide, a cause of the endometrium inflammation (Sheldon et al., 2010), supporting its importance in immune response. In significantly enriched GO terms in Abergelle, Arado and Raya cattle, we found *TLR6* as an important candidate gene for bovine tuberculosis resistance (Song et al., 2014). Several

studies (Zhang et al., 2009; Seabury et al., 2010; Fisher et al., 2011; Elmaghraby et al., 2018; Maurić Maljković et al., 2023) have reported the importance of toll-like receptor genes for immunity, disease resistance and adaptive immune responses, including mastitis, the most economically important disease in dairy cattle (Elmaghraby et al., 2018; Maurić Maljković et al., 2023). Other genes relevant to oxidative stress (*MST1*) (Xiao et al., 2011), cattle temperament (*PLXNA2*) (Gutiérrez-Gil et al., 2008) and fertility such as sperm motility (*GABRR3*) (Hering et al., 2014) were also present in the six GO terms (Abergelle, Arado and Raya cattle).

In our previous morphological study of the same Tigray cattle populations (Zegeye et al., 2021), we showed that four of the five populations may be separated using morphological criteria. The exception was Erob and Abergelle, with a similar morphology. In particular, Begait cattle had the largest body size, a finding in agreement with the missense variants within genes linked to body height and length. Also, the presence of missense variants in genes involved in olfaction may be attributed to the adaptation of the Tigray cattle to the dry agro-ecology in the region, a characteristic of the Sudano-Saharan ecology with heat and water stress as an issue (Nyssen et al., 2009; Kumasi and Asenso-Okyere, 2011; Abraha, 2013). In addition, the regional landscape is mainly composed of mountains and hills (Kumasi and Asenso-Okyere, 2011) with limited grazing resources. As a result, the Tigray cattle are strongly adapted to feed shortage, as evidenced by the overrepresentation of genes associated with the olfactory and sensory perception of smell to differentiate the edible from non-edible or palatable from non-palatable browse plant species.

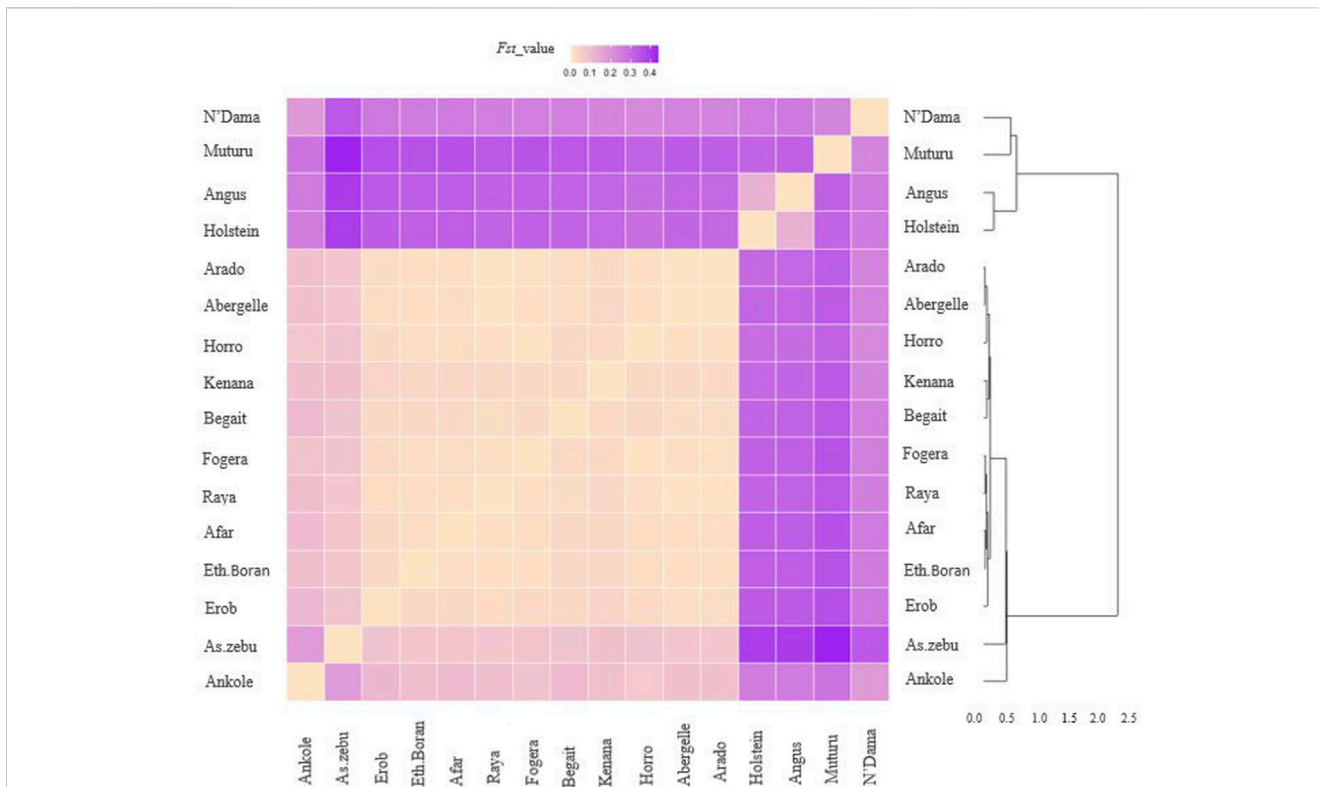


FIGURE 13
Heat map and dendrogram based on pairwise weighted F_{ST} values. The darker colour indicates higher pairwise population differentiation while lighter colour lower population differentiation.

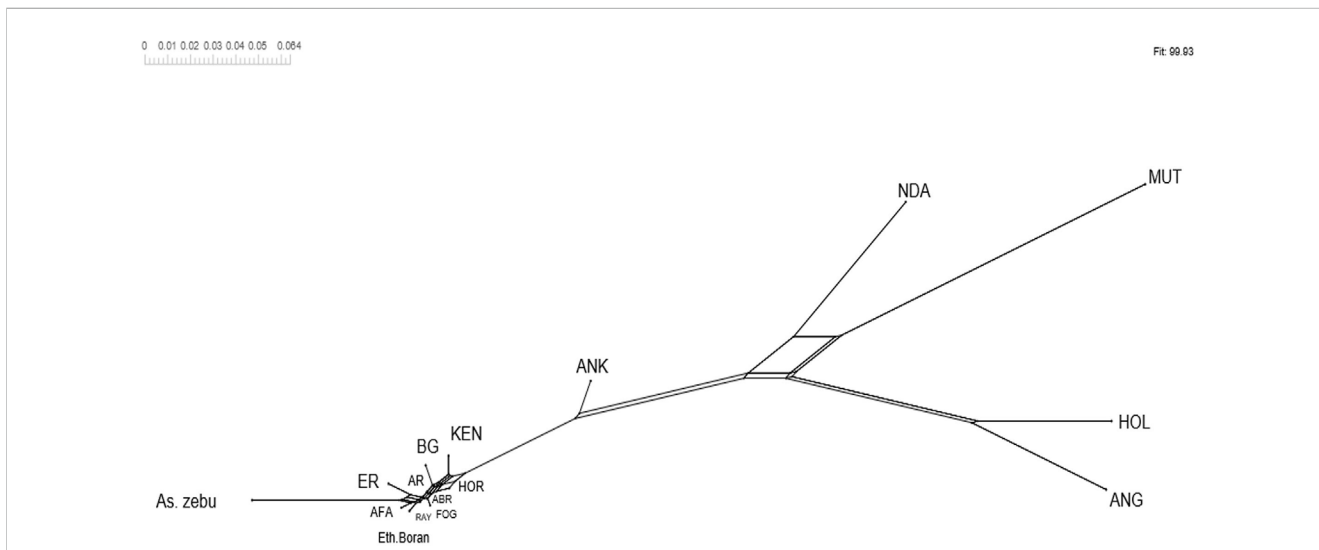


FIGURE 14
Neighbor-Net tree based on pairwise F_{ST} values.

High genetic diversity within and across the Tigray cattle populations

There is a high genome-wide nucleotide diversity (π) in all Tigray cattle, comparable with the values observed in Asian indicine cattle but higher than those in taurine cattle (Muturu, N'Dama,

Angus and Holstein) (Figure 7) and indicine-taurine admixed ($\pi = 2.9 \times 10^{-3}$ for Jiaxian Red) cattle (Xia et al., 2021). Similarly, the observed heterozygosity (H_o), an important indicator of genetic variability in domestic animals (Zhang et al., 2018), ranging from 0.278 to 0.302 among the Tigray cattle was similar with other indicine but higher than African and European taurine cattle.

Historic factors associated with the arrival and admixture of cattle in the Horn of Africa, including the Tigray region, likely shaped today's genome diversity of the Tigray cattle. As an ancient centre of civilisation, the Tigray region probably witnessed the early arrival of taurine cattle, followed by late introductions of indicine cattle in several migration waves, which continuously enriched the genomic landscape of the Tigray cattle. While our results indicated a large proportion of indicine background in Tigray cattle (around 90%), we still found a proportion of African taurine ancestry in their genomes. We may reasonably hypothesise that the rich genetic variation present in modern Tigray cattle is a legacy of multiple introductions, admixture and dispersion of cattle across the Horn of Africa.

We compared the ROH distribution pattern across the Tigray cattle populations and between the Tigray cattle and other reference cattle groups included in this study and found that all the Tigray cattle showed different patterns of ROH as compared to Asian zebu, African taurine (N'Dama and Muturu), African sanga (Ankole), African zebu (Kenana) and European-taurine (Angus and Holstein) cattle. As expected from their high genomic diversity, the number and cumulative length of ROH were smaller in the Tigray cattle compared with previous reports for taurine cattle (Purfield et al., 2012; Xia et al., 2021) and indicine-taurine admixed cattle outside Ethiopia (Xia et al., 2021). However, the number and length of ROH observed in the Tigray cattle were similar to the one reported for other Ethiopian breeds included in our study (Horro, Fogera, Borana, and Afar). It suggests common breeding history among Ethiopian cattle breeds, while PCA and admixture results suggest close genetic relationships among the Ethiopian cattle as recently showed in a genome analysis including 14 Ethiopian indigenous cattle breeds (Terefe et al., 2023).

Inbreeding coefficients were far lower in the Tigray cattle than those reported in other cattle breeds, particularly the Danish dairy cattle breeds (Zhang et al., 2015b), with an inbreeding coefficient five times higher at a population level. An inbreeding coefficient below 5%, as observed for nearly all animals in this study (F_{ROH}), is generally considered to have no consequence on an individual's fitness (Slate et al., 2004). Therefore, the level of inbreeding in the Tigray cattle is within an acceptable range to accommodate within-population improvement of their productivity.

Population genetic structure and relationship

Taurine ancestry was generally low in the Tigray cattle relative to other African humped cattle breeds (Kim et al., 2020). This is particularly expected for cattle populations geographically close to the entry points of Asian indicine cattle into Africa. The unique local ancestries observed in Erob ($K = 7$) or Begait ($K = 10$) and their introgression to all non-taurine African breeds (Supplementary Figure S10) could further confirm the probability of the Tigray region of Ethiopia as a gate of cattle to Africa. Moreover, we observed a closer relationship between Begait to Kenana cattle (a Sudanese cattle breed) than with other Tigray cattle populations. Begait cattle are typically found in the western Tigray regions close to the Sudanese border. Therefore, gene flow from Begait cattle to Sudanese cattle is possible or *vice versa*. Previously, we observed a close morphological relationship between Erob and Abergelle cattle (Zegeye et al., 2021). This result is not supported by our genetic relationship analysis with the two breeds

here clearly separated (e.g., F_{ST} -based dendrogram, Supplementary Figure S11). The two breeds are found at different altitudes (Figure 1B). Henceforth, the relationship between Erob and Abergelle cattle requires further investigation.

Conclusion

Overall, we provided a detailed analysis using whole genome sequencing data of the genetic diversity, relatedness and admixture of five cattle populations indigenous to the Tigray region, the northernmost state of Ethiopia and a major geographic region of ancient civilizations. We found around 36 M SNPs and 3.7 M indels, where around 7% and 34% of them were novel. The contribution of such novel variants increases the number of known cattle genomic variants and prompts our understanding of the genetic diversity of domestic cattle. We found a high within-population diversity based on the incidence, type and distribution of the genomic variants, genome-wide nucleotide diversity, heterozygosity, runs of homozygosity and genomic inbreeding coefficient. Besides, we detected a sign of poor management in a few Begait and Raya cattle having long ROH and strong inbreeding (>10%), possibly resulting from consanguineous mating. So, these two populations may need special attention to maintain their within-population genetic diversity. The admixture analysis confirmed that the Tigray cattle have a common main indicine ancestry, followed by a low African taurine and a rather limited European taurine ancestry. With high within-population genetic diversity, the Tigray cattle represent an important indigenous genetic resource for breeding improvement to enhance their productivity (e.g., milk), while maintaining their environmental adaptability. All the Tigray cattle populations shared highly significant GO and pathway terms associated with sensory perception of smell with overrepresented genes in the olfactory family, which may be relevant to their adaptation to their harsh environments.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI/SRA (<https://www.ncbi.nlm.nih.gov/sra>). Accession No. PRJNA882591.

Ethics statement

Ethical review and approval were not required for the animal study because blood samples were collected with good animal handling practice by a veterinary professional (DVM) during routine activities under the approval of the Ministry of Agriculture, Ethiopia. Hence, no further ethical permission was required for this study.

Author contributions

Conceptualisation, methodology, software, investigation, TZ and OH; validation, GB and OH; resources, HJ, GB, and OH;

facilitating sequence data generation, HJ; data-analysis, writing-original draft preparation, TZ; additional data analysis requested during the review process (VEP annotated private SNPs sorting and Neighbor-Net tree construction), AV-T; writing-review, editing and supervision, GB, HJ, and OH. All authors contributed to the article and approved the submitted version.

Funding

International Livestock Research Institute LiveGene program, supported by the CGIAR Research Program on Livestock (CRP livestock project) sponsored by the CGIAR funding contributors to the Trust Fund (<http://www.cgiar.org/about-us/our-funders/>), the Bill and Melinda Gates Foundation and UK aid from UK Foreign, Commonwealth and Development Office (Grant Agreement OPP1127286) under the auspices of the Centre for Tropical Livestock Genetics and Health (CTLGH), established jointly by the University of Edinburgh, SRUC (Scotland's Rural College), ILRI and the Chinese Government contribution to CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources in Beijing (2018-GJHZ-01).

Acknowledgments

The authors would like to acknowledge Adhanom Baraki (DVM) for his cooperation during blood sample collection,

References

- Abraha, M. E. (2013). *Assessment of drought early warning in Ethiopia: A comparison of wrsi by surface energy balance and soil water balance*. Twente.
- Addo, S., Klingel, S., Hinrichs, D., and Thaller, G. (2019). Runs of homozygosity and NetView analyses provide new insight into the genome-wide diversity and admixture of three German cattle breeds. *PLoS One* 14 (12), e0225847. doi:10.1371/journal.pone.0225847
- Ajmone-Marsan, P., Garcia, J. F., and Lenstra, J. A. (2010). On the origin of cattle: How aurochs became cattle and colonized the world. *Evol. Anthropol.* 19 (4), 148–157. doi:10.1002/evan.20267
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19 (9), 1655–1664. doi:10.1101/gr.094052.109
- Beja-Pereira, A., Caramelli, D., Lalueza-Fox, C., Vernesi, C., Ferrand, N., Casoli, A., et al. (2006). The origin of European cattle: Evidence from modern and ancient DNA. *Proc. Natl. Acad. Sci. U. S. A.* 103 (21), 8113–8118. doi:10.1073/pnas.0509210103
- Bekuma, A., and Hirpha, K. (2018). Productive and reproductive performance of indigenous cows under farmer's management system in and around walmera district, oromia regional state, Ethiopia. *Int. J. Adv. Res. Biol. Sci.* 5 (10), 65–69. doi:10.22192/ijarbs
- Bryant, D., and Moulton, V. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* 21 (2), 255–265. doi:10.1093/molbev/msh018
- Cai, Z., Gulbrandtsen, B., Lund, M. S., and Sahana, G. (2019). Weighting sequence variants based on their annotation increases the power of genome-wide association studies in dairy cattle. *Genet. Sel. Evol.* 51 (1), 20–12. doi:10.1186/s12711-019-0463-9
- CSA (2018). *Agricultural sample survey 2017/18[2010eC] volume II Report on livestock and livestock characteristics (private peasant holdings)*. Addis Ababa, Ethiopia: Statistical Bulletin.
- Cyster, J. G., and Schwab, S. R. (2012). Sphingosine-1-Phosphate and lymphocyte egress from lymphoid organs. *Annu. Rev. Immunol.* 30, 69–94. doi:10.1146/annurev-immunol-020711-075011
- 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. *Anim. Genet.* 39 (4), 425–431. doi:10.1111/j.1365-2052.2008.01748.x
- Abdulfatai Tijjani (PhD) for his assistance during data management and Endashaw Terefe (MSc) for assisting during genomic DNA extraction and sharing some scripts during the data analysis.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi:10.1093/bioinformatics/btr330
- Das, A., Panitz, F., Gregersen, V. R., Bendixen, C., and Holm, L. E. (2015). Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes. *BMC Genomics* 16 (1), 1043. doi:10.1186/s12864-015-2249-y
- de Lima, A. O., Koltjes, J. E., Diniz, W. J. S., de Oliveira, P. S. N., Cesar, A. S. M., Tizioto, P. C., et al. (2020). Potential biomarkers for feed efficiency-related traits in nelore cattle identified by Co-expression network and integrative genomics analyses. *Front. Genet.* 11, 189. doi:10.3389/fgene.2020.00189
- Easa, A. A., Selionova, M., Aibazov, M., Mamontova, T., Sermiyagin, A., Belous, A., et al. (2022). Identification of genomic regions and candidate genes associated with body weight and body conformation traits in karachai goats. *Genes (Basel)* 13 (10), 1773. doi:10.3390/genes13101773
- EBI (2016). *Ethiopian national strategy and plan of action for conservation and sustainable utilization of animal genetic resources*. Addis Ababa, Ethiopia: Ethiopian Biodiversity Institute.
- Edea, Z., Bhuiyan, M. S. A., Dessie, T., Rothschild, M. F., Dadi, H., and Kim, K. S. (2015). Genome-wide genetic diversity, population structure and admixture analysis in african and asian cattle breeds. *Animal* 17 (1), 218–226. doi:10.1017/S1751731114002560
- Edea, Z., Dadi, H., Kim, S. W., Dessie, T., Lee, T., Kim, H., et al. (2013). Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean hanwoo breeds using SNP markers. *Front. Genet.* 4, 35. doi:10.3389/fgene.2013.00035
- Elmaghraby, M. M., El-Nahas, A. F., Fathala, M. M., Sahwan, F. M., and Tag El-Dien, M. A. (2018). Association of toll-like receptors 2 and 6 polymorphism with clinical mastitis and production traits in Holstein cattle. *Iran. J. Vet. Res.* 19 (3), 202–207.
- Eusebi, P. G., Martinez, A., and Cortes, O. (2020). Genomic tools for effective conservation of livestock breed diversity. *Diversity* 12 (1), 8. doi:10.3390/d12010008
- FAO (2015). *The second report on the state of the world's animal genetic resources for food and agriculture*. Editors B. D. Scherf and D. Pilling (Rome, Italy: Commission on Genetic Resources for Food and Agriculture Assessments).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1050365/full#supplementary-material>

- Finneran, N. (2007). *The archaeology of Ethiopia*. Cornwall, Great Britain: Routledge.
- Fisher, C. A., Bhattarai, E. K., Osterstock, J. B., Dowd, S. E., Seabury, P. M., Vikram, M., et al. (2011). Evolution of the bovine TLR gene family and member associations with *Mycobacterium avium* subspecies paratuberculosis infection. *PLoS One* 6 (11), e27744. doi:10.1371/journal.pone.0027744
- Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., et al. (2019). A genomic map of climate adaptation in mediterranean cattle breeds. *Mol. Ecol.* 28 (5), 1009–1029. doi:10.1111/mec.15004
- Freeman, A. R., Bradley, D. G., Nagda, S., Gibson, J. P., and Hanotte, O. (2006). Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. *Anim. Genet.* 37 (1), 1–9. doi:10.1111/j.1365-2052.2005.01363.x
- Guo, J., Zhong, J., Li, L., Zhong, T., Wang, L., Song, T., et al. (2019a). Comparative genome analyses reveal the unique genetic composition and selection signals underlying the phenotypic characteristics of three Chinese domestic goat breeds. *Genet. Sel. Evol.* 51 (1), 70–18. doi:10.1186/s12711-019-0512-4
- Guo, Y., van Schaik, T., Jhamat, N., Niazi, A., Chanrot, M., Charpigny, G., et al. (2019b). Differential gene expression in bovine endometrial epithelial cells after challenge with LPS: specific implications for genes involved in embryo maternal interactions. *PLoS One* 14 (9), e0222081. doi:10.1371/journal.pone.0222081
- Gutiérrez-Gil, B., Ball, N., Burton, D., Haskell, M., Williams, J. L., and Wiener, P. (2008). Identification of quantitative trait loci affecting cattle temperament. *J. Hered.* 99 (6), 629–638. doi:10.1093/jhered/esn060
- Halli, K., Bohlouli, M., Schulz, L., Sundrum, A., and König, S. (2022). Estimation of direct and maternal genetic effects and annotation of potential candidate genes for weight and meat quality traits in a genotyped outdoor dual-purpose cattle breed. *Transl. Anim. Sci.* 6 (1), txac022. doi:10.1093/tas/txac022
- Hanotte, O., Bradley, D. G., Ochieng, J. W., Verjee, Y., Hill, E. W., and Rege, J. E. O. (2002). African pastoralism: Genetic imprints of origins and migrations. *Science* 296 (5566), 336–339. doi:10.1126/science.1069878
- Hedlund, P. B., Danielson, P. E., Thomas, E. A., Slanina, K., Carson, M. J., and Sutcliffe, J. G. (2003). No hypothalamic response to serotonin in 5-HT₁ receptor knockout mice. *Proc. Natl. Acad. Sci. U. S. A.* 100 (3), 1375–1380. doi:10.1073/pnas.0337340100
- Hering, D. M., Olenski, K., and Kaminski, S. (2014). Genome-wide association study for poor sperm motility in holstein-friesian bulls. *Anim. Reprod. Sci.* 146 (3–4), 89–97. doi:10.1016/j.anireprosci.2014.01.012
- Hölter, P., Kunst, S., Wolloscheck, T., Kelleher, D. K., Sticht, C., Wolfrum, U., et al. (2012). The retinal clock drives the expression of *Kcnv2*, a channel essential for visual function and cone survival. *Investig. Ophthalmol. Vis. Sci.* 53 (11), 6947–6954. doi:10.1167/iovs.12-10234
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23 (2), 254–267. doi:10.1093/molbev/msj030
- Iqbal, N., Liu, X., Yang, T., Huang, Z., Hanif, Q., Asif, M., et al. (2019). Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. *PLoS One* 14 (4), e0215065. doi:10.1371/journal.pone.0215065
- Jang, J., Terefe, E., Kim, K., Lee, Y. H., Belay, G., Tijjani, A., et al. (2022). Correction to: Population differentiated copy number variation of *Bos taurus*, *Bos indicus* and their african hybrids. *BMC Genomics* 23 (1), 207. doi:10.1186/s12864-022-08409-8
- Jiang, L., Liu, X., Yang, J., Wang, H., Jiang, J., Liu, L., et al. (2014). Targeted resequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genomics* 15 (1), 1105–1109. doi:10.1186/1471-2164-15-1105
- Keogh, K., Kenny, D. A., and Waters, S. M. (2019). Gene Co-expression networks contributing to the expression of compensatory growth in metabolically active tissues in cattle. *Sci. Rep.* 9 (1), 6093. doi:10.1038/s41598-019-42608-w
- Kern, R. J., Lindholm-Perry, A. K., Freetly, H. C., Snelling, W. M., Kern, J. W., Keele, J. W., et al. (2016). Transcriptome differences in the rumen of beef steers with variation in feed intake and gain. *Gene* 586 (1), 12–26. doi:10.1016/j.gene.2016.03.034
- Kim, J., Hanotte, O., Mwai, O. A., Dessie, T., Salim, B., Diallo, B., et al. (2017). The genome landscape of indigenous african cattle. *Genome Biol.* 18 (1), 34–14. doi:10.1186/s13059-017-1153-y
- Kim, K., Kwon, T., Dessie, T., Yoo, D. A., Mwai, O. A., Jang, J., et al. (2020). The mosaic genome of indigenous african cattle as a unique genetic resource for african pastoralism. *Nat. Genet.* 52 (10), 1099–1110. doi:10.1038/s41588-020-0694-2
- Kobayashi, N., Kawasaki-Nishi, S., Otsuka, M., Hisano, Y., Yamaguchi, A., and Nishi, T. (2018). MFSD2B is a sphingosine 1-phosphate transporter in erythroid cells. *Sci. Rep.* 8 (1), 4969. doi:10.1038/s41598-018-23300-x
- Kour, A., Niranjana, S. K., Jayakumar, S., Pukhrabam, M., and Sarkar, M. (2022). “517. Olfaction: A critical force driving adaptive evolution in yaks,” in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges* (Wageningen Academic), 2147–2150. doi:10.3920/978-90-8686-940-4_517
- Kukučková, V., Moravčíková, N., Ferenčáková, M., Simčíč, M., Mészáros, G., Sölkner, J., et al. (2017). Genomic characterization of pinzgau cattle: Genetic conservation and breeding perspectives. *Conserv. Genet.* 18 (4), 893–910. doi:10.1007/s10592-017-0935-9
- Kumasi, T. C., and Asenso-Okyere, K. (2011). Responding to land degradation in the highlands of Tigray, northern Ethiopia. IFPRI Discussion Paper 01142.
- Leal-Gutiérrez, J. D., Elzo, M. A., Johnson, D. D., Hamblen, H., and Mateescu, R. G. (2019). Genome wide association and gene enrichment analysis reveal membrane anchoring and structural proteins associated with meat quality in beef. *BMC Genomics* 20 (1), 151. doi:10.1186/s12864-019-5518-3
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352
- Li, M. H., Zerabruk, M., Vangen, O., Olsaker, I., and Kantanen, J. (2007). Reduced genetic structure of North Ethiopian cattle revealed by Y-chromosome analysis. *Hered. (Edinb.)* 98 (4), 214–221. doi:10.1038/sj.hdy.6800931
- Low, W. Y., Tearle, R., Liu, R., Koren, S., Rhie, A., Bickhart, D. M., et al. (2020). Haplotype-resolved genomes provide insights into structural variation and gene content in Angus and brahman cattle. *Nat. Commun.* 11 (1), 2071. doi:10.1038/s41467-020-15848-y
- Maruo, Y., Nagasaki, K., Matsui, K., Mimura, Y., Mori, A., Fukami, M., et al. (2016). Natural course of congenital hypothyroidism by dual oxidase 2 mutations from the neonatal period through puberty. *Eur. J. Endocrinol.* 174 (4), 453–463. doi:10.1530/EJE-15-0959
- Mauki, D. H., Tijjani, A., Ma, C., Ng’ang’a, S. I., Mark, A. I., Sanke, O. J., et al. (2022). Genome-wide investigations reveal the population structure and selection signatures of Nigerian cattle adaptation in the sub-saharan tropics. *BMC Genomics* 23 (1), 306–314. doi:10.1186/s12864-022-08512-w
- Maurić Maljković, M., Vlahek, I., Piplica, A., Ekert Kabalin, A., Sušić, V., and Stevanović, V. (2023). Prospects of toll-like receptors in dairy cattle breeding. *Anim. Genet.* 54, 425–434. doi:10.1111/age.13325
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110
- McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83 (3), 359–372. doi:10.1016/j.ajhg.2008.08.007
- Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C. E., Taberlet, P., Luntz, B., et al. (2009). Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Mol. Ecol.* 18 (16), 3394–3410. doi:10.1111/j.1365-294X.2009.04286.x
- MOA (1998). *Agroecological Zones of Ethiopia*. Addis Ababa, Ethiopia: MOA.
- Moulédous, L., Barthas, F., and Zajac, J. M. (2010). Opposite control of body temperature by NPFF1 and NPFF2 receptors in mice. *Neuropeptides* 44 (5), 453–456. doi:10.1016/j.npep.2010.05.003
- Mwai, O., Hanotte, O., Kwon, Y. J., and Cho, S. (2015). African indigenous cattle: Unique genetic resources in a rapidly changing world. *Asian-Australasian J. Anim. Sci.* 28 (7), 911–921. doi:10.5713/ajas.15.0002R
- Nyssen, J., Haile, M., Naudts, J., Munro, N., Poesen, J., Moeyersons, J., et al. (2009). Desertification? Northern Ethiopia Re-photographed after 140 years. *Sci. Total Environ.* 407 (8), 2749–2755. doi:10.1016/j.scitotenv.2008.12.016
- Pagani, L., Kivisild, T., Tarekgn, A., Ekong, R., Plaster, C., Gallego Romero, I., et al. (2012). Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* 91 (1), 83–96. doi:10.1016/j.ajhg.2012.05.015
- Pedrosa, V. B., Schenkel, F. S., Chen, S. Y., Oliveira, H. R., Casey, T. M., Melka, M. G., et al. (2021). Genomewide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. *Genes (Basel)*. 12 (11), 1830. doi:10.3390/genes12111830
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795
- Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13 (1), 70. doi:10.1186/1471-2156-13-70
- R Development Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rosen, B. D., Bickhart, D. M., Schnabel, R. D., Koren, S., Elisk, C. G., Tseng, E., et al. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. *Genetics* 9 (3), giaa021. doi:10.1093/genetics/giaa021
- Rowlands, J., Nieves, C., Hanotte, O., and Ayalew, W. (2006). “Cattle breed distributions across districts as determined from cluster analysis of phenotypic data collected in the oromiya region, Ethiopia,” in *8th world congress on genetics applied to livestock production*. Editors P. Leroy, E. Leroy, and R. Cassart (Brasil: Belo Horizonte), 18–33.

- Seabury, C. M., Seabury, P. M., Decker, J. E., Schnabel, R. D., Taylor, J. F., and Womack, J. E. (2010). Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. *Proc. Natl. Acad. Sci. U. S. A.* 107 (1), 151–156. doi:10.1073/pnas.0913006107
- Sheldon, I. M., Rycroft, A. N., Dogan, B., Craven, M., Bromfield, J. J., Chandler, A., et al. (2010). Specific strains of *Escherichia coli* are pathogenic for the endometrium of cattle and cause pelvic inflammatory disease in cattle and mice. *PLoS One* 5 (2), e9192. doi:10.1371/journal.pone.0009192
- Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol.* 26 (10), 1135–1145. doi:10.1038/nbt1486
- Slate, J., David, P., Dodds, K. G., Veenvliet, B. A., Glass, B. C., Broad, T. E., et al. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: Theoretical expectations and empirical data. *Hered. (Edinb.)* 93 (3), 255–265. doi:10.1038/sj.hdy.6800485
- Song, Y., Sun, L., Guo, A., and Yang, L. (2014). Toll-like receptor 6 gene polymorphisms increase the risk of bovine tuberculosis in Chinese Holstein cattle. *Acta histochem.* 116 (7), 1159–1162. doi:10.1016/j.acthis.2014.06.004
- Soria-Gómez, E., Bellocchio, L., Reguero, L., Lepousez, G., Martin, C., Bendahmane, M., et al. (2014). The endocannabinoid system controls food intake via olfactory processes. *Nat. Neurosci.* 17 (3), 407–415. doi:10.1038/nn.3647
- Spiegel, S., and Milstien, S. (2011). The outs and the ins of sphingosine-1-phosphate in immunity. *Nat. Rev. Immunol.* 11 (6), 403–415. doi:10.1038/nri2974
- Stafuzza, N. B., Zerlotini, A., Lobo, F. P., Yamagishi, M. E. B., Chud, T. C. S., Caetano, A. R., et al. (2017). Single nucleotide variants and InDels identified from whole-genome Re-sequencing of guzerat, gyr, girolando and Holstein cattle breeds. *PLoS One* 12 (3), e0173954. doi:10.1371/journal.pone.0173954
- Taussat, S., Boussaha, M., Ramayo-Caldas, Y., Martin, P., Venot, E., Cantalapiedra-Hijar, G., et al. (2020). Gene networks for three feed efficiency criteria reveal shared and specific biological processes. *Genet. Sel. Evol.* 52 (1), 67–14. doi:10.1186/s12711-020-00585-z
- Terefe, E., Belay, G., Han, J., Hanotte, O., and Tijjani, A. (2022). Genomic adaptation of Ethiopian indigenous cattle to high altitude. *Front. Genet.* 13, 960234. doi:10.3389/fgene.2022.960234
- Terefe, E., Belay, G., Tijjani, A., Han, J., and Hanotte, O. (2023). Whole genome resequencing reveals genetic diversity and selection signatures of Ethiopian indigenous cattle adapted to local environments. *Diversity* 15 (4), p540. doi:10.3390/d15040540
- Tijjani, A., Utsunomiya, Y. T., Ezekwe, A. G., Nashiru, O., and Hanotte, O. (2019). Genome sequence analysis reveals selection signatures in endangered trypanotolerant west african Muturu cattle. *Front. Genet.* 10, 442. doi:10.3389/fgene.2019.00442
- Wang, K., Liu, D., Hernandez-Sanchez, J., Chen, J., Liu, C., Wu, Z., et al. (2015). Genome wide association analysis reveals new production trait genes in a male duroc population. *PLoS One* 10 (9), e0139207. doi:10.1371/journal.pone.0139207
- Wang, X., Fan, Y., He, Y., Han, Z., Gong, Z., Peng, Y., et al. (2021). Integrative analysis of mirna and mrna expression profiles in mammary glands of Holstein cows artificially infected with *Staphylococcus aureus*. *Pathogens* 10 (5), 506. doi:10.3390/pathogens10050506
- Weir, B. S., and Clark Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. *Evol. (N. Y.)* 38 (6), 1358–1370. doi:10.1111/j.1558-5646.1984.tb05657.x
- Woldekiros, H. S., and D'Andrea, A. C. (2017). Early evidence for domestic chickens (*Gallus gallus domesticus*) in the Horn of Africa: Early evidence for domestic chickens. *Int. J. Osteoarchaeol.* 27 (3), 329–341. doi:10.1002/oa.2540
- Xia, X., Zhang, S., Zhang, H., Zhang, Z., Chen, N., Li, Z., et al. (2021). Assessing genomic diversity and signatures of selection in jiaxian red cattle using whole-genome sequencing data. *BMC Genomics* 22 (1), 43–11. doi:10.1186/s12864-020-07340-0
- Xiao, L., Chen, D., Hu, P., Wu, J., Liu, W., Zhao, Y., et al. (2011). The C-Abl-MST1 signaling pathway mediates oxidative stress-induced neuronal cell death. *J. Neurosci.* 31 (26), 9611–9619. doi:10.1523/JNEUROSCI.0035-11.2011
- Zegeye, T., Belay, G., and Hanotte, O. (2021). Multivariate characterization of phenotypic traits of five native cattle populations from Tigray, northern Ethiopia. *Trop. Anim. Health Prod.* 53 (2), 212–215. doi:10.1007/s11250-021-02652-z
- Zerabruk, M., Bennewitz, J., Kantanen, J., Olsaker, I., and Vangen, O. (2007). Analysis of genetic diversity and conservation priorities for six North Ethiopian cattle breeds. *J. Anim. Breed. Genet.* 124 (4), 236–241. doi:10.1111/j.1439-0388.2007.00660.x
- Zerabruk, M., Li, M. H., Kantanen, J., Olsaker, I., Ibeagha-Awemu, E. M., Erhardt, G., et al. (2011). Genetic diversity and admixture of indigenous cattle from North Ethiopia: Implications of historical introgressions in the gateway region to Africa. *Anim. Genet.* 43 (3), 257–266. doi:10.1111/j.1365-2052.2011.02245.x
- Zhang, C. L., Yuan, J., Wang, Q., Wang, Y. H., Fang, X. T., Lei, C. Z., et al. (2012). The association of bovine T1R family of receptors polymorphisms with cattle growth traits. *Res. Vet. Sci.* 93 (2), 783–787. doi:10.1016/j.rvsc.2012.01.014
- Zhang, F., Qu, K., Chen, N., Hanif, Q., Jia, Y., Huang, Y., et al. (2019). Genome-wide SNPs and indels characteristics of three Chinese cattle breeds. *Animals* 9 (9), 596. doi:10.3390/ani9090596
- Zhang, M., Peng, W. F., Hu, X. J., Zhao, Y. X., Lv, F. H., and Yang, J. (2018). Global genomic diversity and conservation priorities for domestic animals are associated with the economies of their regions of origin. *Sci. Rep.* 8 (1), 11677. doi:10.1038/s41598-018-30061-0
- Zhang, Q., Calus, M. P. L., Guldbandsen, B., Lund, M. S., and Sahana, G. (2015a). Estimation of inbreeding using pedigree, 50k SNP Chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16 (1), 88–11. doi:10.1186/s12863-015-0227-7
- Zhang, Q., Guldbandsen, B., Bosse, M., Lund, M. S., and Sahana, G. (2015b). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16 (1), 542. doi:10.1186/s12864-015-1715-x
- Zhang, Y. B., He, F. L., Fang, M., Hua, T. F., Hu, B. D., Zhang, Z. H., et al. (2009). Increased expression of toll-like receptors 4 and 9 in human lung cancer. *Mol. Biol. Rep.* 36 (6), 1475–1481. doi:10.1007/s11033-008-9338-9