Exploring the genomes of East African Indicine cattle breeds reveals signature of selection for tropical environmental adaptation traits

Mengistie Taye¹,²*, Wonseok Lee⁴, Kelsey Caetano-Anolles¹, Tadelle Dessie³, Seoae Cho⁴, Sung Jong Oh⁵, Hak-Kyo Lee⁶ and Heebal Kim¹,⁴,⁷

Abstract: African indigenous cattle breeds have been reared within the continent for millennia. Due to the harsh tropical environmental conditions (e.g. severe disease and parasite prevalence, high temperature, feed and water scarcity) where they evolved, they have developed various levels of tropical environmental adaptation attributes. In order to explore the genomic signatures of tropical environmental adaptation in African cattle, we compared the whole genomes of East African Indicus cattle breeds with European and Asian Taurine cattle breeds using XP-EHH and XP-CLR population statistical methods. Several genes involved in various biological processes and pathways related to domestication and behavior (dopaminergic and glutamatergic synapses), feeding and metabolism (gastric acid secretion, metabolic pathways), thermotolerance (sphingolipid and Wnt signaling), immune system response (T cell receptor signaling), and growth and reproduction.

ABOUT THE AUTHORS
Mengistie Taye The first (corresponding) author is a PhD holder specialized in Agriculture Biotechnology. His PhD research was focusing on bioinformatics analysis of the genomes of animals for adaptive signature of natural and artificial selection in cattle. He has published his findings in different peer reviewed international journals. Analysis of signature of selection is a genotype to phenotype approach and helps to identify and understand whether the phenotypic traits manifested are due environmental acclimatization or genetic effects. Nowadays, there are a number of research projects working on the analysis of signature of selection for different adaptation and production traits in different species of animals. This manuscript, I believe, will add a lot to the knowledge base of population genetics and animal breeding in developing countries where adaptation traits are very important.

PUBLIC INTEREST STATEMENT
Cattle are an important livestock species that provide food (milk, meat, blood), fiber (hide), power (as a draft animal), and security for the smallholder poor farm households in developing countries including Africa. African cattle are known for their adaptation to tropical environmental conditions like thermotolerance, resistance to tropical diseases and have the ability to thrive under poor quality feed and water conditions. This is because, they are evolved under tropical environmental conditions of high year-round temperature, poor quality and availability of feed and water, high prevalence of infectious and vector-borne diseases. Characterizing the genome of adaptive indigenous cattle populations and identifying genomic regions contributing to tropical environment adaptation is an important step in developing a breeding strategy for increased meat and milk production.

Accordingly, in this study, we analyzed the genome of African Indicus cattle breeds and identified genomic regions that are associated with tropical environmental adaptation traits including light and shiny coat color, better sweating ability, higher feeding efficiency, better reproductive performances, and superior disease and parasite resistance.
(osteoblast differentiation, fibroblast migration) were identified from our genome analysis. Genes associated with tick and parasite resistance traits such as keratin genes, collagen genes, calcium signaling, and tumor necrosis factor proteins were also identified. The genes and pathways identified in this study improve our understanding of the biological mechanisms of tropical environment adaptation of African Indicus cattle breeds, which may allow us to use them for genomic selection programs. This result presents a basis for further study and may help to develop vaccines for tick and gastrointestinal parasite challenge.

**Subjects:** Agriculture and Food; Biodiversity; Bioinformatics; Biotechnology

**Keywords:** Indicus cattle; positive selection signature; thermotolerance; tick resistance; XP-CLR; XP-EHH

1. Introduction

Sub-Saharan Africa is endowed with diverse (>145 cattle breeds) and a huge number of cattle populations (Rege, 1999). The genetic makeup of African cattle breeds is comprised primarily of genetic information originating from two subspecies: the humpless *B. taurus* (taurine) and the humped *B. indicus* (indicine). The contribution of the North African subspecies of wild cattle or aurochs *B. primigenius* to the genome of some African cattle populations is a possibility as well (Decker et al., 2014). Taurine cattle, known to have been domesticated in the near east and introduced to Africa, are the earliest African cattle which are, nowadays, inhabiting mainly the humid and sub-humid areas in West and Central Africa. Indicine cattle, domesticated in the Indus valley, were introduced to Africa later most likely after 700 AD, during Arabic migrations into north and east Africa (Ajmone Marsan, Garcia, & Lenstra, 2010; Hanotte et al., 2002; Rege, 1999). These days African cattle breeds are the result of human and natural selection forces, and introgression between the subspecies has shaped their genetic diversity resulting in breeds of diverse morphological and physiological characteristics, as well as adaptation to different agro-ecological and sociocultural conditions (Bahbahani, Afana, & Wragg, 2018; Bahbahani et al., 2015, 2018; Hanotte, Dessie, & Kemp, 2010; Mwai, Hanotte, Kwon, & Cho, 2015; Rege, 1999).

Environmental adaptation attributes including disease and parasite resistance as well as heat tolerance are important traits for cattle survival and production in tropical and subtropical regions (Neto, Jonsson, Michael, & Barendse, 2011; Shyma, Gupta, & Singh, 2015). High prevalence of disease and parasites exert a significant effect on cattle production in Africa causing death and production losses (Mwai et al., 2015). External parasites such as ticks, which feed on blood and transmit various types of diseases impact livestock production and productivity. The estimated economic cost of tick infestation control is high (Mapholi et al., 2014; Neto et al., 2011). Similarly, high environmental temperature compromises animal production and productivity through its effect on growth, reproduction, and feed intake (Hansen, 2004). Moreover, the stringency of available environmental resources in tropical regions make the introduction of productive European breeds difficult (Wang et al., 2007).

Humped zebu cattle are Indicus cattle which are abundantly distributed throughout East Africa and the drier parts of West Africa. These cattle breeds are easily identified by their characteristic hump and pendulous dewlap and are thought to have descended from the secondary cattle introduction that occurred in the arid regions of Africa. Zebu cattle are known for their superior adaptation to harsh environments, including resistance to various types of diseases and parasites as well as thermotolerance (Bahbahani et al., 2015, 2018; Chan, Nagaraj, & Reverter, 2010; Kongsuwan et al., 2010; Rege, 1999).
Characterizing the genome of indigenous cattle populations and identifying genomic regions contributing to tropical environment adaptation is an important step in developing a breeding strategy for increased meat and milk production (Bahbahani et al., 2015, 2018; Chan et al., 2010). Exploring signatures of positive selection has been used to identify genomic regions affected by natural and artificial selection that are associated with the phenotypes of specific populations. Previously, several authors have reported the positive selection of genes associated with tick resistance and other environmental adaptation traits in tropical cattle breeds. For example, Chan et al. reported genes directly and/or indirectly associated with tropical adaptation attributes such as keratins, heat shock proteins, and heat tolerance genes (Chan et al., 2010). By comparing the differential expression of genes in resistant and susceptible cattle, several candidate genes responsible for tick resistance were identified (Kongsuwan et al., 2016; Piper et al., 2008; Wang et al., 2007). The major histocompatibility complex genes have been the center of focus for genetic variation of disease resistance studies in cattle (Martinez et al., 2006; Neto et al., 2011). Gautier et al. reported genes involved in the immune system, nervous system, and skin and hair properties in West African cattle breeds (Gautier et al., 2009). Several genes involved in various biological pathways such as immunity, reproduction, development and heat tolerance have also been reported for East African Shorthorn Zebu cattle (Bahbahani et al., 2015, 2017) and Ethiopian Sheko cattle (Bahbahani et al., 2018). However, studies on East African cattle breeds based on whole genome next-generation sequencing (NGS) data are very scarce.

Several studies reported the positive selection signature of genes in various cattle breeds using different methods (Bahbahani et al., 2018, 2015, 2017; Gautier et al., 2009; Lee et al., 2014; Noyes et al., 2011). Here, we used a cross-population extended haplotype homozygosity (XP-EEH) and cross-population composite likelihood ratio (XP-CLR) statistical methods. The XP-EEH statistic assesses haplotype differences between two populations. It is designed to detect alleles that have increased in frequency to the point of fixation or near fixation in one of the populations being compared (Pickrell et al., 2009; Sabeti et al., 2007). XP-CLR is a likelihood method for detecting selective sweeps that involve jointly modeling the multilocus allele frequency differentiation between two populations (Chen, Patterson, & Reich, 2010). Using these two statistics, we compared the whole genomes of three East African Indicine cattle breeds with four (Three European and one Asian) Taurine cattle breeds in order to identify positively selected genomic regions in African cattle breeds that are associated with tropical environmental adaptation traits.

2. Materials and methods

2.1. Data description and preparation

This study is based on previously published data; detailed information about sampling and re-sequencing of DNA samples can be found within the manuscript (Kim et al., 2017). DNA samples extracted from whole blood samples of three East African Indicus cattle breeds (9 Boran, 9 Ogaden, and 10 Kenana) were individually sequenced using the Illumina HiSeq 2000 platform. Additionally, previously published genomic data of three European (10 Holstein, 10 Angus, and 10 Jersey) and a single Asian (11 Hanwoo) Taurine cattle breeds was used. Together, a per-base sequence quality check was performed using the FastQC software (Andrews, 2010). Bowtie2 (Langmead & Salzberg, 2012) was used to map pair-end sequence reads to the reference bovine genome (UMD 3.1). There was an overall alignment rate of 98.5% and average read depth of 10.8x. Picard tools (http://picard.sourceforge.net) to filter potential PCR duplicates, and SAMTools was used to create index files for reference and bam files (Li et al., 2009). Genome analysis toolkit 1.4 (GATK) performed local realignment of reads (McKenna et al., 2010). The “UnifiedGenotyper” and “SelectVariants” arguments of GATK were used to call candidate SNPs. In order to filter variants and avoid possible false positives, the “VariantFiltration” argument of the same software was adopted using the following options: 1) SNPs with a phred-scaled quality score of less than 30 were filtered; 2) SNPs with MQ0 (mapping quality zero; total count across all samples of mapping quality zero reads) >4 and quality depth (unfiltered depth of non-reference samples; low scores are indicative of false positives and artifacts) <5 were filtered; and 3) SNPs with FS (Phred-scaled p-value using Fisher’s exact test) > 200 were filtered since FS represents variation on either the forward or the reverse strand, which is indicative of false positive calls. BEAGLE (Browning & Browning,
2007) was used to infer the haplotype phase and impute missing alleles for the entire set of cattle populations simultaneously. After performing all the filtering processes, a total of ~37 million SNPs were retained and used for further analysis in this study. The sequences used in this study are available from GenBank with the Bioproject accession number of African cattle breeds (PRJNA312138—SRX1858079, SRX1858078, SRX1858077, SRX1858076, SRX1858075, SRX1858074, SRX1858073, SRX1858072, SRX1858071, SRX1802254, SRX1802253, SRX1802251, SRX1775710, SRX1775700, SRX1766099, SRX1766098, SRX1766097, SRX1766096, SRX1765344, SRX1765338, SRX1762810, SRX1762809, SRX1752210, SRX1751992, SRX1640149, SRX1631639, SRX1625858, SRX1765311), Holstein (PRJNA210521—SRX322362, SRX322361, SRX322360, SRX322359, SRX322358, SRX322357, SRX322356, SRX322355, SRX322354, SRX322353), Jersey (PRJNA318089—SRX1756427, SRX1756425, SRX1756424, SRX1756311, SRX1756308, SRX1756302, SRX1756292, SRX1756266, SRX1756143, SRX1756073), Angus (PRJNA318087—SRX1762572, SRX1762571, SRX1762569, SRX1762564, SRX1762377, SRX1762017, SRX1762016, SRX1758657, SRX1757535, SRX1756016), and Hanwoo (PRJNA210523—SRX322347, SRX322352, SRX322351, SRX322350, SRX322349, SRX322348, SRX322346, SRX322345, SRX322344, SRX318521, SRX318497).

2.2. Phylogenetic tree and population structure

In order to infer the evolutionary relationship between the breeds considered, we constructed a phylogenetic tree using genome-wide autosomal SNPs from 69 individual sequences. We used the SNPhylo pipeline (Lee, Guo, Wang, Kim, & Paterson, 2014) to construct the phylogenetic tree with a sub sample of 17,243 SNPs and options of Minor allele frequency (MAF) >0.05 and number of bootstrap samples of 1000. FigTree (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the tree. Additionally, we used STRUCTURE software which implements Bayesian algorithms in order to detect the true number of clusters, K (the number of ancestral populations), to identify groups of individuals corresponding to the uppermost hierarchical levels (Evanno, Regnaut, & Goudet, 2005). A total number of 16,181 loci were used for the structure analysis, using the following options of Length of Burnin Period of 2000, Number of MCMC Reps after Burnin of 100,000 and MAF of 0.05. We used VCFtools (Price et al., 2006) and PLINK (Purcell et al., 2007) with thin option (0.000675) to prepare input data used by STRUCTURE.

2.3. Detection of signature of positive selection

In order to detect the positive selective sweep regions in East African Indicus cattle, the whole genomes of three East African Indicus cattle breeds—Boran (Kenya), Ogaden (Ethiopia), and Kenana (Sudan)—grouped together as a test population were compared with four (three European and one Asian) Taurine cattle breeds—Holstein, Angus, Jersey and Hanwoo—grouped together as a reference population. East African Indicus cattle breeds are the result of the second introduction of Zebu cattle to Africa and have been bred for millennia under the harsh African conditions. Therefore, comparing the genomes of these breeds with European and Asian breeds and detecting recent positive selection signatures might give an insight into the selective forces shaping the genomes that help the African cattle adapting to the harsh environmental conditions.

We used the XP-EHH statistical method to assess haplotype differences between the test and reference populations (Sabeti et al., 2007). XP-EHH compares haplotype lengths between populations to control for local variation in recombination rates. It detects alleles that have increased in frequency to the point of fixation or near fixation in one of the populations (Pickrell et al., 2009). XP-EHH is directional, meaning that the direction of selection is determined by the sign of the values; positive values indicate selective sweep in the test populations, whereas a negative value indicates selection in the reference populations. We followed previously published procedures (Kim, Cho, Caetano-Anolles, Kim, & Ryu, 2015; Pickrell et al., 2009). XP-EHH values were calculated using the XP-EHH software, which can be accessed through the following link: (http://hgdp.uchicago.edu/Software/). In order to compare the genomic regions across populations, the genome was split into non-overlapping segments of 50 kb and the maximum XP-EHH value was computed for each segment. In order to define the empirical P-value, segments were grouped into three clusters according to the number of SNPs.
In each cluster, the regions with P-values less than 0.01 (1%) were considered as strong signals of selection in the African cattle populations.

Additionally, we performed an XP-CLR statistical test in order to identify potential regions differentially selected between the two populations compared (Chen et al., 2010). XP-CLR is a likelihood method for detecting selective sweep regions based on allele frequency differentiation. The script available at (http://genepath.med.harvard.edu/reich) was used to calculate XP-CLR scores. Non-overlapping sliding windows of 50 kb and a maximum number of 600 SNPs within each window were used. The correlation level from which the SNPs contribution to XP-CLR result was down-weighted to 0.95. The top 1% (0.01) of the empirical distributions were designated as candidate sweeps and genes that span the window regions were defined as candidate genes (Kim et al., 2015). Significant genomic regions identified from the XP-EHH and XP-CLR tests were annotated based on UMD 3.1.

### 2.4. Characterization of candidate genes under selection

The genes identified from XP-EHH and XP-CLR statistics were submitted to the Database for Annotation, Visualization, and Integrated Discovery (DAVID) gene ontology and annotation tool for gene enrichment analysis (Huang, Sherman, & Lempicki, 2009). Gene Ontology (GO) Biological Process (BP) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of the DAVID tool were used to cluster the genes of similar biological functions and pathways. All genes identified from XP-EHH and XP-CLR were submitted together.

We used SNPEff, a genetic variants (amino acid changes) annotation and effect prediction tool, for the candidate genes (Cingolani et al., 2012). Missense variants of candidate genes were extracted and carried on for the logistic and association analysis. By employing the Chi-squared test (plink—file mydata—assoc option) and logistic model (plink—file mydata—logistic option) employed in PLINK V1.07 software, significant SNPs that are specific for African Indicus cattle were identified and highly significant variants are reported.

The gene names and descriptions used in this manuscript are based on genecards (http://www.genecards.org/). The Manhattan plots of the −log_{10} transformed XP-EHH values and XP-CLR scores were drawn using R software.

### 3. Result and discussion

#### 3.1. Phylogenetic tree and population structure

In order to assess the historical relationship among the cattle breeds considered, we constructed a non-rooted phylogenetic tree of 69 individual cattle using autosomal SNPs. As expected, the phylogenetic tree revealed that African Indicus cattle breeds are clustered together separately from European and Asian Taurine cattle breeds (Figure 1a). Additionally, we performed population structure (Evanno et al., 2005) at different population assumptions, K (Figure 1b; Additional file 2: Figure S1). When K is 2, consistent with the phylogenetic tree, African Indicine and European and Asian Taurine cattle breeds showed clear differences, but at K 3, African Indicus breeds showed an admixture of each other. At K ≥ 4 population assumptions, except Jersey cattle, European and Asian Taurine cattle breeds showed some admixture level. In general, our result is consistent with previous reports which indicated that Indicine cattle are divergent from Taurine cattle breeds (Decker et al., 2014).

#### 3.2. Signature of positive selection in African Indicus cattle populations

In order to detect positive selection sweep regions in East African Indicus cattle populations, we used XP-EHH (Pickrell et al., 2009; Sabeti et al., 2007) and XP-CLR (Chen et al., 2010) population statistical methods. We compared the genomes of three East African Indicus cattle breeds (grouped together as a test population) with the genomes of four (three European and one Asian) Taurine cattle breeds (grouped together as a reference population). The Manhattan plot of the −log_{10} transformed XP-CLR and XP-EHH P-values is presented in Figure 2. From the signature of selection...
Figure 1. Population structure and relationship of cattle breeds considered. (a) Phylogenetic tree (Green Branch, Green Node—Kenana; Green Branch, Orange Node—Boran; Green Branch, Red Node—Ogaden; Red Branch, Blue Node—Hanwoo; Red Branch, Green Node—Angus; Red Branch, Cyan Node—Holstein; Red Branch, Rose Node, (b) Population structure at 2, 3, 4 and 7 population assumptions, (c) Number of genes identified from XP-CLR and XP-EHH statistics.

Figure 2. Manhattan plot of the $-\log_{10}$ transformed XP-EHH (a), and XP-CLR (b) p-values. The y-axis shows the $-\log_{10}$ (P-value) of XP-EHH and XP-CLR p-value, and x-axis shows chromosomal positions. The horizontal dotted lines represent the 1% outlier regions in both of the statistical methods.
regions were identified under selection. Annotation of these outlier regions resulted in a total of 765 (XP-EHH = 338; and XP-CLR = 499) genes (Figure 1c; Additional file 1: Table S1-S2).

Next, we performed DAVID gene enrichment analysis to identify enriched GO BP terms and KEGG pathways (Huang et al., 2009) using all the genes (765) identified from both statistics, and 15 significant ($p < 0.05$) BP terms and 19 KEGG pathways were found enriched. We clustered significantly enriched BP terms based on genes involved in the terms (Figure 3). These BP terms are involved in various biological functions including immune response (NFAT signaling cascade), growth and development (cell proliferation, cell shape, myoblast differentiation, neural tube closure, canonical Wnt signaling pathway), and behavior (walking behavior).

Enriched KEGG pathways are presented in Figure 4 (see Additional file 2: Table S3 for genes in the pathways). Dopaminergic synapse was among the KEGG pathways enriched. Dopamine is an organic chemical of the catecholamine and phenethylamine families that mediates a wide range of brain functions. It is involved in the domestication process of animals (Nikulina, 1990). Glutamatergic synapse pathway plays an important role in the behavioral adaptation of stress and fear responses (Kamprath et al., 2009). These pathways might also be related to domestication, which affects tameness and fear response in domestic animals (Mirkina et al., 2010; Zeder, 2012). Dopamine and glutamate modulate foraging and feeding-related behaviors of animals (Hills, 2006). The dopaminergic system aids free-ranging animals in identifying grass/forage species that are nutritious and non-poisonous (Berthoud, 2007; Jensen, 2002). The pastoral mode of livestock production in Africa might contribute for the positive selection of the gene regions involved in these pathways. African pastoralists move from place to place with their cattle in search of feed and water, which forces cattle to adapt to new vegetation and feed types (Hanotte et al., 2002; Mwai et al., 2015).

The KEGG pathways of gastric acid secretion, and fructose and mannose metabolism enriched are related to feeding digestion and metabolism. Gastric acid is vital for normal upper gastrointestinal functions, including protein digestion, and calcium and iron absorption, as well as...
providing some protection against bacterial infections. The physical and chemical nature of livestock feed affects gastric function to which coarse textured feeds result in higher gastric acid secretion (Low, 1990). The poor nutritional quality and seasonal availability of livestock feed in tropical Africa can act as a selective pressure which contributes to the adaptation and efficient utilization of low-quality feed by indigenous African cattle breeds. Lower metabolic requirements or reducing metabolism and higher digestive efficiency, and ability to utilize poor quality feed are adaptations of tropical cattle to feed scarcity (Mirkena et al., 2010). The higher efficiency of Indicus cattle to ferment nitrogen-deficient poor-quality feed and capture energy has been reviewed (Hegarty, 2004; Warwick & Cobb, 1975).

The mammalian heat shock response involves the sphingolipid signaling pathway, which activates sphingomyelin hydrolysis and/or the de novo biosynthesis of sphingolipids (Jenkins, 2003). The sphingolipid signaling pathway has two main metabolites: the ceramide and sphingosine-1-phosphate, which are important for heat stress response. Ceramides signals cells to undergo apoptosis during a severe heat stress (Jenkins, 2003). Sphingosine-1-phosphate has been found to protect oocytes from physiologically relevant heat shock and affect oocyte maturation (Roth & Hansen, 2004). Wnt signaling is also required for sweat gland development and sweating (Cui et al., 2014). Evaporative cooling through the involvement of sweat glands and other skin components is an important mechanism of heat tolerance (Jian, Duangjinda, Vajrabukka, & Katawatin, 2014). The genes in these pathways (PLCB1, MAPK12, and SGPL1) have been found to be involved in different aspects of thermotolerance. PLCB1 is associated with heat tolerance in catfish (Jin et al., 2017) and has been taken as an adaptation to hot arid environments in sheep and goats (Kim et al., 2016). Mitogen-activated protein kinases (MAPKs) elicit cellular response favoring survival or apoptosis (Sugimoto et al., 2012). Renin secretion, enriched in the KEGG pathways, is associated with adaptation to heat stress (Ali et al., 2012). It regulates the body’s water balance through the secretion of aldosterone and animals adapted to hot arid environments display lower overall water intake and turnover (Mirkena et al., 2010; Warwick & Cobb, 1975). Additionally, heat shock protein families (DNAJC11, DNAJC8), heat shock factor proteins (HSF5) and genes associated with heat stress response (PPP2R5E) identified in this study might contribute to the superior heat tolerance ability of African Indicus cattle breeds (Bahbahani et al., 2018, 2015, 2018, 2017; Wang, Dzama, Rees, & Muchadeyi, 2015). The BP term temperature homeostasis, a strategy by the animal to cope with climate variability, was also enriched.
Renin is found in high concentrations in reproductive tissues indicating its role in reproduction function. It affects ovulation, angiogenesis, and steroidogenesis (Yoshimura, 1997). Hedgehog signaling is an important pathway for proper development of the reproductive system (Franco & Yao, 2012). It is implicated in male germ cell development and differentiation in chicken (Chen et al., 2017). In addition to pathways, genes involved in reproduction function such as IGF-1, ESR2, and FGFR2 were identified (Table 1). IGF-1 affects follicular development and oocyte maturation and also enhances ovarian angiogenesis II production (Yoshimura, 1997). ESR2 is a protein-coding gene that controls many cellular processes including growth, differentiation, and function of the reproductive system. A polymorphism in ESR2 gene has been shown to be associated with sperm quality and boar fertility traits (Gunawan et al., 2012). FGFR2 is among the mitogenic signaling molecules with an essential role in the regulation of embryonic development. It is involved in postnatal development and function of the uterus (Filant, DeMayo, Pru, Lydon, & Spencer, 2013). The significant effect of high temperature on reproductive function of livestock species has been previously observed (Hansen, 2009). The pathways and genes identified here might contribute to the better reproductive efficiency of African Indicus cattle breeds to produce and reproduce in the tropical harsh environmental conditions of the continent (Hansen, 2009; Makina et al., 2015).

The T cell receptor signaling pathway, an important pathway for immune system development, was identified in this study. T cell receptors play a key role in the functioning of T cells and formation of an immunological synapse between T cells and the antigen-presenting cells. They initiate activation of T cells which in turn causes clonal expansion, differentiation, cytotoxic killing, or induction of programmed cell death (Bromley et al., 2001). The positive selection of genes involved in gene ontologies and pathways related to immune response have been previously reported for African cattle (Makina et al., 2015; Wang et al., 2015). Under tropical environments, indigenous African cattle breeds tend to better resist disease and parasite challenge as compared to European cattle breeds (Wang et al., 2015). The high and diverse disease and parasite prevalence in tropical Africa (Mirkena et al., 2010) where these breeds

### Table 1. Candidate genes identified as positively selected in African cattle breeds affecting different traits in relation to tropical environment adaptation

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th>Gene position (kbp)</th>
<th>XP-CLR score</th>
<th>XP-EHH value</th>
<th>XP-EHH p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLPH</td>
<td>3</td>
<td>117600.0–117650.0</td>
<td>221.32</td>
<td>1.30</td>
<td>0.0042</td>
</tr>
<tr>
<td>IGF-1</td>
<td>5</td>
<td>66525.6–66575.6</td>
<td>120.11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SLC25A48</td>
<td>7</td>
<td>48850.0–48900.0</td>
<td>-</td>
<td>2.25</td>
<td>0.0001</td>
</tr>
<tr>
<td>TGM1</td>
<td>10</td>
<td>20750.0–20800.0</td>
<td>-</td>
<td>1.45</td>
<td>0.0052</td>
</tr>
<tr>
<td>ESR2</td>
<td>10</td>
<td>76676.2–76726.2</td>
<td>187.69</td>
<td>1.19</td>
<td>0.0070</td>
</tr>
<tr>
<td>PLCB1</td>
<td>13</td>
<td>850.0–900.0</td>
<td>-</td>
<td>1.20</td>
<td>0.0068</td>
</tr>
<tr>
<td>TGM3</td>
<td>13</td>
<td>53300.0–53350.0</td>
<td>-</td>
<td>1.15</td>
<td>0.0088</td>
</tr>
<tr>
<td>PRG3</td>
<td>15</td>
<td>81900.0–81950.0</td>
<td>-</td>
<td>1.66</td>
<td>0.0009</td>
</tr>
<tr>
<td>ATP2A2</td>
<td>17</td>
<td>56477.0–56527.0</td>
<td>157.57</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KRT33A</td>
<td>19</td>
<td>42150.0–42200.0</td>
<td>-</td>
<td>1.45</td>
<td>0.0022</td>
</tr>
<tr>
<td>SLC45A2</td>
<td>20</td>
<td>39850.0–39900.0</td>
<td>-</td>
<td>1.34</td>
<td>0.0037</td>
</tr>
<tr>
<td>BoLA</td>
<td>23</td>
<td>27850.0–27900.0</td>
<td>110.91</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MC5R</td>
<td>24</td>
<td>43,975.7–44025.7</td>
<td>121.10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FGFR2</td>
<td>26</td>
<td>41925.6–41975.6</td>
<td>121.62</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DMBT1</td>
<td>26</td>
<td>42750.0–42800.0</td>
<td>-</td>
<td>1.54</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Chr.: Chromosome; XP-CLR: cross-population composite likelihood ratio; XP-EHH: cross-population extended haplotype homozygosity
have evolved might be the selective pressure for the genes involved in pathways and BP terms related to immune responses.

Additionally, genes that contribute to the superior tick and parasite resistance mechanisms in African Indicus cattle were identified (Table 1). Host tick resistance is defined as the ability of the host to reduce damage caused by the parasite through diverse immune and non-immune structural components (Neto et al., 2011). Immune response genes identified in relation to tick resistance include BoLA, TNFAIP8L3, and SLC25A48. BoLA, also known as the bovine major histocompatibility complex, has a major role in antigen processing and presentation. It plays a vital role in host resistance to an ecto- and endoparasite infestation (Martínez et al., 2006; Shyma et al., 2015). BoLA is highly polymorphic; several scholars reported the association of different alleles to tick resistance (Martínez et al., 2006; Neto et al., 2011). TNFAIP8L3 and SLC25A48 genes are involved in antigen recognition and have been previously identified to be under selection in African cattle breeds in relation to tick resistance (Makina et al., 2015).

The physical structure of the epidermal layers of cattle skin represents the first line of defense against ectoparasite invasion (Gautier et al., 2009; Kongswan et al., 2010). In relation to this, keratin and keratin associated genes (KRT33A, KRTAP27-1, and KRTAP9-1) were identified under selection. Keratin genes are heteropolymeric structural proteins that form the structural framework of the skin and hair cells contributing to tick resistance via their action as a barrier to the external environment (Nakamura, Kanemaru, & Fukami, 2013). KRT33A gene was previously found overrepresented in highly tick resistant cattle (Kongswan et al., 2010). Keratinocytes in the epidermis also have an immune response function— they secrete cytokines that initiate local inflammatory responses (Nakamura et al., 2013). Tick resistance in African cattle has been understood to be due to superior skin immunity (Marufu, Ockweni, Chimonyo, & Dzama, 2011). The positive selection of keratin-related genes has been previously reported in African cattle (Babahani et al., 2015; Chan et al., 2010; Makina et al., 2015). Collagens (COL12A1, COL8A1) provide structural integrity to the skin (Wang et al., 2007). Structural integrity was found over-represented from genes differentially expressed in tick resistant Brahman cattle (Kongswan et al., 2008). Extracellular matrix proteins (PRG3, VWA2, and ATRN) form a protective barrier producing an unfriendly environment for tick attachment and feeding (Kongswan et al., 2008). PRG3 stimulates neutrophil superoxide production and histamine release from basophils. It is involved in histamine biosynthetic process and activates basophils that are important effectors of tick rejection and a major component of acquired resistance of the host (Falcone, Pritchard, & Gibbs, 2001; Wikel, 1996).

The contribution of light coat color to cattle tick resistance has been previously reported (Ibelli et al., 2012; Mapholi et al., 2014; Marufu et al., 2011). This might be due to the fact that on light colored animals ticks can be easily seen by predator birds and picked up easily (Ibelli et al., 2012; Mapholi et al., 2014; Marufu et al., 2011). Melanogenesis, defined as the production of melanin pigments, plays a role in the pigmentation of skin and hair (Cieslak, Reissmann, Hofreiter, & Ludwig, 2011). Several genes (SLC45A2, MLPH, RAB17, RAB37, RAB42, RAB7A, and ATRN), which are involved at different stages of melanocyte biology (Guilbert, Girardot, Leveziel, Julien, & Oulmouden, 2004), were detected in this study. Mutations in these genes cause dilution of phenotypes (Cieslak et al., 2011), which might result in light coat color of African Indicine cattle breeds (Marufu et al., 2011). Mutations in MLPH gene have been found to cause dilution of coat color (Philipp et al., 2005). In addition to color, cattle hairs are coated with an emulsion of sweat and sebum, which prevents tick attachment (Kongswan et al., 2010). Searching for non-synonymous mutations, we identified five (three known—rs136185632, rs209852661, rs209010309 and two novel—3:117616155, 3:117632216), and two known (rs134604394, rs378039235) significant missense variants on MLPH and SLC45A2 gene regions, respectively (Table 2). MC5R controls the secretion of sebum (Zhang, Li, Anthonavage, & Eisinger, 2006), a complex mixture of lipid that contributes to skin integrity and inflammatory processes, transport antioxidants to the skin surface, and has innate antimicrobial activity (Zhang et al., 2011). In this
gene region, one novel (24:43984677) and one known (rs134437472) significant missense variants were identified (Table 2). The smoother coat and shorter hair of indigenous African cattle breeds have been reported to contribute to their superior tick resistance (Marufu et al., 2011).

Genes involved in calcium signaling (ATP2A2, TGM3, and TGM1) have also been found to play an important role in host response to tick challenge (Bagnall, Gough, Cadogan, Burns, & Kongsuwan, 2009; Kongsuwan et al., 2010, 2008). The ATP2A2 gene, which encodes Ca2+-dependent ATPase, is reported to be highly expressed in highly tick resistant cattle (Bagnall et al., 2009). Transglutaminases (TGM3, TGM1) are Ca2+-dependent enzymes expressed in terminally differentiating keratinocytes that are involved in the apoptosis and cornification of the epidermis. TGMs are required for crosslinking the main component of the epidermal cornified envelope in the epidermis (Candi, Schmidt, & Melino, 2005) and provide a defense function against ectoparasites (Kongsuwan et al., 2010). In TGM1 and TGM3 gene regions, we identified two novel (10:20736590, 10:20739046) and three known (rs41695720, rs136283113, rs211468449) significant missense variants, respectively (Table 2).

Gastrointestinal (GI) parasites are economically significant livestock health problems causing production losses (Benavides, Sonstegard, & Van Tassell, 2016; McManus, Do Prado Paim, de Melo, Brasil, & Paiva, 2014). Gastric acid secretion, one of the identified enriched KEGG pathways, has a role in protecting the host against several parasitic diseases in humans (Martinsen, Bergh, & Waldum, 2005). Genes involved in GI parasite resistance identified in this study include TNFAIP3, TNFAIP8L3, and DMBT1. TNF proteins encode a multifunctional pro-inflammatory cytokine related to nematode resistance (Araujo et al., 2009). DMBT1 is a protein-coding gene that plays a crucial role in mucosal defense, cellular immune defense, and epithelial differentiation. It has been found highly expressed in the mesenteric lymph node of intestinal nematode resistant animals (Araujo et al., 2009). Through observation of the genetic variation relating to gastrointestinal parasite resistance between and within breeds, it is evident (McManus et al., 2014) that Indicus cattle have evolved for superior parasite resistance as compared to Taurine cattle and their crosses (Frisch, O’Neill, & Kelly, 2000).

4. Limitation of this study
As false positive results are not uncommon in these kinds of studies, additional validation procedures are important. Additionally, although the statistical methods used are robust enough to

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr.</th>
<th>SNP ID</th>
<th>aa change</th>
<th>Logistic p-value</th>
<th>X² P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLPH</td>
<td>3</td>
<td>rs136185632</td>
<td>p.Ser131Arg</td>
<td>1.78E-05</td>
<td>7.93E-18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>:117616155</td>
<td>p.Ala301Thr</td>
<td>2.62E-05</td>
<td>8.77E-20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>:117632216</td>
<td>p.Pro518Ser</td>
<td>0.000105</td>
<td>7.83E-19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs209852661</td>
<td>p.Leu519Ser</td>
<td>9.51E-06</td>
<td>6.49E-16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs209013090</td>
<td>p.Thr571Ala</td>
<td>9.51E-06</td>
<td>6.49E-16</td>
</tr>
<tr>
<td>TGMI</td>
<td>10</td>
<td>10:207369046</td>
<td>p.Arg103Gln</td>
<td>0.003422</td>
<td>0.00017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10:20739046</td>
<td>p.Val349Ile</td>
<td>3.64E-05</td>
<td>6.62E-18</td>
</tr>
<tr>
<td>TGM3</td>
<td>13</td>
<td>rs41695720</td>
<td>p.Tyr420His</td>
<td>0.000498</td>
<td>2.69E-13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs136283113</td>
<td>p.Ala412Ser</td>
<td>0.006084</td>
<td>0.00247</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs211468449</td>
<td>p.Ser239Asn</td>
<td>3.90E-06</td>
<td>5.50E-13</td>
</tr>
<tr>
<td>SLC45A2</td>
<td>20</td>
<td>rs134604394</td>
<td>p.Phe182Tyr</td>
<td>0.002168</td>
<td>3.23E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>rs378039235</td>
<td>p.Ala240Thr</td>
<td>4.30E-05</td>
<td>1.65E-07</td>
</tr>
<tr>
<td>MCSR</td>
<td>24</td>
<td>rs134437472</td>
<td>p.Ala171Thr</td>
<td>8.98E-05</td>
<td>1.28E-23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24:43984677</td>
<td>p.His266Gln</td>
<td>8.98E-05</td>
<td>1.28E-23</td>
</tr>
</tbody>
</table>
identify recent signatures due to natural and artificial selection, the genes identified here might be due to the divergence between the two species which occurred millions of years ago.

5. Conclusion

Results of our analysis identified important putative genes and pathways that are positively selected in East African Indicus cattle in response to different selection pressures of the African tropical environment. The interaction of African cattle breeds with the high disease and parasite prevalence, high temperature, and seasonal feed and water scarcity prevailing in the region might be the main selective pressure underlying the selection signature of genes. Identification of these genes allows us to better understand the biological process and mechanisms of tropical environmental adaptation in African cattle and can be used in future genomic selection programs. The genes identified in relation to tick and parasitic resistance can be used for the development of tick and parasite control methods such as designing selection programs and development of vaccines.

Additional file

Additional file 1: Table S1. List of genes identified as positively selected in African cattle from a genome-wide comparison of three East African Indicine cattle breeds and four (three European and one Asian) Taurine cattle breeds using XP-CLR statistics; Table S2. List of genes identified as positively selected in African cattle from a genome-wide comparison of three East African cattle breeds and four (three European and one Asian) Taurine cattle breeds using XP-EHH statistics (Common genes are indicated with red text).

Additional file 2: Table S3. Enriched KEGG Pathways overrepresented from DAVID Gene ontology analysis for genes identified from East African Indicine cattle breeds. Table S4. Gene name and description of candidate genes that contribute to tropical environment adaptation mechanisms of East African Indicine cattle (from geneCards [provided by RefSeq, Apr 2014] (http://www.genecards.org/) and literature). Figure S1. Population structure at K = 3, 5 and 6 population assumptions in seven cattle breeds.

Supplementary material

Supplemental data for this article can be accessed here.

Acknowledgements

This work was supported by Agenda (PJO1134905) of the National Institute of Animal Science, Rural Development Administration (RDA), Korea.

Funding

The authors received no direct funding for this research.

Author details

Mengistie Taye1,2
E-mail: mengistietaye@yahoo.com
ORCID ID: http://orcid.org/0000-0001-6795-9943
Wonseok Lee1
E-mail: rznanka@snu.ac.kr
ORCID ID: http://orcid.org/0000-0002-5103-3334
Kelsey Caetano-Anolles1
E-mail: kelseyca@gmail.com
Tadelle Dessie3
E-mail: T.DESSION@CGIAR.ORG
Seoae Cho4
E-mail: seoae@c&kgenomics.com
Sung Jong Oh5
E-mail: shjong@jejunu.ac.kr
Hak-Kyo Lee5
E-mail: breedlee@jejunu.ac.kr
ORCID ID: http://orcid.org/0000-0002-5103-3334
Heebal Kim1,4,7
E-mail: heebal@snu.ac.kr
1 College of Agriculture and Environmental Sciences, Bahir Dar University, P. O. Box 79, Bahir Dar, Ethiopia.
2 Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea.
3 International Livestock Research Institute (ILRI), P.O. Box 5689, Addis Ababa, Ethiopia.
4 C&K genomics, Main Bldg. #514, SNU Research Park, Seoul 151-919, Republic of Korea.
6 The Animal Molecular Genetics & Breeding Center, Department of Animal Biotechnology, Chonbuk National University, Jeonju 561-756, Republic of Korea.
7 Interdisciplinary Program in Bioinformatics, Seoul National University, Kwan-ak St. 599, Kwan-ak Gu, Seoul, 151-741, Republic of Korea.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Citation information

Cite this article as: Exploring the genomes of East African Indicine cattle breeds reveals signature of selection for tropical environmental adaptation traits, Mengistie Taye, Wonseok Lee, Kelsey Caetano-Anolles, Tadelle Dessie, Seoae Cho, Sung Jong Oh, Hak-Kyo Lee & Heebal Kim, Cogent Food & Agriculture (2018), 4: 1552552.

References


Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... Daly, M. J. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559–575. doi:10.1016/j.ajhg.2007.05.015


