Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef

Citation for published version:

Digital Object Identifier (DOI):
10.1186/s12863-016-0467-1

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
BMC Genetics

Publisher Rights Statement:
© The Author(s). 2017 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef

Mengistie Taye1,2, Jaemin Kim3, Sook Hee Yoon1, Wonseok Lee1, Olivier Hanotte4,5, Tadelle Dessie5, Stephen Kemp6,7, Okeyo Ally Mwai6, Kelsey Caetano-Anollés8, Seoae Cho9, Sung Jong Oh10, Hak-Kyo Lee11 and Heebal Kim1,9,12*

Abstract

Background: Africa is home to numerous cattle breeds whose diversity has been shaped by subtle combinations of human and natural selection. African Sanga cattle are an intermediate type of cattle resulting from interbreeding between Bos taurus and Bos indicus subspecies. Recently, research has asserted the potential of Sanga breeds for commercial beef production with better meat quality as compared to Bos indicus breeds. Here, we identified meat quality related gene regions that are positively selected in Ankole (Sanga) cattle breeds as compared to indicus (Boran, Ogaden, and Kenana) breeds using cross-population (XP-EHH and XP-CLR) statistical methods.

Results: We identified 238 (XP-EHH) and 213 (XP-CLR) positively selected genes, of which 97 were detected from both statistics. Among the genes obtained, we primarily reported those involved in different biological process and pathways associated with meat quality traits. Genes (CAPZB, COL9A2, PDGFRα, MAP3K5, ZNF410, and PKM2) involved in muscle structure and metabolism affect meat tenderness. Genes (PLA2G2A, PARK2, ZNF410, MAP2K3, PLCΔ3, PLCΔ1, and ROCK1) related to intramuscular fat (IMF) are involved in adipose metabolism and adipogenesis. MB and SLC48A1 affect meat color. In addition, we identified genes (TIMP2, PKM2, PRKG1, MAP3KS, and ATP8A1) related to feeding efficiency. Among the enriched Gene Ontology Biological Process (GO BP) terms, actin cytoskeleton organization, actin filament-based process, and protein ubiquitination are associated with meat tenderness whereas cellular component organization, negative regulation of actin filament depolymerization and negative regulation of protein complex disassembly are involved in adipocyte regulation. The MAPK pathway is responsible for cell proliferation and plays an important role in hyperplastic growth, which has a positive effect on meat tenderness.

Conclusion: Results revealed several candidate genes positively selected in Ankole cattle in relation to meat quality characteristics. The genes identified are involved in muscle structure and metabolism, and adipose metabolism and adipogenesis. These genes help in the understanding of the biological mechanisms controlling beef quality characteristics in African Ankole cattle. These results provide a basis for further research on the genomic characteristics of Ankole and other Sanga cattle breeds for quality beef.

Keywords: African cattle, Ankole cattle, Meat quality, Sanga cattle, XP-CLR, XP-EHH
Background

Africa, with its diverse agro-ecological zones, is a home to diverse cattle breeds adapted to their local environments. African cattle breeds are derived from *Bos taurus* and *Bos indicus* subspecies introduced to the continent at different times, and through interbreeding between them [1, 2]. Since the introduction, their diversity has been shaped by subtle combinations of human and natural selection. Selection in African cattle is mainly for sociocultural concerns and to survive the heterogeneous environment [3]. African cattle have been evolved to adapt to the poor feed availability, high environmental temperature, and high prevalence of internal and external parasite and disease conditions of the continent. These cattle breeds display better heat tolerance, adaptability, tick resistance, reproductive longevity, and maternal characteristics such as fertility, low inter-calf periods and cow efficiency [1, 4–7].

African Sanga cattle, sometimes referred to as *Bos africana*, are an intermediate type of cattle believed to be the result of interbreeding between *Bos taurus* and *Bos indicus*, which dwell in eastern, central and southern Africa [1, 8]. Generally, Sanga cattle can be identified by their long and slender horns, small cervicothoracic hump, and small and unfolded dewlap [9]. There are 30 Sanga cattle breeds/strains in Africa subdivided into Sanga of eastern and Sanga of southern Africa based on geographical location [8]. Recently, research outputs are asserting the potential of African Sanga and Sanga derived breeds to produce carcass and meat quality attributes that favorably compare to British and Continental breeds and are often better than those of the *Bos indicus* breeds [6, 10–13]. Sanga breeds in south Africa (eg., Bonsmara, Drakensberger and Nguni) were found to produce beef with lower shear force, shorter myofibrillar fragment length, larger rib fat thickness, and larger soluble collagen when compared with indicus (Brahman) cattle [11].

Meat quality is a general term used to describe the attributes of meat which include carcass composition and conformation, the eating quality of meat, health issues associated with meat, and production and environmental issues [14]. Meat sensory characteristics such as tenderness, flavor, juiciness, and color are important meat quality parameters which are affected by biological characteristics and proteolytic activities of muscle [15, 16]. The biological characteristics of muscles such as fiber type, collagen, intramuscular adipose tissue and protease activities regulate meat tenderness and flavor and are known to be affected by genetic and rearing factors [15]. The heritability of beef quality traits is low to moderate which varies between breed groups, methods of estimation, number of records, and other factors [17, 18]. The genetic variation within and between breeds is because of the positive selection of gene regions caused by beneficial polymorphisms in the genes affecting the traits. Identification of selection signatures in the genome provide information about the evolutionary processes involved in shaping genomes and functional information about genes/genomic regions [19].

Studies attempting to detect positive selection signatures in African cattle have reported several genes involved in immune system, reproduction, energy metabolism, coat coloration, thermoregulation and tick resistance [20, 21]. The detection of immune related genes might be related to the selective pressure that has been exerted by the long-term presence of pathogens in the continent [21], whereas signatures of selection associated with reproduction and thermoregulation is an adaptation to perform under heat stress conditions [20]. However, there have been no previous studies attempting to identify genes affecting meat quality traits in African cattle in general and Sanga cattle in particular.

In this study, we reported genes that are positively selected in Ankole cattle population associated with meat quality traits. This was done by scanning the whole genome of four African cattle breeds (African Sanga cattle: Ankole; and three indicus breeds: Boran, Ogaden, and Kenana) [22, 23]. We employed cross-population extended haplotype homozygosity (XP-EHH) and cross population composite likelihood ratio (XP-CLR) statistics in order to detect selection signatures from different data patterns; two approaches were used as each has its own advantages. XP-EHH compares haplotype lengths of populations to detect selective sweeps when the allele has approached or achieved fixation in one population but remains polymorphic in the other population [24]. XP-CLR is a statistic based on allele frequency differentiation across populations. It is not affected by ascertainment biases and has the advantage of being able to detect older signals and selection on standing variation [25].

Results and discussion

Data description

DNA samples extracted from whole blood samples of four African cattle breeds (Boran, Ogaden, Kenana and Ankole) were sequenced to ~11 × genome coverage each. Using a standard sample preparation and whole genome re-sequencing pipeline, an overall alignment rate of 98.84% covering 98.56% of the taurine reference genome was obtained. After filtering false positive calls using several filtering steps, a total of ~37 million SNPs were retained and used for detection of positive selection signature analysis.

Phylogenetic tree

Maximum likelihood (ML) and neighbor-joining (NJ) methods produced consistent features regarding the
genetic distance between the breeds considered (Fig. 1). Ankole cattle are clearly separated from the three indicus breeds (100% bootstrap values/quartet puzzling reliability values). Within indicus cattle, each of the three breeds were also depicted as a monophyletic group with highly significant values.

While African indicus cattle have been present on the continent since 1500 BC [8], Sanga cattle were derived through hybridization between taurine cattle and zebu cattle around 700 AD [1]. More than 30 breeds/strains of Sanga cattle can be found distributed throughout Eritrea, Ethiopia, southern Sudan, the Great Lakes region of East Africa and southern Africa. The Ankole group is one of the three groups of Sanga cattle representing Sanga cattle in Uganda, Rwanda, Burundi, Tanzania and Democratic Republic of Congo [8]; it is a valuable and widely used genetic resource in these regions.

**Positive selective signature in Ankole cattle population**

XP-EHH and XP-CLR tests were performed in order to detect positive selection signatures in Sanga (Ankole) cattle. The genome of Ankole population was compared with the genomes of three indicus cattle breeds grouped together into one population. Based on the analysis, we obtained 238 and 213 putatively advantageous positively selected genes from XP-EHH (Additional file 1: Table S1) and XP-CLR test statistics (Additional file 2: Table S2), respectively; of these, 98 genes were detected in both statistics (Additional file 3: Table S3). Gene Ontology Biological Processes (GO BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways within DAVID were used to build on biological modules consisting of clusters of functional terms [26]. All the 354 genes obtained from both XP-EHH and XP-CLR statistics were included, after removing duplicates, for the analysis. Gene ontology analysis resulted in 44 significantly ($p < 0.05$)
enriched GO BP categories (Additional file 4: Table S4) and the KEGG-pathway analysis resulted in three significantly enriched pathways \((p < 0.05; \text{Table 1})\). The ClueGO plugin [27] created a functionally organized pathway term networks (Fig. 3).

We also analyzed Tajima’s D for the candidate gene regions which revealed a significant departure from neutrality and indicated the selective maintenance of alleles within the Ankole population as compared to its indicus counterparts (Table 2). The negative Tajima’s D values obtained for the candidate gene regions indicate the presence of an excess of rare alleles in the population. It is known that low-frequency alleles contribute less to the number of pair-wise differences in a sample set than alleles of moderate frequency do; a surplus of rare alleles inflates the latter value disproportionately to the former value [28]. Similarly, population differentiation analysis supported the positive selection of candidate genes (Table 2); candidate gene regions produced higher values of fixation index [29]. \(F_{ST}\) has been widely used to identify selective sweep regions in different livestock species [30]. The Tajima’s D and \(F_{ST}\) plot of candidate gene regions are presented in Fig. 4 and Additional file 5: Figure S1.

### Biological process and pathways related to meat quality traits

Meat quality is a multifactorial and complex trait affected by different factors at different levels ranging from molecular to mechanical. Molecularly, genes involved in many cellular mechanisms such as muscle growth, glycolysis, muscle contraction, stress reaction, cell cycle, proteolysis, protein ubiquitination and apoptosis have been reported to be associated with meat quality characteristics [16, 31, 32]. Previous studies reported that, as compared with indicus breeds, Sanga breeds produce better quality beef [11–13] with lower shear force, shorter myofibrillar fragment length, larger rib fat thickness, larger soluble collagen, and higher percent drip loss [11]. Additionally, Sanga cattle have better feed conversion efficiency, reproductive performances, and tick resistance in the tropics [33].

From DAVID gene ontology analysis, 44 significant \((p < 0.05)\) GO BP terms were enriched (Fig. 2; Additional file 4: Table S4). The BP terms and gene clusters related to meat quality characteristics were chosen based on their biological function and previous literature. Accordingly, among the enriched GO BP terms (Fig. 2), actin cytoskeleton organization (represented by nine genes; \(\text{FMNL1}, \text{FMNL3}, \text{DOCK2}, \text{LIMA1}, \text{ROCK1}, \text{MRAS}, \text{PRKG1}, \text{CAPZB}, \text{and ADD1}\) and actin filament-based process (additionally contains \(\text{MYO7A}\)) are related to meat tenderness [32, 34, 35]. Cellular component organization, a cellular level process which results in the assembly and arrangement of constituent parts or disassembly of a cellular component, is important for beef tenderness [32]. It is also significantly differentially expressed in relation to pork IMF and tenderness [36]. Five genes (\(\text{WWP1}, \text{MDM2}, \text{CANDI}, \text{PARK2}, \text{and LNX1}\)) were involved in protein ubiquitination, which is a key step in protein degradation [37]. Ubiquitination pathway affects muscle properties that are relevant for the quality of meat at postmortem [38], and are expressed in relation to tenderness [36]. GO terms of negative regulation of actin filament depolymerization and negative regulation of protein complex disassembly are involved in adipocyte regulation [34].

The KEGG MAPK pathway \((p = 0.0215; \text{Table 1})\), represented by eight genes (\(\text{MAP4K4}, \text{ACVR1B}, \text{FGF18}, \text{MAP3K5}, \text{MAP2K3}, \text{MRAS}, \text{PDGFRA}, \text{PLA2G2A}, \text{NR4A1}, \text{MAPKAPK2}, \text{and NFATC2}\)) is responsible for cell proliferation and plays an important role in hyperplastic growth [39], which has a positive effect on meat tenderness [31]. Gap junction, regulation of actin cytoskeleton and MAPK signaling pathway also are important in residual feed intake [40]. The ClueGO plugin created a functionally organized pathway term network (Fig. 3), that the networks actin filament bundle assembly and positive regulation of proteolysis were among enriched networks in relation to meat quality characteristics [32].

#### Table 1 KEGG pathways obtained from DAVID gene enrichment \((p < 0.05)\) analysis

<table>
<thead>
<tr>
<th>KEGG pathway term</th>
<th>P-value</th>
<th>Genes</th>
<th>Fold enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leukocyte transendothelial migration</td>
<td>0.019</td>
<td>(\text{ROCK1, PK3CB, CLDN10, TXX, RAPGEF4, RAPGEF3, CTNNA3})</td>
<td>3.24</td>
</tr>
<tr>
<td>MAPK signaling pathway</td>
<td>0.021</td>
<td>(\text{MAP4K4, ACVR1B, FGF18, MAP3K5, MAP2K3, MRAS, PDGFRA, PLA2G2A, NR4A1, MAPKAPK2, NFATC2})</td>
<td>2.25</td>
</tr>
<tr>
<td>Melanoma</td>
<td>0.038</td>
<td>(\text{FGF18, PK3CB, MITF, PDGFRA, MDM2})</td>
<td>3.85</td>
</tr>
</tbody>
</table>

We used 354 genes obtained from both XP-EHH and XP-CLR statistics after removing duplicate genes.

#### Genes affecting meat quality traits in Ankole cattle

Here, we described genes positively selected in Ankole Sanga cattle that are potentially associated with meat quality and feed conversion efficiency traits based on previous studies and their biological functions (Table 2; Additional file 6: Table S5).

#### Genes related to meat tenderness

Meat tenderness is an important meat eating quality trait. It is mainly affected by the quantity and solubility of connective tissue, composition and contractile state of muscle fibers, and the extent of proteolysis in rigor muscle [11, 31, 41]. Tender meat contains higher levels of soluble
collagen, more fat, and lower water content. Myofibril fragmentation index also has a positive correlation with beef loin tenderness [42]. Sanga breeds have a lower percentage of white muscle fiber and a higher myofibrillar fragmentation index [6, 11], which results in lower shear force and more tender beef compared to indicus cattle [11]. In this study, we have identified genes (CAPZB, COL9A2, PDGFRA, MAP3K5, ZNF410, LIMA1, and PKM2) that may potentially affect muscle structure and development thereby affecting meat tenderness in Ankole cattle.

The CAPZB (XP-CLR = 142.50) gene encodes the beta subunit of the barbed-end actin binding protein, which belongs to the F-actin capping protein family. It is involved in skeletal muscle development and growth [43], and cell signaling and regulation of actin in myofilament contractility [38]. When up-regulated, it increases the ability of muscle accretion in pigs [43]. CAPZB contributes to muscle metabolic and structural properties and proteolytic processes providing a link between these functional networks which are important for maturation of muscle to meat [38]. A previous functional analysis of meat tenderness revealed a positive correlation between CAPZB expression and beef tenderness [32]. In the pig, CAPZB is an essential element for protein kinase signaling to the myofilaments and, as a structural protein, it has been shown to influence muscle biochemistry and its postmortem abundance is related to meat quality [44]. The Tajima’s D and FST plot of the CAPZB gene region (Fig. 4a) shows the presence of an excess of rare alleles in Ankole population and the differentiation of the region between the compared breeds, respectively.

### Table 2

<table>
<thead>
<tr>
<th>Candidate genes</th>
<th>Chr.</th>
<th>Window (Mbp)</th>
<th>XP-CLR</th>
<th>Max XP-EHH</th>
<th>XP-EHH P-value</th>
<th>Tajima’s D</th>
<th>Weighted FST</th>
<th>Species, trait and reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK3C8</td>
<td>1</td>
<td>131.50–131.55</td>
<td>121.48</td>
<td>2.21</td>
<td>1.69E-03</td>
<td>–1.06</td>
<td>0.27</td>
<td>RFI [79]</td>
</tr>
<tr>
<td>MIR5</td>
<td>1</td>
<td>131.73–131.78</td>
<td>112.25</td>
<td>-</td>
<td>-</td>
<td>–1.87</td>
<td>0.16</td>
<td>FCE [79]</td>
</tr>
<tr>
<td>PLA2G2A</td>
<td>2</td>
<td>133.28–133.33</td>
<td>117.48</td>
<td>-</td>
<td>-</td>
<td>0.70</td>
<td>0.27</td>
<td>IMF [58], [59]</td>
</tr>
<tr>
<td>CAPZB</td>
<td>2</td>
<td>133.78–133.83</td>
<td>142.50</td>
<td>-</td>
<td>-</td>
<td>–0.47</td>
<td>0.22</td>
<td>Tenderness [32]</td>
</tr>
<tr>
<td>COL9A2</td>
<td>3</td>
<td>106.38–106.43</td>
<td>107.14</td>
<td>1.93</td>
<td>5.95E-03</td>
<td>–1.30</td>
<td>0.15</td>
<td>Tenderness [46]</td>
</tr>
<tr>
<td>LIMA1</td>
<td>5</td>
<td>29.78–29.83</td>
<td>117.18</td>
<td>-</td>
<td>-</td>
<td>–0.33</td>
<td>0.39</td>
<td>Tenderness [45]</td>
</tr>
<tr>
<td>SLC48A1</td>
<td>5</td>
<td>32.63–32.68</td>
<td>101.45</td>
<td>1.80</td>
<td>6.96E-03</td>
<td>–0.35</td>
<td>0.26</td>
<td>Meat color [68]</td>
</tr>
<tr>
<td>MB</td>
<td>5</td>
<td>74.18–74.23</td>
<td>129.86</td>
<td>1.96</td>
<td>5.00E-03</td>
<td>–0.98</td>
<td>0.20</td>
<td>Meat color [41, 67]</td>
</tr>
<tr>
<td>APOC6</td>
<td>5</td>
<td>74.20–74.25</td>
<td>-</td>
<td>2.00</td>
<td>5.00E-03</td>
<td>–0.91</td>
<td>0.13</td>
<td>IMF [65]</td>
</tr>
<tr>
<td>ATP8A1</td>
<td>6</td>
<td>62.90–62.96</td>
<td>264.27</td>
<td>2.02</td>
<td>4.00E-03</td>
<td>–1.39</td>
<td>0.36</td>
<td>FCE [76]</td>
</tr>
<tr>
<td>PDGFRB</td>
<td>6</td>
<td>71.35–71.40</td>
<td>408.60</td>
<td>2.61</td>
<td>3.00E-04</td>
<td>–2.29</td>
<td>0.45</td>
<td>Tenderness [53]</td>
</tr>
<tr>
<td>MAP3K5</td>
<td>9</td>
<td>75.55–75.60</td>
<td>147.37</td>
<td>2.09</td>
<td>3.00E-03</td>
<td>–0.23</td>
<td>0.15</td>
<td>RFI [75]</td>
</tr>
<tr>
<td>PARK2</td>
<td>9</td>
<td>99.13–99.18</td>
<td>132.40</td>
<td>2.34</td>
<td>2.36E-03</td>
<td>–0.27</td>
<td>0.30</td>
<td>IMF [62]</td>
</tr>
<tr>
<td>PKM2</td>
<td>10</td>
<td>18.95–19.00</td>
<td>-</td>
<td>1.97</td>
<td>2.00E-03</td>
<td>–1.18</td>
<td>0.25</td>
<td>Tenderness [32], [52]; IMF [64]; Drip loss [70, 71]</td>
</tr>
<tr>
<td>ZNF410</td>
<td>10</td>
<td>85.68–85.73</td>
<td>165.86</td>
<td>2.18</td>
<td>1.94E-03</td>
<td>–0.67</td>
<td>0.28</td>
<td>Tenderness [32], [35]</td>
</tr>
<tr>
<td>AHS1</td>
<td>10</td>
<td>89.70–89.75</td>
<td>-</td>
<td>2.23</td>
<td>2.00E-03</td>
<td>–1.47</td>
<td>0.31</td>
<td>IMF [66]</td>
</tr>
<tr>
<td>MAP4K4</td>
<td>11</td>
<td>6.65–6.70</td>
<td>-</td>
<td>1.85</td>
<td>9.00E-03</td>
<td>0.72</td>
<td>0.08</td>
<td>Drip loss [70, 71]</td>
</tr>
<tr>
<td>NFATC2</td>
<td>13</td>
<td>80.00–80.05</td>
<td>-</td>
<td>2.01</td>
<td>4.24E-03</td>
<td>2.53</td>
<td>0.25</td>
<td>Tenderness [55]</td>
</tr>
<tr>
<td>WWPP1</td>
<td>14</td>
<td>78.63–78.68</td>
<td>133.90</td>
<td>-</td>
<td>-</td>
<td>1.03</td>
<td>0.18</td>
<td>Tenderness [50]</td>
</tr>
<tr>
<td>OR2D2, OR10A4,</td>
<td>15</td>
<td>46.40–46.45</td>
<td>-</td>
<td>2.07</td>
<td>3.45E-03</td>
<td>2.61</td>
<td>0.18</td>
<td>Feed intake [78], [77]</td>
</tr>
<tr>
<td>OR2D3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP2K3</td>
<td>19</td>
<td>35.85–35.90</td>
<td>-</td>
<td>1.93</td>
<td>4.00E-03</td>
<td>–0.20</td>
<td>0.16</td>
<td>IMF [49]</td>
</tr>
<tr>
<td>PLCD3</td>
<td>19</td>
<td>45.40–45.45</td>
<td>-</td>
<td>1.89</td>
<td>5.00E-03</td>
<td>–0.18</td>
<td>0.03</td>
<td>IMF [56]</td>
</tr>
<tr>
<td>TM2P</td>
<td>19</td>
<td>54.10–54.15</td>
<td>-</td>
<td>1.88</td>
<td>7.46E-03</td>
<td>–0.73</td>
<td>0.18</td>
<td>RFI [73]</td>
</tr>
<tr>
<td>PLCD1</td>
<td>22</td>
<td>11.45–11.50</td>
<td>195.30</td>
<td>2.12</td>
<td>3.00E-03</td>
<td>–1.43</td>
<td>0.14</td>
<td>IMF [56]</td>
</tr>
<tr>
<td>ROCK1</td>
<td>24</td>
<td>35.48–35.53</td>
<td>130.40</td>
<td>2.22</td>
<td>1.61E-03</td>
<td>0.93</td>
<td>0.29</td>
<td>Tenderness [48]; IMF [61]</td>
</tr>
<tr>
<td>PKRG1</td>
<td>26</td>
<td>7.34–7.48</td>
<td>166.61</td>
<td>-</td>
<td>-</td>
<td>–1.27</td>
<td>0.13</td>
<td>Tenderness [54]; IMF [36]; RFI [72]</td>
</tr>
</tbody>
</table>

Note: Chr. Chromosome; Window: start and end positions of the gene region; RFI residual feed intake, FCE feed conversion efficiency, IMF intramuscular fat; Superscripts in the Species, Trait and Reference column indicate the species that the trait has been previously reported for as chicken, pork, beef, general (not for a specific species).
**LIMA1** (called **EPLIN**) encodes a cytoskeleton-associated protein that inhibits actin filament depolymerization and cross-links filaments in bundles. It is associated in pigs with functions regarding muscle development and metabolism [45]. **ZNF410**, also known as **APA-1**, is an essential component of the stress pathway involved in the meat tenderization process [32]. In previous muscle transcriptome analyses, **ZNF410** has been shown to be highly expressed in the longissimus muscle of Basque pigs that are known to produce pork with higher intramuscular fat and tenderness compared to Large White pigs [35]. **COL9A2**, a fibrillar collagen, constitutes the largest component of extracellular matrix (ECM) to which its amount, type, and solubility present in muscle tissue have a strong effect on meat tenderness [39]. This gene was found to be upregulated in the longissimus dorsi muscle of Jeju native piglets [46], whose meat is known for its preferable taste, tenderness and superior marbling [47]. **ROCK1**, a gene that regulates actin cytoskeleton and cell polarity, is associated with body weight, carcass weight, shank length, shank circumference and other carcass weight traits in chicken [48].

Genes involved in MAPK signaling (**MAP3K5**, **MAP2K3**, **MAP4K4**, and **MAPKAPK2**) were also identified. MAPK signaling is one of the major intracellular signaling pathways affecting myogenesis [49] and is relevant to postmortem meat quality [38]. **MAP2K3** shows associations with loin muscle area and fat traits in pigs, implying roles in muscle differentiation and growth [49].

E3 ubiquitin ligase genes (**WWP1**, and **PARK2**) play an important role in the regulation of a wide variety of cellular functions such as protein degradation, transcription, and RNA splicing. These genes catalyze protein...
ubiquitylation resulting in the targeting of proteins toward various cellular fates, with proteasome-mediated proteolytic degradation [50]. The ubiquitin–proteasome system is one of the proteolytic systems responsible for the majority of the protein degradation in muscle that is relevant for meat quality postmortem [51].

The expression of PKM2 (XP-EHH = 1.9696; \( p = 2.00 \times 10^{-3} \)), a gene involved in energy metabolism, is positively correlated with WBSF and has been reported as a functional protein marker for meat tenderness in Thai indigenous chicken [52] and beef [32]. PDGFRA also has an effect on shear force and Loin Eye Area in pig [53].

The Tajima’s \( D \) and \( F_{ST} \) plot of PKM2 and PDGFRA gene regions is shown in Fig. 4b and c, respectively. PRKG1 is reported to be important in the conversion of muscle to meat [54]. NFATC2 is a calcineurin substrate expressed in skeletal muscle which is responsible for activating new myotubes [39]. Calcineurin is crucial for myocyte differentiation and determination of the slow oxidative fibre phenotype [55].

**Genes related to meat intramuscular fat (IMF)**
IMF is a heritable meat quality trait which affects flavor, juiciness, visual characteristics and meat tenderness. It is
positively correlated with body fat and red muscle fiber [41]. Steak from Ankole cattle has been found to be juicier than those Ankole-Boran crossbreds [13], and Strydom et al. showed higher levels of rib fat thickness in Sanga as compared to indicus cattle [10].

We identified several genes (PLA2G2A, PARK2, ZNF410, PKM2, MAP2K3, PLCD3, PLCD1, ROCK1, and AHSA1) which affect the fat content of meat in Ankole cattle. PLA2G2A (XP-CLR = 117.48) is a member of the phospholipase A2 family (PLA2), which is involved in the hydrolysis of phospholipids into fatty acids and phosphatidylinositol and phospholipid metabolism [56]. Also referred to as Adipose-Specific Phospholipase A2 (AdPLA), it is involved in adipocyte metabolism and catalyzes the efficient release of fatty acids from triglycerides.

Fig. 4 Tajima’s D and FST plot of positively selected gene regions in Sanga and indicus cattle populations. a CAPZB gene; b PKM2 gene; c PDGFRα gene; d AHSA1 gene; and e MB gene. For other genes, please see Additional file 6: Figure S1. The Tajima’s D plot for each gene region (upper plot for each gene) show the Tajima’s D value within a 50 kb window plotted for both populations. The smaller (negative) Tajima’s D value in the Sanga population shows that the gene region considered is under positive selection. The FST plot (lower plot for each gene) represents the FST values calculated within 50 kb windows separated by 5 kb window steps.
of free fatty acids and lysocephospholipid from phosphatidylcholine [57]. It has been reported in the literature that PLA2 has a positive effect on porcine fat deposition (IMF) and potentially regulates lipolysis and increases the MUFA deposition rather than the SFA deposition [58]. It is also associated with intramuscular fat in beef cattle [59]. Pla2g2a has been reported to be a candidate gene in relation to obesity in mice [60].

ROCK1 is involved in pathways relevant to muscle/adipose tissue function in pigs with divergent phenotypes for fatness traits [61]. E3 ubiquitin ligase enzymes have been identified to be involved in the modulation of lipid biology [50, 62]. PARK2 is a strong positional candidate for adiposity in chicken and a positive regulator of fat metabolism [62]. PRKGI is involved in gap junction and is a candidate gene for intramuscular fat in the pigs [36]. MAP2K3 has been shown to be associated with loin muscle and fat traits in pigs [49]. MAP4K4 is involved in adipogenesis, triglyceride storage, fatty acid release, fatty acid oxidation and mitochondrial oxidative phosphorylation [63]. PKM2 is significantly associated with back fat thickness, an economically important trait in pigs [64]. APOL6 is one of the most important known genes involved in lipoprotein metabolism [65]. Phospholipase C family genes (PLCD1 and PLCD3) generate diacylglycerol and are involved in phophatidylinositol catabolism and phospholipid synthesis [56]. The transcription of AHS1 (AHAI) is related to Omega-3 fatty acids in skeletal muscle, which influence meat tenderness, juiciness, and flavor, and are beneficial to human health [66]. The positive selection of the AHS1 gene region is shown in the Tajima’s D and FST plot in Fig. 4d.

Genes related to meat color, drip loss, and feed conversion efficiency (FCR)
Meat color and water holding capacity of meat are among the quality parameters used as an indicator of freshness and wholesomeness [41, 67]. These characteristics are related to variations in the glycolysis rate and muscle temperature decline postmortem. Myoglobin (MB; XP-CLR = 129.86; XP-EHH = 1.9640; p = 5.00.E-03), a globular single chain protein located in the sarcoplasm, is the principle protein responsible for the red color of meat. MB serves as a reserve supply of oxygen and facilitates the movement of oxygen within muscles [41, 67]. Figure 4e shows the Tajima’s D and FST plot of MB gene region in Sanga and B. indicus populations. The Solute Carrier Family 48 (Heme Transporter), Member 1 (SLC48A1) is responsible for the transport of heme from endosome to the cytosol [68] and may also have a function in meat color. In general, beef from Sanga cattle breeds showed higher chorma than that of indicus cattle breeds [11].

The loss of reddish fluid mainly consisting of water and proteins from meat, called drip loss, is an important meat quality characteristics which is affected by several ante- and post-mortem factors [69]. A small but significant difference in drip loss is reported between Sanga and indicus cattle breeds; meat from Sanga cattle showed higher drip loss [10, 11]. Higher expression of PKM2 and MAP4K4 suppresses the glucose content of muscle cells promoting the onset of anaerobic production of lactate post mortem, thereby facilitating the decline in pH resulting in higher drip loss [70, 71].

Feed intake and efficiency, measured as residual feed intake (RFI), are economically important traits affecting the cost of beef production [72]. Variation in RFI (animals with lower RFI are more efficient) has a genetic component with moderate heritability [73]. We identified positively selected genes (TIMP2, PKM2, PRKGI, MAP3K5, ATP8AI) that are reported in the literature to be related to RFI and feed conversion efficiency. TIMP2 has been shown to be upregulated in low RFI animals in gene expression profiling studies on genes expressed differentially in cattle with high and low RFI [73]. PKM2 was associated with average daily gain, and feed to body weight gain ratio, with a significant additive and/or dominance effects on these traits [74]. PRKGI is involved in gap junction and is also a candidate gene for RFI in cattle [72]. MAP3K5, also known as apoptosis signal-regulating kinase 1 (ASK1), is a candidate gene for residual feed intake in pigs [75]. ATP8AI is also related to feed intake, feed conversion ratio, residual feed intake and weight gain [76]. Olfactory receptor genes (OR2D2, OR10A4, and OR2D3) have been shown to affect the perception of taste and smell [77, 78] and therefore can be related to feed intake and feeding behavior [77]. PIK3CB, and MRAS genes involved in the Akt/PI3K and MAPK signaling pathways, respectively, are important for high feed efficiency in chicken [79]. The positive selection of these genes may provide clues as to why Ankoile cattle are able to use and survive on poor quality feed and withstand severe droughts [80].

Implication of the results of this study on Ankoile population
The Ankoile group is one of the three groups of Sanga cattle representing Sanga cattle in east and central Africa [8]. Ankoile breed is a valuable and widely used genetic resource in the region due to its better adaptability. However, there have been no well-designed breed improvement programs for Ankoile and other Sanga breeds of eastern Africa [80, 81]. Selective breeding efforts in other South African Sanga cattle breeds (e.g., Mashona, Tuli, and Afrikander) have resulted in local cattle showing higher beef productivity [82]. As cattle genetic resources are being depleted [1, 3] and given the importance of this vital genetic resource, designing breeding programs that would help improve and conserve Ankoile cattle is crucial.
[81]. With this regard, the results provide a basis for further research on the genomic characteristics of Ankole cattle in relation to meat quality traits.

Limitations of the present study
As is typical in this kind of study, there is a possibility of obtaining false positive results. Therefore, validation with other methods such as GWAS, candidate gene approach and gene expression analysis are suggested. In addition, given the multifactorial nature of meat quality traits, limited published literature is available on genes affecting beef quality characteristics.

Conclusions
Results from the whole genome scan revealed several positively selected genes involved in different biological and cellular functions including those affecting meat quality characteristics. The genes identified in relation to meat quality characteristics are involved in muscle and lipid metabolism that affect tenderness and intramuscular fat content of meat; and help to improve our understanding of the biological mechanisms controlling meat quality traits in beef cattle production. These results provide a basis for further research on the genomic characteristics of Ankole and other Sanga cattle breeds for quality beef production.

Methods
Sample preparation and whole genome re-sequencing
The data used for this paper was obtained from a project: “The genome landscape of indigenous African cattle” [Kim et al., accepted]. DNA extracted from whole blood samples (10 ml) taken from four African cattle breeds (ten Ankole, nine Boran, nine Ogaden and ten Kenana) was used for this analysis. G-DEXTMIIb Genomic DNA Extraction Kit (iNtRoN Biotechnology, Seoul, Korea) was used to isolate DNA according to the manufacturer’s protocol. To generate inserts of ~300 bp, 3 μg of genomic DNA was randomly sheared using Covaris System. Using the TruSeq DNA Sample Prep. Kit (Illumina, San Diego, CA), we constructed the library following the manufacturer’s guidelines and whole genome sequencing was performed using the Illumina HiSeq 2000 platform. To check the quality of the raw sequence data, we used fastQC software [83]. Pair-end sequence reads were mapped to the reference bovine genome (UMD 3.1) using Bowtie2 [84] with default parameters except the “-no-mixed” option. The overall alignment rate of reads to the reference sequence was 98.50% with an average read depth of 10.8×. On average across the whole samples, the reads covered 98.51% of the genome.

We used open source software packages of Picard tools (http://broadinstitute.github.io/picard), SAMtools [85], and Genome Analysis ToolKit 1.4 (GATK) [86] for downstream processing and variant calling. Picard tools was used to filter potential PCR duplicates. SAMtools was used to create index files for reference and bam files. Genome analysis toolkit 1.4 performed local realignment of reads to correct misalignments due to the presence of indels (“RealignerTargetCreator” and “IndelRealigner” arguments). We used the “UnifiedGenotyper” and “SelectVariants” arguments of GATK to call candidate SNPs. To filter variants and avoid possible false positives, the “VariantFiltration” argument of the same software was adopted with 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 [87] was used to infer the haplotype phase and impute missing alleles for the entire set of cattle populations simultaneously. After all the filtering processes, a total of ~37 million SNPs were retained and used for further analysis.

Phylogenetic reconstruction
To understand the genetic distance between the breeds considered, we conducted phylogenomic analyses using neighbor-joining (NJ) and maximum likelihood (ML) methods. A total of 26,427,196 autosomal SNPs from the genomes of 38 individuals of four breeds were used for the phylogenetic tree construction.

ML analyses [88] were performed using the program TREE-PUZZLE 5.2 [89] with the GTR model. For the quartet puzzling method (1000 puzzling steps), nucleotide frequencies and Ts/Tv ratios (3.18) were estimated from the dataset. Quartet puzzling provided reliability values for maximum likelihood analysis [90].

NJ analysis [91] was performed using the PHYLIP package 3.69 [92] based on Kimura’s [93] 2-parameter distance. Ts/Tv ratios (3.18) were estimated from the dataset using TREE-PUZZLE 5.2 [89] and were used as inputs for the SEQBOOT, DNADIST, NEIGHBOUR, and CONSENSE programs of the PHYLIP package. A bootstrap test (with 1000 pseudoreplicates) [94] was performed to obtain statistical support for each node of the NJ tree.

Detection of positive selection signals
To detect genome-wide selective sweeps, we used the XP-EHH [24] and XP-CLR [25] statistics. XP-EHH assesses haplotype differences between two populations and is designed to detect alleles that have increased in frequency to the point of fixation or near fixation in one of the two populations being compared [24, 95].
We compared the genome of Ankole cattle, used as a test population, with indicus cattle (Boran, Ogaden, and Kenana grouped into one population), used as a reference population. XP-EHH compared the integrated EHH between the two populations for each SNP and the sign of the XP-EHH score determines the direction of selection with extreme values indicating selection in the test population genome. To facilitate comparison of genomic regions across populations, we then split the genome into non-overlapping segments of 50 kb and computed the maximum XP-EHH score in each segment. In order to define the empirical P-value, genomic windows were binned in increments of 500 SNPs (combining all windows ≥ 1000 SNPs into one) according to the method used previously [95]. Regions with P-values less than 0.01 (1%) were considered strong signals in the Ankole population.

We also performed XP-CLR to identify potential regions differentially selected between the two populations [25]. XP-CLR is a likelihood method for detecting selective sweeps that involve jointly modeling the multilocus allele frequency differentiation between two populations. XP-CLR scores were calculated using XP-CLR software package [25]. We used non-overlapping sliding windows of 50 kb, maximum number of SNPs within each window as 600, and correlation level from which the SNPs contribution to XP-CLR result was down weighted to 0.95. The regions with the XP-CLR values in the top 1% of the empirical distribution (XP-CLR > 97.86) were designated as candidate sweeps and the genes that span the window regions were defined as candidate genes [22]. Significant genomic regions identified from XP-EHH and XP-CLR were annotated to the closest genes (UMD 3.1).

In order to confirm the positive selection of detected genes using these two statistics, we calculated Tajima’s D and F_{ST} for the candidate gene regions. Detecting the same gene regions using different methods can provide cogent evidence for selective influences in the region [30]. Tajima’s D is used to detect selective sweeps going to fixation in the population that makes rare alleles in excess in the population, which results in a negative Tajima’s D [28]. Population differentiation (F_{ST}) is based on the principle that natural selection can change the amount of differentiation between different populations of a species. When populations are differentiated, the amount of genetic differentiation within the region that includes selected locus will increase during when the genetic differentiation in the genomic region is greater than the level expected under neutrality, which can be a consequence of natural selection [29]. VCFtools was used in a window size of 50 kb at an interval of 5 kb steps to calculate the Tajima’s D and F_{ST} values of the candidate gene regions [96].

Characterization of candidate genes under selection
We used the Database for Annotation, Visualization, and Integrated Discovery (DAVID; version 6.7) gene ontology and annotation tool for gene enrichment analysis to further understand the biological functions and pathways of selected genes [26]. Significant GO terms provide insight into the functional characteristics of annotated genes. The KEGG database was also cross-referenced within DAVID to identify significant pathways. R software (version 3.2.1) was used for hierarchical clustering of GO terms from DAVID. Additionally, Cytoscape software’s (version 3.2.0) ClueGO plugin was used to visualize the integration of Gene Ontology (GO) terms as well as KEGG pathways and create a functionally organized GO/pathway term network [27] with default settings.

Additional files

Additional file 1: Table S1. A summary of genes from Sanga vs. Zebu (B. Indicus) XP-EHH test statistics comparison. (XLS 64 kb)

Additional file 2: Table S2. A summary of genes from Sanga vs. Zebu (B. Indicus) XP-CLR test statistics comparison. (XLS 54 kb)

Additional file 3: Table S3. Summary of genes common for both XP-EHH and XP-CLR test statistics. (XLS 45 kb)

Additional file 4: Table S4. Gene Ontology Biological Process terms obtained from DAVID gene ontology analysis using all XP-EHH and XP-CLR gene lists. (XLs 47 kb)

Additional file 5: Figure S1. Tajima’s D and F_{ST} plot of positively selected gene regions in Sanga and indicus cattle populations. The Tajima’s D plot for each gene region (upper plot for each gene) is the Tajima’s D value within 50 kb window plotted for both populations. The smaller (negative) Tajima’s D value in Sanga population shows that the gene region considered is under positive selection. The F_{ST} plot (lower plot for each gene) is the F_{ST} values within 50 kb windows separated by 5 kb steps. (DOCX 300 kb)

Additional file 6: Table S5. Names and descriptions of major candidate meat quality related genes. All gene names and descriptions are based on RefSeq. (DOCX 24 kb)
Author's contributions
MT conceived and designed the study, analyzed the data, and wrote the paper; JK designed the study and helped analyzing the data; SHY analyzed part of the data; WL helped analyzing the data; KCA helped writing the paper; OH, TD, SK, QAM, SC, SJJ, HKL, & HK designed the project; HK organized and supervised the project. All authors read and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent to participate
Not applicable.

Ethics approval
Blood samples from African indigenous cattle breeds were collected after consent from the local authorities and owners of the animals. No further specific permissions were required from the Ethics Committee of the International Livestock Research Institute at the time of the sampling.

Author details
1Department of Agricultural Biotechnology, Animal Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea. 2Bahir Dar University, College of Agriculture and Environmental Sciences, PO Box 79, Bahir Dar, Ethiopia. 3National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Building 50 Room 3551, Bethesda, MD 20892, USA. 4The University of Nottingham, School of Life Sciences, Nottingham NG7 2RD, UK. 5International Livestock Research Institute (ILRI), PO Box 5089, Addis Ababa, Ethiopia. 6International Livestock Research Institute (ILRI), PO Box 30709-00100, Nairobi, Kenya. 7The Centre for Tropical Livestock Genetics and Health, The Roslin Institute, The University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, Scotland. 8Department of Animal Sciences, University of Illinois, Urbana, IL 61801, USA. 9CGIAR genomics, Main Bldg. #514, SNU Research Park, Seoul 151-919, Republic of Korea. 10National Institute of Animal Science, RDA, Wanju, Republic of Korea. 11The Animal Molecular Genetics & Breeding Center, Department of Animal Biotechnology, Chonbuk National University, Jeonju 561-756, Korea. 12Institute for Biomedical Sciences, Shinshu University, Nagano, Japan.

Received: 6 July 2016 Accepted: 20 December 2016

Published online: 09 February 2017

References