Genomic analysis of Nigerian indigenous chickens reveals their genetic diversity and adaptation to heat-stress

Genomic analysis of Nigerian indigenous chickens reveals their genetic diversity and adaptation to heat-stress

Mifta P. Rachman[1*], Oladeji Bamidele[2], Tadelle Dessie[3], Jacqueline Smith[4], Olivier Hanotte[3,5*], Almas A. Gheyas[4*]

[1]School of Biosciences, University of Nottingham, LE12 5RD, United Kingdom; [2]African Chicken Genetic Gains (ACGG), Department of Animal Sciences, Obafemi Awolowo University, Ile Ife 220282, Nigeria; [3]LiveGene – CTLGH, International Livestock Research Institute (ILRI), P.O. Box 5689, Addis Ababa, Ethiopia; [4]Centre for Tropical Livestock Genetics and Health (CTLGH), Roslin Institute, University of Edinburgh, EH25 9RG, Edinburgh, United Kingdom; [5]School of Life Sciences, University of Nottingham, NG7 2RD, Nottingham, United Kingdom.

*Correspondence: Mifta P. Rachman, email: mmiftarachman@gmail.com; Olivier Hanotte, email: o.hanotte@cgiar.org, olivier.hanotte@nottingham.ac.uk; Almas A. Gheyas, email: almas.gheyas@roslin.ed.ac.uk (current address: almas.gheyas@stir.ac.uk).

Abstract

Indigenous poultry breeds from Africa can survive in harsh tropical environments (such as long arid seasons, excessive rain and humidity, and extreme heat) and are resilient to disease challenges, but they are not productive compared to their commercial counterparts. Their adaptive characteristics are in response to natural selection or to artificial selection for production traits that have left selection signatures in the genome. Identifying these signatures of positive selection can provide insight into the genetic bases of tropical adaptations observed in indigenous poultry and thereby help to develop robust and high-performing breeds for extreme tropical climates. Here, we present the first large-scale whole-genome sequencing analysis of Nigerian indigenous chickens from different agro-climatic conditions, investigating their genetic diversity and adaptation to tropical hot climates (extreme arid and extreme humid conditions). The study shows a large extant genetic diversity but low level of population differentiation. Using different selection signature analyses, several candidate genes for adaptation were detected, especially in relation to thermotolerance and immune response (e.g., cytochrome P450 2B4-like, *TSHR*, *HSF1, CDC37*, *SFTPB, HIF3A, SLC44A2,* and *ILF3* genes). These results have important implications for conserving valuable genetic resources and breeding improvement of chickens for thermotolerance.

Introduction

It is important to recognize the value of indigenous livestock populations from various geographic regions. These animals have adapted to their local agro-climatic conditions, making them important genetic resources for conservation efforts. By protecting these populations, we can help preserve their unique genetic traits and ensure the sustainability of our agricultural practices. Native tropical breeds are particularly crucial. As climate change and global warming are forcing many temperate regions to experience tropic-like conditions, such breeds may hold genetic solutions for climate resilience.

Nigeria is a tropical lowland country where poultry farming plays a crucial role in the economy and livelihood of local people. About 45% of the Nigerian population is involved in poultry production, mostly small or medium-scale farming, and Nigeria ranks second for its chicken population size (180 M birds) within Africa[1,2]. However, despite the importance of poultry farming for the country's economy, over half of its chickens are still raised in extensive backyard farming systems. Moreover, about 80% of the chickens reared in backyard farming in Nigeria are represented by unimproved local breeds[3]. Being unimproved, they have poor productivity, but otherwise have very desirable qualities such as hardiness to thrive under harsh tropical environments, the ability to forage for food, the ability to hatch on their own and brood, and considerable tolerance to endemic disease challenges[4]. Besides this, their egg and meat products are preferred by local people[5]. These local chickens, commonly called Nigerian Indigenous Chickens (NICs), represent important genetic resources for the sustainable development of the poultry programme in Nigeria to cater for future needs arising from climate challenges and consumer demands.

The NICs surviving in varied Nigerian agro-climatic conditions offer an excellent opportunity to dissect tropical environmental adaptation, particularly thermotolerance in chickens, both under hot-humid and hot-arid conditions. Nigeria's landscape has been classified into several agroecological zones (AEZs). Transiting from a South to North direction, these include Mangrove Swamp and Coastal Vegetation, Freshwater Swamp Forest, Lowland Rain Forest, Derived Savanna, Guinea Savanna, Sudan Savanna, and Sahel Savanna[6]. In addition, there are a few mountainous areas found in the Jos Plateau, Adamawa, Taraba, and the Northern part of Cross River State (Figure 1). Whilst most Nigerian geographic regions experience very high temperatures (except in the high plateaus), the climate varies from very wet conditions in the coastal South (annual rainfall > 3500 mm,

temperature up to 32ºC) to extreme arid conditions in the Sahel region of the North-West and North-East (annual rainfall < 600 mm, temperature up to 41ºC)[7]. NICs can be classified into ecotypes in different ways. Based on geographical location, the NICs are classified into two major breeds or ecotypes: Fulani and Yoruba[8]. The Fulani ecotype is found in the Sahel and Guinea savanna, the cattle Kraals and Montane parts of northern Nigeria, whereas the Yoruba ecotype is located around the rainforest, swamps, and derived savanna areas[8]. Alternatively, agro-climatic regions can also be used to classify the NICs into potential ecotypes, such as mangrove, freshwater swamp forest, rainforest, derived savanna, Guinea savanna, Sudan savanna, and Sahel savanna[9]. The genetic characterisation of NICs from different ecotypes, particularly those based on agroecological zones, is crucial for conserving genetic and adaptive diversity and elucidating the molecular mechanisms of environmental adaptation. In particular, the prevalence of a very high temperature across most parts of the Nigerian landscape provides an excellent opportunity to investigate the genetic basis of heat stress adaption in general, and those specific to hot-humid and hot-arid conditions.

Until now, most of the genetic studies on NICs have been based either on mtDNA or microsatellite markers[10]. No study so far reports genetic and adaptive diversity based on whole-genome sequence (WGS) data. In this study, we examine the genetic diversity in NICs with WGS data from a large number of samples representing different agro-climatic zones. Using the dense genetic variants detected in this study, we investigate the signatures of positive natural selection in the NICs in response to heat stress in humid and arid conditions.

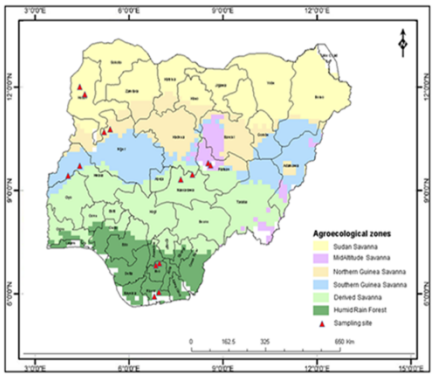
Results

Whole-genome sequencing shows a large within-population genetic diversity in NICs

In the present study, we sequenced and analysed 120 village chicken samples from 14 different populations (4 - 10 samples/population) which represent diverse AEZs across Nigeria (Figures 1A and 1B). There is little variation in the mean annual temperature among these AEZs except in the mid-altitude region with a slightly lower temperature (Figure 1C). However, the AEZs show a large variation in mean annual rainfall patterns (Figure 1D). The genomes of 120 chickens were sequenced with 1.5 billion and 6 billion clean reads for each sample. The reads were then aligned to the chicken reference genome (GRCg6a) at an average mapping rate of 99% with a mean genomic coverage of ~64X after mapping (Table S1). Using a joint analysis of all the samples, we detected over 17 M SNPs, of which ~11%

(1.9 M) are novel. For downstream analysis, population wise SNPs were extracted which resulted in 8.9 million to 11.8 million SNPs per population (Table S1).

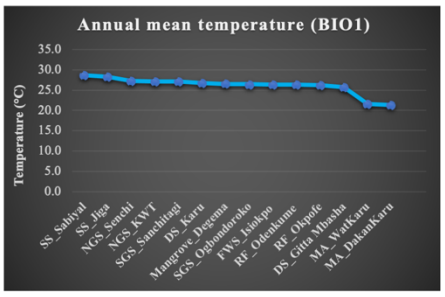
A



B

Population	#Samples	Agroecological zones	Climatic condition
Jiga	6	Sudan Savana/Arid/Semi-arid	Hot/Arid
Sabiyal	4	Sudan Savana/Arid/Semi-arid	Hot/Arid
Senchi	4	Northern Guinea Savana	Hot/Arid
Kwendo/Doro/Tadurga (KWT)	6	Northern Guinea Savana	Hot/Arid
Wat-Karu	10	MidAltitude	Cool/Subhumid
Dakan-Karu	10	MidAltitude	Cool/Subhumid
Ogbondoroko	10	Southern Guinea Savana	Hot/Semiarid
Sanchitagi	10	Southern Guinea Savana	Hot/Semiarid
Gitta Mbasha	10	Derived Savana	Hot/Semiarid
Karu	10	Derived Savana	Hot/Semiarid
Okpofe	10	Humid/Rain Forest	Hot/Semiarid
Odenkume	10	Humid/Rain Forest	Hot/Semiarid
Isiokpo	10	Humid savana/Fresh water Swamp	Hot/Humid
Degema	10	Humid Savana/Mangrove	Hot/Humid

C



D

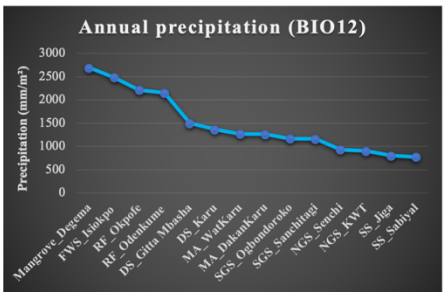


Figure 1: (A) Nigerian agro-ecological map showing sampling locations (figure modified from https://redd.unfccc.int/files/nigeria_national_frel_modified_revised__for_posting.pdf), (B) details of the studied chicken populations, (C) ordering of populations based on mean annual temperature, (D) ordering of populations based on mean annual precipitation. The means in (C) and (D) are based on 40 years of data (1960 - 2000) from the Worldclim database[11].

NICs show low levels of population differentiation.

Population structure and between-population differentiation were investigated using different approaches: Principal Component Analysis (PCA), admixture analysis, and *Fst* analysis. PCA plots show most populations clustered together except for those of Degema, DakanKaru and Odenkume (Figures 2A, 2B). Pairwise *Fst* analyses show a generally low level of population differentiation (*Fst* < 0.05; indicating no or negligible differentiation), except in a few cases where moderate differentiation (0.05 – 0.08) was observed (Figure 2C); these few cases involved the same populations (i.e., Degema, Dakan-Karu and Odenkume). Admixture analysis predicted contributions from four ancestral gene pools (Figure 2D), but closer inspection indicates that it may have been due to the presence of some exotic birds in

different populations. While most samples (at K = 4) show a mixture of two ancestral gene pools (shown by green and yellow colours), a few samples from different populations appear to have a completely different origin (shown either by black or blue colours).

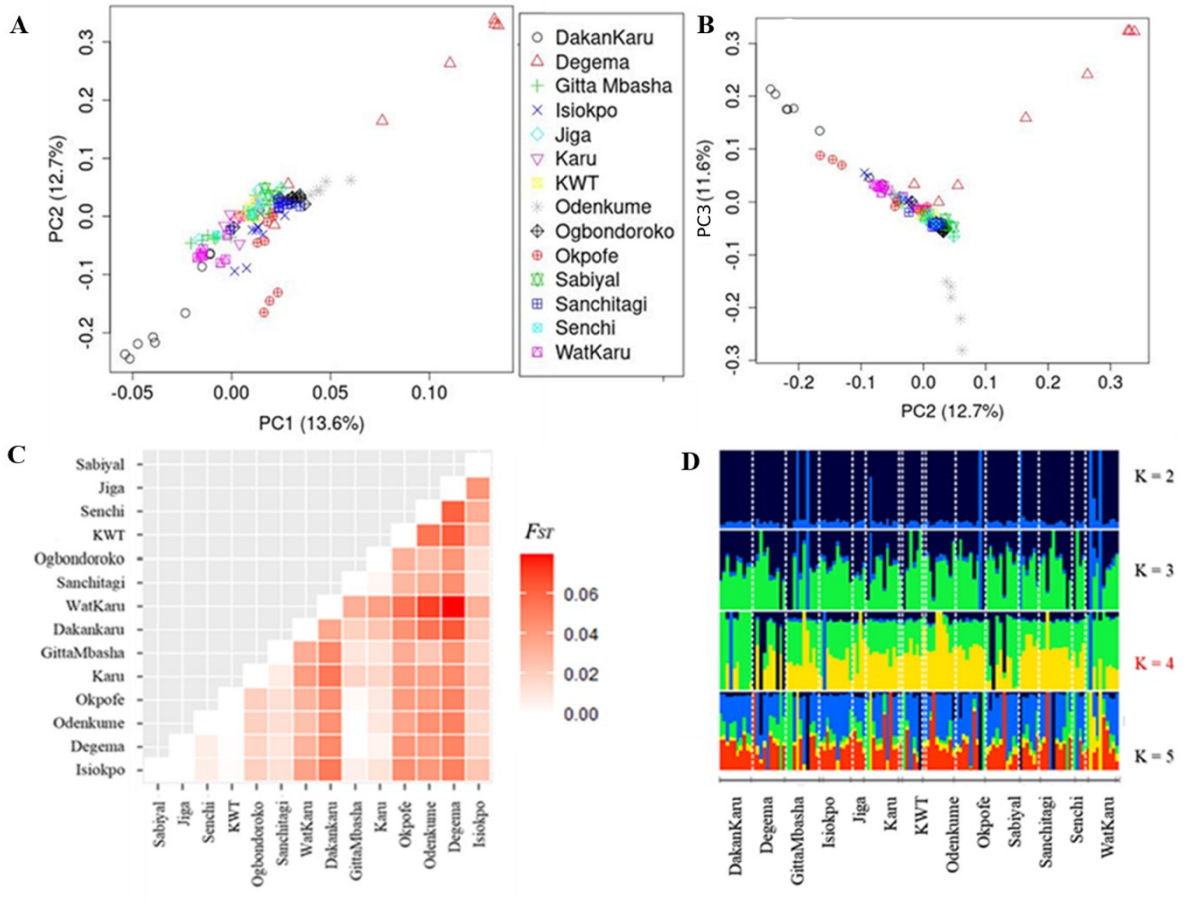


Figure 2: Population structure and genetic diversity: (A & B) PCA plots showing population structure, (C) heat map of pairwise *Fst* values, (D) admixture analysis (K = 4 is the best prediction).

Genome-wide linkage disequilibrium (LD) structure confirms large genetic diversity

LD decay analysis in our study shows that LD (r^2) in NIC genomes drops rapidly with distance. The LD (r^2) values ranged from 0.20 to 0.30 in 10 Kbp distance for all chromosomes except for the Z-chromosome, which has the slowest LD decay rate. LD decay was estimated in four groups of chickens (10-90 samples in each group) identified based on the PCA results with tightly clustered populations considered as a single group. In Figure 3, we show the LD decay plot of one large group (Group 4 = 90 samples) as representative of all populations (Table S1 and Figure S1).

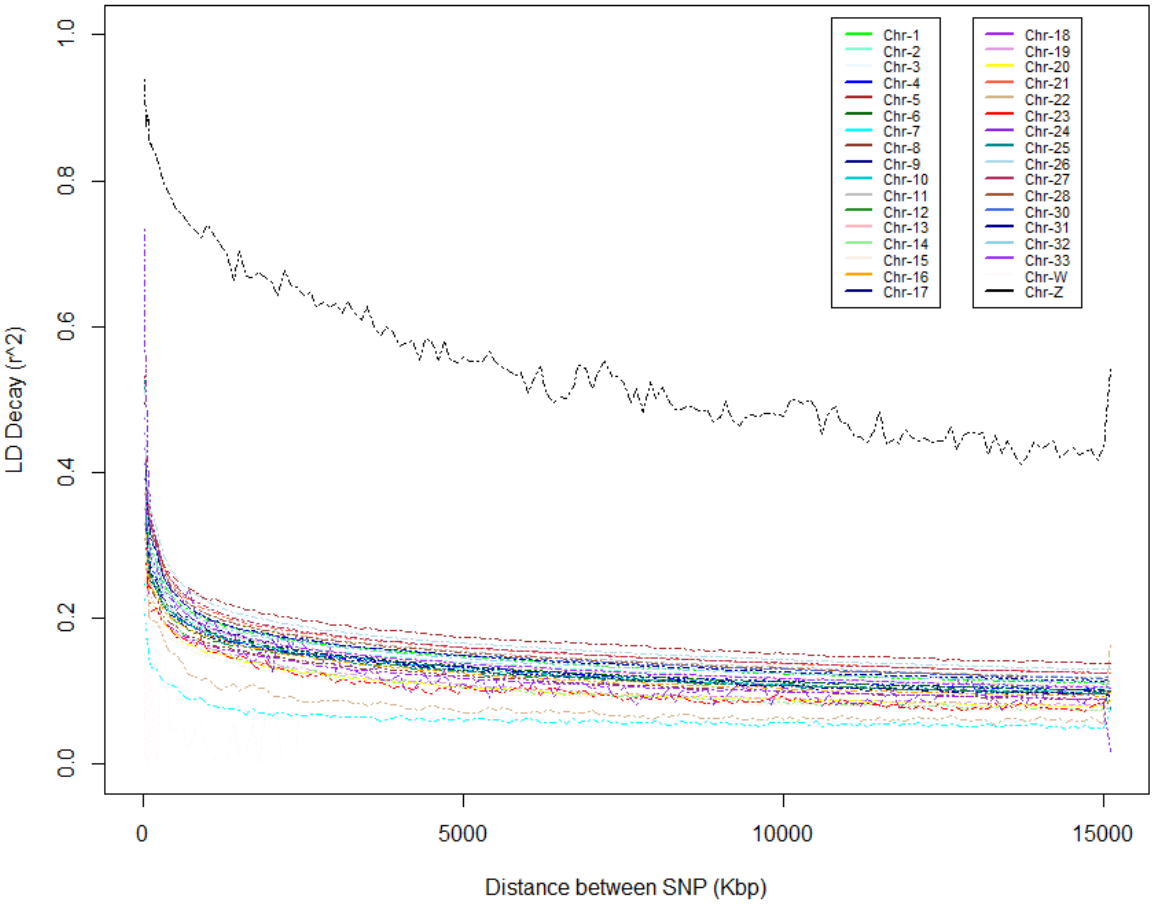


Figure 3: Chromosome-wise LD decay in a representative Nigerian chicken population groups.

Identification of genome-wide selective sweep signals

Since the geographical regions representing the studied populations showed little variation in annual mean temperature (Figure 1C) but showed a large variation in annual rainfall (Figure 1D), selection signature analyses (SSA) to uncover heat-stress adaptation were undertaken using two different approaches.

In the first approach, the genomes of 12 of the 14 populations (omitting Wat-Karu and Dakan-Karu from the Mid-altitude region with lower temperatures) were combined as a single population. A within-population genomic search for low heterozygosity was performed using the "Pooled Heterozygosity" (*Hp*) approach described by Rubin et al.[12] and was done in

overlapping sliding windows of 20 Kb size and a 10 Kb step size. The goal of this analysis was to identify candidate genomic regions that are putatively under selection for heat-stress adaptation irrespective of other environmental conditions (e.g., arid or humid conditions or different agroecological conditions). The combined analysis of many populations allowed the reduction of spurious signals from any population structure and the detection of genomic regions with extremely low heterozygosity ($ZHp \leq -4$) for the all hot-climate populations (Table S2, S3 and Figure 4, S2).

The second approach entailed a comparison of population groups from extreme hot-humid (Degema and Isiokpo) and extreme hot-arid (Jiga and Sabiyal) climates (Figure 1D) to identify candidate regions that show large differentiation either in allele frequency spectrum (using *Fst* method) or LD pattern (using *XP-EHH* method) (Figure 5). Again, two populations were combined in each extreme group to reduce any population structure effect and the analyses were performed using overlapping sliding windows of the same size as above. The rationale behind using two approaches for cross-group comparison was to gain more confidence in the results as these analyses had a much smaller sample size (10 samples per group) than the *Hp* approach used above. Candidate windows were generated from selection signature analysis with empirical P-value < 0.01, which had the thresholds as standardized *Fst* (Z*Fst*) > 3.7 or absolute standardized *XP-EHH* (|*XP-EHH*_std|) > 2.6 (Table S2, S4, S5 and Figure 5B, 5C, S3, S4). Only common windows or regions from both analyses were considered as putative selective sweeps (Table S6, S7). *XP-EHH* analysis can indicate the directionality of selection. In our analysis, a positive *XP-EHH* value indicated selection in the hot-arid group while a negative value indicated selection in the hot-humid group (Table S6).

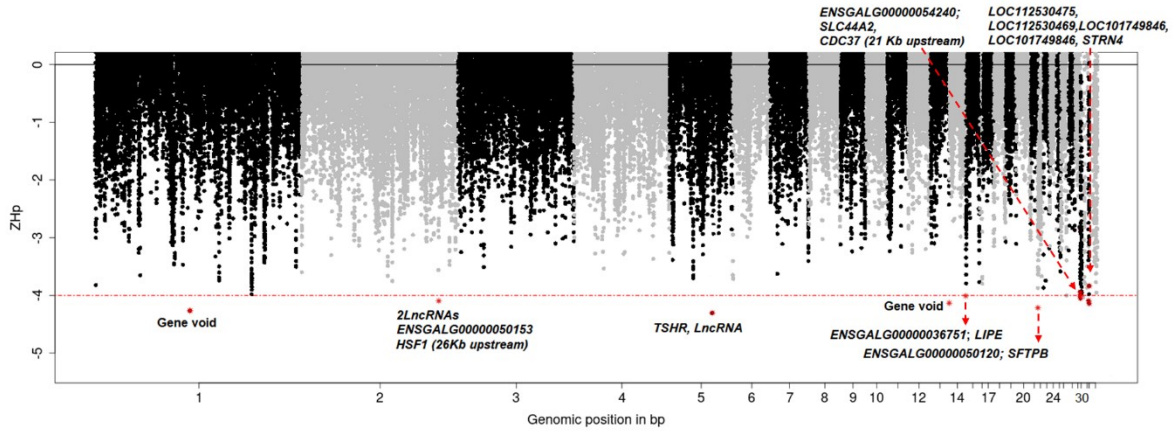


Figure 4: Manhattan plots from *Hp* analysis based on 12 hot climate populations.

Candidate loci and genes for heat-stress adaptation irrespective of the agro-ecologies

A total of 92,936 windows were analysed. Only 11 windows (0.02%) passed the genome-wide significance threshold of $ZHp \leq -4$ with mean of $Hp = 0.022$ (Figure S2). These candidate selective sweep windows are located on chromosomes 1, 2, 5, 14, 22, 30, and 32 and overlap with 16 genes (except two windows which are gene-void) (Table 1 and S3). These overlapping genes – to be considered here as the candidate genes under positive selection - have highly relevant biological functions (Table 1) associated with thermotolerance, e.g., *TSHR* – has a role in thermogenesis (possibly regulated by an epigenetic mechanism, as fixed in most chickens[13,14]. *RYR1*-like genes (*LOC112530475* and *LOC101748756*) – involved in hyperthermia, *CYP450*-like genes (*LOC112530469* and *LOC101749846*) – role in oxidative stress response, *SFTPB* – role in respiratory gaseous exchange (affecting heat loss from the body), *LIPE* - lipid metabolism, *STRN4* and *SLC44A2* – roles in nervous system processes and immunity, and two long non-coding RNAs (lncRNAs) with possible cis-regulatory effect on the nearby Heat Shock Factor 1 (*HSF1*) gene[15,16,17,18]. The analysis of GO terms and KEGG pathways associated with these genes sheds further light on their potential relevance to heat stress adaptation (Table S8). For instance, *LIPE, LOC101749846,* and *LOC112530469* are involved in the glycerolipid catabolic process (GO:0046503) and organic acid metabolic process (GO:0006082), *LOC101749846* and *LOC112530469* are also associated with the molecular function GO term 'oxidoreductase activity' (GO:0016491), *TSHR* is involved in neuroactive ligand-receptor interaction (gga04080) and *LOC101748756* is involved in the calcium signalling pathway (gga04020) (Table S8).

Table 1. Candidate windows and genes under positive selection signatures in Nigerian indigenous chickens in relation to heat stress adaptation based on *Hp* analysis.

Sweep Regions	Candidate genes and functions
Chr 2:131020000-131040000	Contains multiple lncRNAs with possible cis-regulatory function on the nearby gene (*HSF1*). *HSF1* (Heat Shock Factor 1): 26 kb upstream, encodes a transcription factor that is rapidly induced after temperature stress and binds heat shock promoter elements[18,19]
Chr 5:41000000-41020000	*TSHR*: Role in thermogenesis[20,21]

Chr 14:16010000-16030000	*LIPE*: Hormone-sensitive lipase activity[22,23]
Chr 22:40000-60000	*SFTPB*: Pulmonary surfactant protein; role in respiratory gaseous exchange[24,25]
Chr 30:10000-30000	*SLC44A2*: Choline transport (important for the nervous system); involved in positive regulation of I-kappaB kinase[17,22,26] Nearby gene: *ILF3* (16 kb downstream) has a role in chronic stress adaptation[27]
Chr 30:200000-220000	*LOC107050992*: iron-sulpher binding and electron transfer activity[17,22] Nearby gene: *CDC37* (21 kb upstream)[28,29]
Chr 32:0-20000	*LOC112530475* & *LOC101748756*: both RYR1 like genes; RYR1 is involved in calcium channel activity and calmodulin-binding[17,22]
Chr 32:590000-610000	*LOC112530469* & *LOC101749846*: Both are cytochrome P450 like genes; oxidoreductase activity and heme-binding[17,22] *STRN4*: Calmodulin binding & calcium channel activities[22] Nearby gene: *HIF3A* (16 kb downstream)[30,31]

Candidate loci and genes for heat-stress adaptation in the hot-arid condition

Twenty-eight putative sweep regions were commonly detected by both *Fst* and *XP-EHH* analyses. The size of these regions ranged between 20 kb and 140 kb (average 35 kb), with a combined total length of 1,100 kb and an average length of 35 kb. These regions are distributed across chromosomes 1, 2, 3, 5, 6, 7, 8, 11, and 19 (Table S6-S7). Chromosome 1 has the longest length of selection signature region, and chromosomes 11 and 19 the shortest length (20 kb) compared to other regions. A total of 34 genes are found within these regions including 10 long non-coding RNA (lncRNA), and 21 protein-coding genes, with five regions being gene deserts (Table 2, Table S6, Figure 5B). Table 2 shows that the overlapped genes are involved in cytokine activities, inflammatory responses, and immune responses (e.g., *TAFA5, TRIM24, AGR2, CHID1, ARG2 CKLF, CLTM3, SUPT4H1*) and include a sub-set of protein-coding genes with highly relevant stress response functions. Those genes also showed certain GO terms, pathways, and QTLs that relate to disease resistance and immune responses (Table S9, Figure S5-S6). Other gene functions include those of the nervous system (*DPY19L1* and *BICD1*), calcium ion transport (*TSPAN13*), abdominal fat deposition (*KALRN)* and bone formation (*MRPS18*).

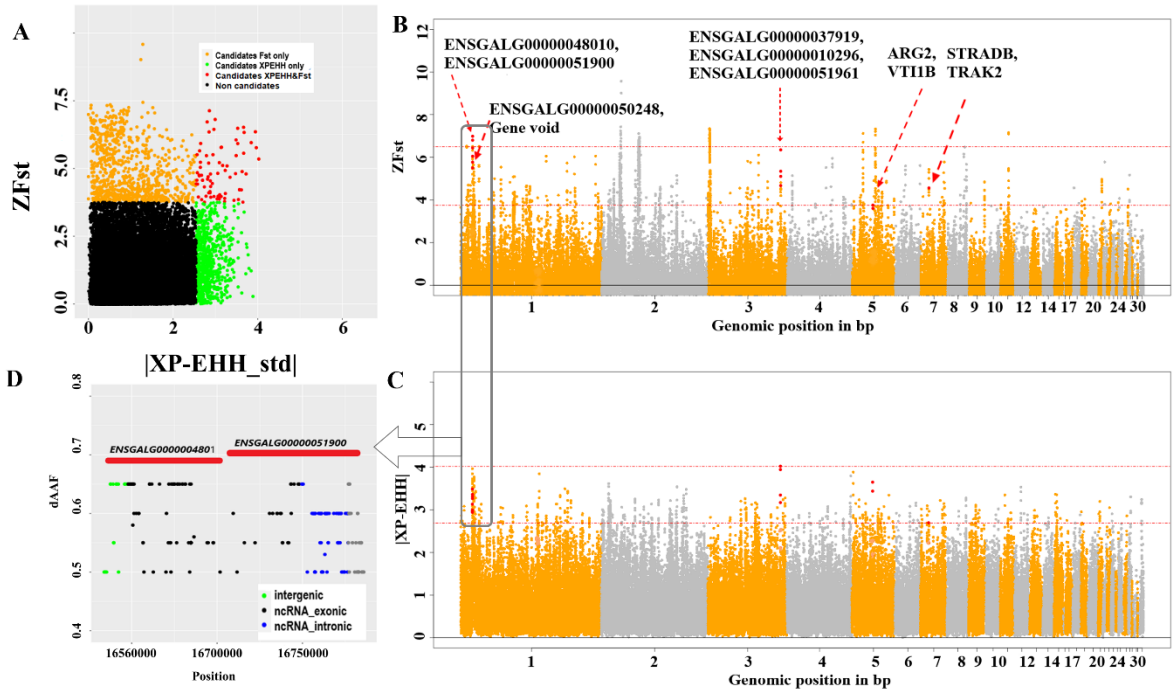


Figure 5: Selection signature analysis results for birds from regions of high and low precipitation. (A) Scatter plot of standardized values of *XP-EHH* versus *Fst*. (B, C) Manhattan plots for the *Fst* and *XP-EHH* analyses; common windows are marked with an asterisk along with gene names from common windows; the red dashed line represents *ZFst* and |*XP-EHH*| threshold. (D) Closer look at the common *Fst*/*XP-EHH* region - chr1:16,630,000 -16,790,000 with SNPs showing allele frequency difference (dAAF) > 0.5 between the hot arid and hot humid groups.

Table 2. **Candidate genes under positive selection signatures in hot-arid conditions**

Sweep Regions	**Candidate genes and functions**
Chr1: 17600000-17680000	*TAFA5:* role in cytokine activity[22]
Chr1: 56420000-56440000	*TRIM24:* involved in cytokine pathways and the inflammatory response[32,33]
Chr1: 59080000-59110000	*BICD1:* participates in the development and function of the nervous system[34]
Chr2: 28560000-28580000	*TSPAN13:* involved in the regulation of calcium ion transmembrane transport[22] *AGR2:* involved in the inflammatory response[22]
Chr2: 47230000-47260000	*DPY19L1:* role in neuronal migration in the developing mouse cerebral cortex[35]

Chr3: 30900000-30920000	*MRPS18A:* cortical bone formation[36]
Chr5: 15250000-15290000	*CHID1:* involved in negative regulation of cytokine production; inflammatory response[22]
Chr5: 15320000-15380000	*MUC6:* involved in the maintenance of gastrointestinal epithelium; intestinal barrier function in chicken[22,37]
Chr5: 29080000-29100000	*ARG2*: anti-inflammation associated[38]
Chr7: 27830000-27850000	*KALRN:* key regulatory role in abdominal fat deposition[39]
Chr7: 32950000-32980000	*ARHGAP15:* involved in signal transduction[22]
Chr8: 27480000-27510000	*NFIA:* necessary for articular cartilage differentiation[40]
Chr11: 11440000-11460000	*CMTM3:* involved in positive regulation of B cell receptor signalling pathway[22,41] *CKLF:* this may play an important role in the inflammation and regeneration of skeletal muscle[42] *TK2:* kinase activity[22]
Chr19: 710000-730000	*SUPT4H1:* viral infection pathway[43]

Candidate loci and genes for heat-stress adaptation in the hot-humid condition

Only three common selective sweep regions (i.e., regions detected by both *Fst* and *XP-EHH* analysis as candidates) were detected for the hot-humid climate. These are located on chromosomes 1, 5 and 7 (all are 20 kb in size) and overlap with six genes - comprising two lncRNAs and four protein-coding genes (Table 3, Table S6 and S7) involved in immune response (*ARG2, VTI1B* and *STRADB)* and intracellular transport (*TRAK2*). A summary of the molecular functions, biological processes, and pathways is presented in Table S10 and Figure S7 and S8.

Table 3. **Protein coding genes overlapping with putative selection signatures in populations from hot-humid conditions**

Sweep Regions	Relevant Biological Functions for the Candidate Genes
Chr5: 29080000-29100000	*ARG2:* mitochondrial type of arginase that leads to an increase in oxygen-free radical formation and endothelial dysfunction[44,] arginine metabolism is also a critical regulator of innate and adaptive immune responses[45].
Chr7: 11480000-11500000	*VTI1B:* concerned with increased secretion of cytokines associated with cellular senescence[16]. *STRADB*: among its related pathways are MTOR signalling and AMP-activated protein kinase signalling[22]. *TRAK2*: Predicted to be involved in several processes, e.g., mitochondrion distribution; organelle transport along the microtubule and protein targeting[22].

Discussion

This is the first study performing a large-scale WGS analysis on NICs to assess genetic diversity and identify genomic signatures of adaptive selection in relation to hot (humid or arid) climates. The study has high coverage of chicken populations representing the diversity of the Nigerian landscape, providing the opportunity to investigate both within and between population genetic diversity, as well as adaptive diversity. Our study generates and utilizes a powerful and robust variant dataset by jointly analysing a large number of samples (n=120) using established bioinformatic workflow of GATK. Such joint analysis is known to improve the sensitivity and accuracy of the detection greatly[46]. Therefore, the variant dataset along with the WGS data contributes major genomic resources for further research on chicken.

By comparing SNPs from NICs and those available in public databases as well as those recently detected in Ethiopian indigenous chickens by our group[47], we have detected ~11% novel chicken variants. Our study reveals large within-population genetic diversity in NICs, but a low level of genetic differentiation between populations. This result corroborates the findings of previous diversity studies based on mitochondrial DNA (mtDNA) D-loop in NICs, which reported that, currently, all the sequences belong to a single clade or haplogroup, predominantly found in South Asia (Indian subcontinent). It supports a single geographic origin in Asia (Indian subcontinent) and suggest a extensive genetic intermixing within the country, thus resulting in a lack of mtDNA phylogeographic structure among the NICs[10,48].

The value and pattern of LD decay detected in the present study is similar to that observed in a previous study of Korean native chicken[49], and some Chinese indigenous chicken[50] (Wenchang, Beijing You, Taihe Silkies, and Shouguang). They showed a rapid decay in LD structure that is generally common in local breeds or populations that experienced less intensive breeding programs compared to commercial chicken breeds, [51].Genomic LD structure can be affected by various factors including effective population size, non-random mating, admixture, genetic drift, selection, mutation, and recombination rate. The rate of LD decay can therefore be used to measure the evolutionary history of populations[52] and are also helpful for determining the resolution of association mapping or assessing the desired number of SNPs to be used for genome-wide association analysis.

A major focus of our study was to identify candidate genes and pathways related to thermotolerance in the NICs. Important candidate genes and their involvement in thermoregulation can be summarized schematically as Figure 6. Chicken activates thermoregulation mechanisms to lose heat when the environmental temperature is above the thermoneutral zone by showing three types of responses: behavioural, biochemical, and physiological. It is also notable that adaptive response to heat stress occurs not only with high temperatures but is also affected by the relative humidity of the environment.

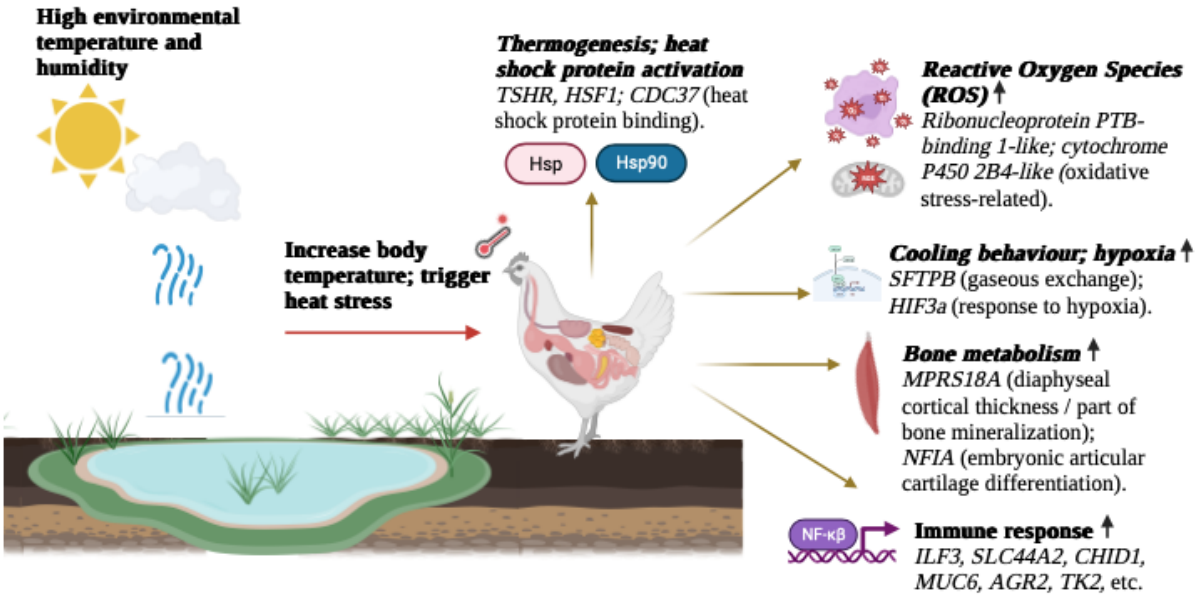


Figure 6: Summary of the main effects of heat stress on Nigerian indigenous chickens. (Created with BioRender.com).

312 The *TSHR* gene, detected in the *Hp* analysis, overlapped with the strongest peak (*ZHp*= - 4.3)
313 and is considered here as a major candidate gene for tropical heat-stress adaptation
314 irrespective of arid or humid conditions. Many studies have demonstrated that *TSHR* may be
315 involved in reproduction, regulating energy balance, metabolism, and thermoregulation[53,54].
316 Moreover, recent studies found that it functionally contributed to the chicken response and
317 adaptation to hot and tropical environments[21,54]. This gene also has been detected in selection
318 signature analyses from most domestic chicken populations[12,55]. The potential role of
319 epigenetic regulation of thermotolerance in chicken has also been expressed by Karlsson et
320 al.[57] and Gheyas et al.[47] The study from Guo et al.[21] reported that a missense mutation in the
321 *TSHR* gene might regulate the metabolic rate to enhance heat tolerance and contribute to
322 chicken adaptation to high ambient temperature in tropical climates.

323 In this study, several lncRNAs were found to overlap with putative sweep regions. LncRNAs
324 have potential regulatory functions on gene expression exerted in either a cis-acting or trans-
325 acting manner on their target genes. Cis-acting lncRNAs regulate the expression of target
326 genes that are located at or near the same genomic locus while trans-acting lncRNAs either
327 inhibit or activate gene transcription at independent chromosomal loci[58]. In chickens,
328 lncRNAs have been reported to regulate muscle development, lipid metabolism, egg
329 production and disease resistance[59,60]. In our study, the heat shock transcription factor 1
330 (*HSF1*) gene, is found 26 kb upstream of a candidate region on chromosome 2 which
331 overlaps with three lncRNAs. This gene functions as a stress-inducible and DNA-binding
332 transcription factor that plays a role in the transcriptional activation of heat shock response
333 (HSR) leading to the expression of a large class of molecular chaperones of heat shock
334 proteins (HSPs) that protect cells from cellular damage in the chicken[18,61]. Consequently,
335 *HSF1* is associated with several gene ontology terms like 'cellular response to heat
336 (GO:0034605)' and 'heat shock protein binding (GO:0031072)'[17] and has been proposed as a
337 marker during acute heat stress in chickens[61]. Another gene, *CDC37* (Cell Division Cycle 37)
338 which is located 21 kb upstream of another candidate region on chromosome 30 also overlaps
339 with a lncRNA. *CDC37* has the molecular function of 'heat shock protein binding
340 (GO:0031072)' and probably acts as a co-chaperone of *HSP90* or non-client protein binding
341 partner that also assists in repairing denatured proteins or promoting their degradation caused
342 by heat stress[29].
343

Previous studies have explored the links between putative novel lncRNAs and previously reported QTLs; for instance, a novel lncRNA (*LncFAM*) was found located in a chicken growth QTL[62], while another study revealed the association of certain lncRNAs with response to Marek's Disease Virus (MDV) in commercial egg production lines[63]. In our study, we found regions from *Hp* and shared *Fst-XPEHH* analyses that overlapped with both lncRNAs and QTLs (related to immune response, fear behaviour, and Mareks's disease susceptibility (Table S6). However, the role and function of these lncRNAs with nearby protein-coding genes and the overlapping QTLs in relation to heat stress adaptation still requires further investigation.

Various kinds of stress, including extreme environmental temperatures, lead to the generation of reactive oxygen species (ROS), causing oxidative stress and lipid peroxidation. Through *Hp* analysis, we have found several candidate genes on chromosomes 30 and 32 (*LOC107050992* (ribonucleoprotein PTB-binding 1-like), and cytochrome P450 2B4-like*: LOC112530469* & *LOC101749846*) which are associated with oxidative stress. These genes are involved in several GO terms[17] associated with oxidoreductase activity (GO:0016491) and oxidation-reduction process (GO:0055114). Oxidative stress from heat exposure can manifest in all parts of the body, but mitochondrial dysfunction is central to oxidative stress. In the initial acute heat stress phase, mitochondrial substrate oxidation and electron transport chain activity are increased, resulting in excessive superoxide (a type of ROS) production[64]. During gaseous exchange in heat stress at high ambient temperature, a bird's respiratory rate is enhanced to dissipate heat. From the *Hp* analysis (of combined population) we found a sweep region on chromosome 22 that overlaps with the Surfactant Protein B (*SFTPB*) gene which is involved in the biological process of gaseous exchange between an organism and its environment (GO:0007585). The respiratory system of birds exposed to heat stress operates both for gaseous exchange and as the evaporative cooling system[65].

A study by Varasteh et al.[66] reported that chicken's critical adaptive response to heat stress increases the peripheral blood flow, resulting in reduced blood supply in the intestines and a hypoxia-induced oxidative stress response. A sweep region detected on chromosome 32 from the *Hp* analysis overlaps with gene *LOC101749846* and is nearby to *HIF3a* (Hypoxia-inducible factor 3 subunit alpha), which is involved in the epoxygenase P450 pathway (GO:0019373), response to hypoxia (GO:0001666) as well as angiogenesis (GO:0001525) in

chicken[17]. This finding is in line with the study from Zahoor et al.[67] who reported angiogenic pathways are involved in hypoxia-induced angiogenesis in chickens.

One of the most noticeable developmental problems associated with heat stress in poultry is a pronounced induction of leg abnormalities, as shown in broilers, layers, and turkeys[68]. Elevated temperatures impair gut integrity, thereby increasing systemic inflammation that elicits osteoclastic bone resorption that is related to bone metabolism such as cortical thickness in diaphysis in tibia bone as suggested by Zhang et al[68]. The *MRPS18A* gene, which overlaps with a sweep region from the arid region population was found to be a candidate for the diaphyseal cortical thickness (part of bone mineralization) that is associated with heat shock factors[36] which are involved in bone formation. *MPRS18A* encodes a mitochondrial ribosomal protein. Since mitochondrial function is crucial for cellular metabolism, such mutation can therefore affect the functions of diverse organ and tissue systems, including bone and active osteoclasts that are rich in mitochondria[36].

Immunity is suppressed under heat stress conditions as has been observed in previous studies[69,70]. To protect the body from the adverse effects of heat stress, a defence mechanism is activated in chickens. Initially, the early response system stimulates the central nervous system, and eventually, the immune system is involved. Based on *Hp* and shared *Fst*-XP-EEH regions we have found two genes (*SLC44A2* and *BICD1*, respectively) which participate in the function of the nervous system[22,34]. These genes are potentially involved in immunity as the central nervous system modulates immune responses, which are mediated by a complex network of signals. These signals interplay across the nervous, endocrine, and immune systems which affect metabolism and immune responses[71]. Many immunity-related genes have been detected in all three selection signature analyses. Some of the genes located nearby to the sweep regions based on *Hp* analysis are immune-related genes that might have cis-acting or trans-acting interaction with their target genes. These include Interleukin Enhancer Binding Factor 3 (*ILF3*), and Solute Carrier Family 44 Member 2 (*SLC44A2*) - both contributing to the negative regulation of viral genome replication (GO:0045071) and are involved in innate immune system pathways and positive regulation of I-kappa B kinase/IκKβ (GO:0043123). Several genes related to immune response have also been found in the common sweeps from *Fst* and *XP-EHH* analyses, for instance, the cytokine-related genes: *TAFA5, TRIM24, CHID1, MUC6* (from arid climate)*; VTI1B* (from humid climate); infection and inflammation-related genes: *AGR2, TK2, CMTM3, SUPT4H1* (arid climate);

SETD4, ARG2, VTI1B and *STRADB* (humid climate). Heat stress significantly impacts the immunity and cytokine expression of chickens. Heat stress was found to modulate the gene expression of a range of different cytokines in chickens and many studies have demonstrated that bacteria are exploiting neuroendocrine alterations following stress response in the host to promote growth and pathogenicity.

Conclusions

This study has generated and characterized over 17 million high-quality genome-wide SNPs, of which ~11% are novel variants. These large numbers of SNPs provide an additional resource for future applications and characterization of NICs. The identification of candidate genes/genomic regions under selection will help in understanding their evolution and functional roles in relation to environmental challenges. The small number of highly plausible candidate genes detected for hot climate adaptation, are seen to be involved in relevant biological processes and pathways related to oxidative stress (e.g., *cytochrome P450 2B4-like: LOC112530469* & *LOC101749846, SFTPB)*, cellular responses to heat and hypoxia (e.g., *HSF1*, *CDC37*, *HIF3A*), transcriptional regulation (*TSHR*), immune response (*SLC44A2, ILF3*), and metabolic activities – all of which are important for thermal adaptation. This study also enhances our understanding of the role of natural selection in shaping the genome of NICs for adaptation to both hot-arid and hot-humid tropical conditions. Apart from the genetic adaptation, this study dissects the within and between-population genetic diversity in Nigerian indigenous chicken populations. Understanding genetic diversity is a prerequisite in setting up an effective breeding program and selecting the population to use. Our study shows that there is large genetic diversity within Nigerian chicken populations that can be harnessed for breeding improvement of locally adapted birds.

Taken together, these findings will help guide the improvement of indigenous chickens by helping design specific breeding programs as well as poultry management strategies to minimise heat stress and enhance disease resistance and productive performance. This will increase the contribution of poultry products to global food security.

Materials & Methods

Chicken sampling

Sampling was performed to represent diverse Nigerian agro-climatic zones. From each zone, two villages were selected and from each village 4 - 10 scavenging chickens were sampled by drawing blood (50 - 250 µl) from the wing vein with the logistical support and agreement of the Department of Animal Sciences, Obafemi Awolowo University (Ife Ife, Nigeria). All animal works were approved by the Institutional Animal Care and Use Committee of the International Livestock Research Institute (IREC2017-26) and were handled strictly in compliance with the guidelines of this committee. A geographic coordinate (latitude and longitude) was collected for each sampled village for the extraction of environmental data from public databases.

All the collected blood samples were processed for DNA using the Qiagen DNeasy blood and tissue kit protocol (https://www.qiagen.com/ca/resources/download.aspx?id=63e22fd7-6eed-4bcb-8097-7ec77bcd4de6&lang=en). The genomic DNA (gDNA) from each sample was then normalized to a final volume of 100 µl and final concentration of 50 ng/µl and was sent to Edinburgh Genomics (http://genomics.ed.ac.uk/) in the UK for whole-genome sequencing.

Sequencing and variant calling

Whole genome sequencing was performed on the Illumina HiSeqX platform, with an average 64x paired-end coverage. Sequence reads were mapped against the chicken reference genome (GRCg6a) (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_000002315.4/) using the BWA-mem algorithm[72]. Variant calling, filtration, and genotyping were performed by combining all 120 samples together following the best practice protocol of the GATK package for "Germline short variant discovery"[73], involving the Haplotype Caller method and Joint Genotyping of all samples together. Initial filtration on SNP calling was performed using the GATK's machine learning algorithm, the VQSR (Variant Calling Score Recalibration) approach for which 1M validated chicken SNPs were used as a 'training' and 'true' set and ~20M publicly available SNPs from the Ensembl database was used as a 'known' set. Further filtration on SNPs was applied before using for genomic analysis as: biallelic SNPs, minor allele frequency > 0.05, genotype quality > 15.0, depth of coverage > 3, maximum missing genotype rate < 20% and Hardy-Weinberg-Equilibrium (HWE)

probability < $1x10^{-7}$. Only autosomal SNPs were used for genetic diversity and selection signature analyses (SSA). Quality checks for samples were performed based on genotype missing rate and relatedness analysis in VCFtools v0.1.15[74]. Thus an individual pair showing higher than expected relatedness (> 0.9) was removed. For downstream analyses, variants were extracted for individual populations using option *"gatk SelectVariants"* from GATK (https://gatk.broadinstitute.org/hc/en-us/articles/13832694334235-SelectVariants).

Genetic diversity analysis

Inbreeding coefficient for individual chickens, nucleotide diversity (π or "pi") for the individual chicken population, and pairwise population-differentiation (*Fst*) between populations were calculated using the filtered SNP variant set in VCFtools (v0.1.15). The average genome nucleotide diversity and *Fst* were estimated in 20 kb windows with 10 kb sliding steps. Population structure among the investigated populations was inferred with PCA using 'smartpca' in 'eigenstrat' version 6.0.1[75]. The proportion of ancestry (admixture) in each individual and population was estimated using ADMIXTURE version 1.3.0[76] considering K values from 1 to 5 with the lowest cross-validation error used to choose the best K value. The PCA analysis was performed with an LD-pruned set of SNPs consisting of about 4M variants. LD pruning was performed in Plink (v1.9) (https://www.cog-genomics.org/plink/1.9/ld) with the options *"--indep-pairwise 50 5 0.5"*. For admixture analysis, a 30% thinning of the SNPs was performed after the LD pruning to reduce the computation burden (retaining 583,700 SNPs). A pairwise r^2 estimation was used to measure LD between pairs of SNPs within a chromosome using the PopLDdecay (v3.40) program[77]. SNPs on both autosomal and sex chromosomes that passed the quality control using options "-MAF 0.05" (minimum minor allele frequency of 0.05) and "-MaxDist 15" (maximum window bin 15 kb) were used. The decay of LD was plotted using the" ggplot" package (https://ggplot2-tidyverse-org) in Rstudio version 3.4.3.

Selection signature analyses

Selection signature analysis (SSA) was performed using Pooled Heterozygosity (*Hp*)[12], *Fst*[78] and *XP-EHH*[79] approaches in overlapping sliding windows (20 kb size with 10 kb step) with at least 10 SNPs/window from the combination of multiple populations (Table S2). The weighted *Fst* values were standardized (Z*Fst*) to allow the setting of the same threshold across analyses. *XP-EHH* analyses were carried out using the Hapbin package[80] after removing SNPs with missing genotypes. *XP-EHH* analyses were first performed for

individual SNPs and then mean values were calculated within windows for both the standardized *XP-EHH* (*XP-EHH*_std) and the absolute value of *XP-EHH*_std. SSA windows with empirical P-value < 0.01 were considered as putative selective sweeps for a standardized *Fst* (Z*Fst*) > 5 or an absolute standardized *XP-EHH* (|*XP-EHH*_std|) > 3.5. Moreover, since the positive and negative values of *XP-EHH* indicate the directionality of selection, all SNPs within an *XP-EHH*-based candidate window are needed to show the same directionality.

Function analysis of candidate gene and SNPs

Bedtools version 2.25.0[81] was used to merge the overlapping selected windows. Chicken genes that overlapped genomic windows passing the significant selective sweep threshold were retrieved from the Ensembl Genes 106 database using the Biomart online tool (http://www.ensembl.org/biomart). The candidate genes were then processed in a web-based PANTHER Classification System[82] and KEGG Pathway Database (https://www.genome.jp/kegg/pathway.html) to map the candidate genes to known biological processes, molecular function, cellular processes, and molecular pathways. Candidate genes were also checked for their overlap with known chicken QTLs (ChickenQTLdb: https://www.animalgenome.org/cgi-bin/QTLdb/GG/index).

References

1. FMARD, The World Bank. Nigeria livestock roadmap for productivity and resilience 2020-2026. http://ngfrepository.org.ng:8080/jspui/handle/123456789/3999 (2020).

2. CSIRO. Poultry production in Nigeria. https://research.csiro.au/livegaps/wp-content/uploads/sites/37/2021/03/1.-LiveGAPS-factsheet-Poultry-production-in-Nigeria-22-April-2020.pdf (2020).

3. Food and Agriculture Organization. Livestock and livelihoods spotlight Nigeria cattle and poultry sectors. https://www.fao.org/documents/card/en/c/CA2149EN/ (2018).

4. Ikpeme, E. *et al*. Genetic relationship among three Nigerian chicken (Gallus *gallus*) genotypes based on cytochrome-b of mitochondrial DNA. *Asian Journal of Animal Sciences*, **15**, 35-42. https://doi.org/10.3923/ajas.2021.35.42 (2020).

5. Bett, H. K., Peters, K. J., Nwankwo, U. M. & Bokelmann, W. Estimating consumer preferences and willingness to pay for the underutilized indigenous chicken products. *Food Policy*, **41**, 218-225. https://doi.org/10.1016/j.foodpol.2013.05.012 (2013).

6. Federal Department of Forestry. National Forest Reference Emission Level (FREL) for the Federal Republic of Nigeria. Abuja, Nigeria. https://redd.unfccc.int/files/nigeria_national_frel_modified_revised__for_posting.pdf (2019).

7. Sowunmi, F. Effect of climatic variability on maize production in Nigeria. *Research Journal of Environmental and Earth Sciences*. **2**(1), 19-30. https://maxwellsci.com/print/rjees/v2-19-30.pdf (2010).

8. Nwogwugwu, C. P., Lee, J., Freedom, E. C. & Lee, S. Review on the genetic potential of Nigerian local chickens. *Journal of Animal Breeding and Genomics*, **2**(2). http://www.jabg.org/view/N0270020401.pdf (2018).

9. Adetayo, A. S., & Sonaiya, E.B. (2001). Comparison of the performance of Nigerian indigenous chickens from three agroecological zones. *Livestock Research for Rural Development*, **13**(2). http://www.lrrd.org/lrrd13/2/aded132.htm (2001).

10. Ajibike, A. B. *et al*. Genetic diversity, phylogeographic structure and effect of selection at the mitochondrial hypervariable region of Nigerian chicken populations. *Journal of genetics*, **96**(6), 959–968. https://doi.org/10.1007/s12041-017-0860-1 (2017).

11. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**, 4302-4315. https://doi.org/10.1002/joc.5086 (2017).

12. Rubin, C. J. *et al*. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, **464** (7288), 587–591. https://doi.org/10.1038/nature08832 (2010).

13. Gheyas, A. A. *et al*. Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA research: an international journal for rapid publication of reports on genes and genomes*, **22**(3), 205–217. https://doi.org/10.1093/dnares/dsv005 (2015).

14. Vallejo-Trujillo, A. *et al*. Ecological niche modelling for delineating livestock ecotypes and exploring environmental genomic adaptation: The example of Ethiopian village chicken. *Frontiers in Ecology and Evolution*, **10**. https://doi.org/10.3389/fevo.2022.866587 (2022).

15. Seo, G., Han, H., Vargas, R. E., Yang, B., Li, X., & Wang, W. MAP4K interactome reveals STRN4 as a key STRIPAK complex component in hippo pathway regulation. *Cell reports*, **32**(1), 107860. https://doi.org/10.1016/j.celrep.2020.107860 (2020).

16. Uniprot. Available at: https://www.uniprot.org/.

17. Biomart. Available at: http://asia.ensembl.org/info/data/biomart/index.html.

18. Xu, J., Yin, B., Huang, B., Tang, S., Zhang, X., Sun, J., & Bao, E. Co-enzyme Q10 protects chicken hearts from in vivo heat stress via inducing HSF1 binding activity and Hsp70 expression. *Poultry science*, **98**(2), 1002–1011. https://doi.org/10.3382/ps/pey498 (2019).

19. ENTREZ. HSF1. Available from: https://www.ncbi.nlm.nih.gov/gene/3297.

20. Xie, S. *et al*. Thyroid transcriptome analysis reveals different adaptive responses to cold environmental conditions between two chicken breeds. *PloS one*, **13**(1), e0191096. https://doi.org/10.1371/journal.pone.0191096 (2018).

21. Guo, X. *et al*. Genome-wide scan for selection signatures and genes related to heat tolerance in domestic chickens in the tropical and temperate regions in Asia. *Poultry science*, **101**(7), 101821. https://doi.org/10.1016/j.psj.2022.101821. (2022).

22. Genecards. Available at: https://www.genecards.org/.

23. Kong, B. W. *et al*. RNA sequencing for global gene expression associated with muscle growth in a single male modern broiler line compared to a foundational Barred Plymouth Rock chicken line. *BMC genomics*, **18**(1), 82. https://doi.org/10.1186/s12864-016-3471-y (2017).

24. To, K. K. W. *et al*. Surfactant protein B gene polymorphism is associated with severe influenza. *Chest*, **145**(6), 1237–1243. https://doi.org/10.1378/chest.13-1651 (2014).

25. Lin, Z. *et al*. Genetic association of pulmonary surfactant protein senes, SFTPA1, SFTPA2, SFTPB, SFTPC, and SFTPD with cystic fibrosis. *Frontiers in immunology*, ***9***, 2256. https://doi.org/10.3389/fimmu.2018.02256 (2018).

26. Cindrova-Davies, T. *et al*. RNA-seq reveals conservation of function among the yolk sacs of human, mouse, and chicken. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(24), E4753–E4761. https://doi.org/10.1073/pnas.1702560114 (2017).

27. Yamashita, A. *et al*. ILF3 prion-like domain regulates gene expression and fear memory under chronic stress. *iScience*, **26**(3), 106229. https://doi.org/10.1016/j.isci.2023.106229 (2023).

28. Tatebe, H. & Shiozaki, K. Identification of Cdc37 as a novel regulator of the stress-responsive mitogen-activated protein kinase. *Molecular and cellular biology*, **23***,* 5132-5142. https://doi.org/10.1128/MCB.23.15.5132-5142.2003 (2003).

29. Dayalan Naidu, S., & Dinkova-Kostova, A. T. Regulation of the mammalian heat shock factor 1. *The FEBS journal*, **284**(11), 1606–1627. https://doi.org/10.1111/febs.13999 (2017).

30. Paul, C., Teng, S. & Saunders, P. T. K. A single, mild, transient scrotal heat stress causes hypoxia and oxidative stress in mouse testes, which induces germ cell death1. *Biology of Reproduction*, **80**, 913-919. https://doi.org/10.1095/biolreprod.108.071779 (2009).

31. Coimbra-Costa, D., Alva, N., Duran, M., Carbonell, T. & Rama, R. Oxidative stress and apoptosis after acute respiratory hypoxia and reoxygenation in rat brain. *Redox biology*, **12**, 216-225. https://doi.org/10.1016/j.redox.2017.02.014 (2017).

32. Quan, R., Zhu, S., Wei, L. *et al.* Transcriptional profiles in bursal B-lymphoid DT40 cells infected with very virulent infectious bursal disease virus. *Virol J*, **14**, 7. https://doi.org/10.1186/s12985-016-0668-2 (2017).

33. Van Tol, S.; Hage, A.; Giraldo, M.I.; Bharaj, P.; Rajsbaum, R. The TRIMendous Role of TRIMs in Virus–Host Interactions. *Vaccines*, ***5***, 23. https://doi.org/10.3390/vaccines5030023 (2017).

34. Chen, X. *et al*. Population Genomic Sequencing Delineates Global Landscape of Copy Number Variations that Drive Domestication and Breed Formation of in Chicken. *Frontiers in genetics*, **13**, 830393. https://doi.org/10.3389/fgene.2022.830393. (2022).

35. Watanabe K, Bizen N, Sato N, Takebayashi H. Endoplasmic Reticulum-Localized Transmembrane Protein Dpy19L1 Is Required for Neurite Outgrowth. *PloS one*, **11**(12): e0167985. https://doi.org/10.1371/journal.pone.0167985. (2016).

36. Johnsson, M., Jonsson, K. B., Andersson, L., Jensen, P., & Wright, D. Genetic regulation of bone metabolism in the chicken: similarities and differences to Mammalian systems. *PLoS genetics*, **11**(5), e1005250. https://doi.org/10.1371/journal.pgen.1005250. (2015).

37. Slawinska, A. *et al*. Modulation of microbial communities and mucosal gene expression in chicken intestines after galactooligosaccharides delivery In Ovo. *PloS one*, **14**(2), e0212318. https://doi.org/10.1371/journal.pone.0212318 (2019).

38. Verwoolde, M. B. *et al*. Innate immune training and metabolic reprogramming in primary monocytes of broiler and laying hens. *Developmental and comparative immunology*, **114**, 103811. https://doi.org/10.1016/j.dci.2020.103811 (2021).

39. Zhu, Y. T. *et al*. Transcriptome analysis reveals steroid hormones biosynthesis pathway involved in abdominal fat deposition in broilers. *Journal of Integrative Agriculture*. https://doi.org/10.1016/j.jia.2023.04.015 (2023).

40. Singh, P. N. P. *et al*. NFIA and GATA3 are crucial regulators of embryonic articular cartilage differentiation. *Development (Cambridge, England)*, **145**(2), dev156554. https://doi.org/10.1242/dev.156554 (2018).

41. Zhong, J. *et al*. Characterization and expression profile of CMTM3/CKLFSF3. *Journal of biochemistry and molecular biology*, **39**(5), 537–545. https://doi.org/10.5483/bmbrep.2006.39.5.537. (2006).

42. Han, W. *et al*. Molecular cloning and characterization of chemokine-like factor 1 (CKLF1), a novel human cytokine with unique structure and potential chemotactic activity. *The Biochemical journal*, **357**(Pt 1), 127–135. https://doi.org/10.1042/0264-6021:3570127 (2001).

43. Reactome. Available at: https://reactome.org/.

44. Kharrati-Koopaee, H. *et al*. Gene network analysis to determine the effect of hypoxia-associated genes on brain damages and tumorigenesis using an avian model. *Journal, genetic engineering & biotechnology*, **19**(1), 100. https://doi.org/10.1186/s43141-021-00184-5.

45. Geiger, R. *et al*. L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-tumor Activity. *Cell*, **167**(3), 829–842.e13. https://doi.org/10.1016/j.cell.2016.09.031 (2016).

46. Anolles, Derek."The logic of joint calling for germline short variants". *GATK-Broad Institute*, 12, Dec. 2023, https://gatk.broadinstitute.org/hc/en-us/articles/360035890431-The-logic-of-joint-calling-for-germline-short-variants.

47. Gheyas, A.A. *et al*. Integrated environmental and denomic analysis reveals the drivers of local adaptation in African indigenous chickens. *Molecular Biology and Evolution*, **38**, 4268–4285. https://doi.org/10.1093/molbev/msab156 (2021).

48. Adebambo, A. O. *et al*. Lack of phylogeographic structure in Nigerian village chickens revealed by mitochondrial DNA D-loop sequence analysis. *International Journal of Poultry Science*, **9**(5), 503-507. https://doi.org/10.3923/ijps.2010.503.507 (2010).

49. Seo, Dongwon, *et al*. Estimation of linkage disequilibrium and analysis of genetic diversity in Korean chicken lines. *PLoS One,*13(2), e0192063. https://doi.org/10.1371/journal.pone.0192063 (2018).

50. Tian, Shuaishuai, *et al*. Identification of Runs of Homozygosity Islands and Functional Variants in Wenchang Chicken. *Animals,* 13(10), 1645. https://doi.org/10.3390/ani13101645 (2023).

51. Qanbari, Saber. On the Extent of Linkage Disequilibrium in the Genome of Farm Animals. *Frontiers in genetics*, 10, 1304. https://doi.org/10.3389/fgene.2019.01304 (2020).

52. Wragg, D., Mwacharo, J. M., Alcalde, J. A., Hocking, P. M. & Hanotte, O. Analysis of genome-wide structure, diversity and fine mapping of Mendelian traits in traditional and village chickens. *Heredity (Edinb)*, **109**, 6-18. https://doi.org/10.1038/hdy.2012.9 (2012).

53. Jiang, D. *et al*. Thyroid-stimulating hormone inhibits adipose triglyceride lipase in 3T3-L1 adipocytes through the PKA pathway. *PloS one*, **10**(1), e0116439. https://doi.org/10.1371/journal.pone.0116439 (2015).

54. Wang, M. S. *et al*. 863 genomes reveal the origin and domestication of chicken. *Cell research*, **30**(8), 693–701. https://doi.org/10.1038/s41422-020-0349-y (2020).

55. Shi, S. *et al*. Whole genome analyses reveal novel genes associated with chicken adaptation to tropical and frigid environments. *Journal of advanced research*, **47**, 13–25. https://doi.org/10.1016/j.jare.2022.07.005 (2023).

56. Qanbari, S. *et al*. Genetics of adaptation in modern chicken. *PLoS genetics*, **15**(4), e1007989. https://doi.org/10.1371/journal.pgen.1007989 (2019).

57. Karlsson, A. C. *et al*. The effect of a mutation in the thyroid Stimulating Hormone Receptor (TSHR) on development, behaviour and TH levels in domesticated chickens. *PloS one*, **10**(6), e0129040. https://doi.org/10.1371/journal.pone.0129040 (2015).

58. Fatica, A., & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews. Genetics*, **15**(1), 7–21. https://doi.org/10.1038/nrg3606. (2014).

59. Li, Z., Ren, T., Li, W., & Han, R. Regulatory mechanism and application of lncRNAs in poultry. *In Poultry-An Advanced Learning*. IntechOpen. https://doi.org/10.1002/jcp.22394 (2019).

60. You, Z. *et al*. Integrated analysis of lncRNA and mRNA repertoires in Marek's disease infected spleens identifies genes relevant to resistance. *BMC genomics*, **20**(1), 245. https://doi.org/10.1186/s12864-019-5625-1 (2019).

61. Xie, J. *et al*. Differential expression of heat shock transcription factors and heat shock proteins after acute and chronic heat stress in laying chickens (*Gallus gallus*). *PloS one*, **9**(7), e102204. https://doi.org/10.1371/journal.pone.0102204 (2014).

62. Li, W. *et al*. Analysis of four complete linkage sequence variants within a novel lncRNA located in a growth QTL on chromosome 1 related to growth traits in chickens. *Journal of animal science*, **98**(5), skaa122. https://doi.org/10.1093/jas/skaa122. (2020).

63. Smith, J., Lipkin, E., Soller, M., Fulton, J. E., & Burt, D. W. Mapping QTL Associated with Resistance to Avian Oncogenic Marek's Disease Virus (MDV) Reveals Major Candidate Genes and Variants. *Genes*, **11**(9), 1019. https://doi.org/10.3390/genes11091019 (2020).

64. Akbarian, A. *et al*. Association between heat stress and oxidative stress in poultry; mitochondrial dysfunction and dietary interventions with phytochemicals. *Journal of animal science and biotechnology*, **7**, 37. https://doi.org/10.1186/s40104-016-0097-5 (2016).

65. Powell, F. L. Chapter 13 - Respiration. In: SCANES, C. G. (ed.) Sturkie's Avian Physiology (Sixth Edition). San Diego: Academic Press. Available at:https://rlmc.edu.pk/themes/images/gallery/library/books/Physiology/Avian%20Physiology.pdf. (2015).

66. Varasteh, S., Braber, S., Akbari, P., Garssen, J., & Fink-Gremmels, J. Differences in Susceptibility to Heat Stress along the Chicken Intestine and the Protective Effects of Galacto-Oligosaccharides. *PloS one*, **10**(9), e0138975. https://doi.org/10.1371/journal.pone.0138975 (2015).

67. Zahoor, I., de Koning, D. J., & Hocking, P. M. Transcriptional profile of breast muscle in heat stressed layers is similar to that of broiler chickens at control temperature. *Genetics, selection, evolution: GSE*, **49**(1), 69. https://doi.org/10.1186/s12711-017-0346-x. (2017).

68. Zhang, H. *et al*. 25-hydroxycholecalciferol reverses heat induced alterations in bone quality in finisher broilers associated with effects on intestinal integrity and inflammation. *Journal of animal science and biotechnology*, **12**(1), 104. https://doi.org/10.1186/s40104-021-00627-6. (2021).

69. Quinteiro-Filho, W. M. *et al*. Heat stress decreases expression of the cytokines, avian β-defensins 4 and 6 and Toll-like receptor 2 in broiler chickens infected with Salmonella Enteritidis. *Veterinary Immunology and Immunopathology*, **186**, 19-28. https://doi.org/10.1016/j.vetimm.2017.02.006. (2017).

70. El-Kassas, S. *et al*. Ameliorative effect of dietary supplementation of copper oxide nanoparticles on inflammatory and immune reponses in commercial broiler under normal and heat-stress housing conditions. *Journal of thermal biology*, **78**, 235–246. https://doi.org/10.1016/j.jtherbio.2018.10.009. (2018).

71. He, S.P. *et al*. Impact of heat stress and nutritional interventions on poultry production. *World's Poultry Science Journal*. **74**, 1-18. https://doi.org/10.1017/S0043933918000727. (2018).

72. Li, H., & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **25**(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324. (2009).

73. Broad Institute. Best Practices for Variant Calling with the GATK. Available from: https://www.broadinstitute.org/partnerships/education/broade/best-practices-variant-calling-gatk-1. (2015).

74. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158, https://doi.org/10.1093/bioinformatics/btr330. (2011).

75. Price, A. L. *et al*. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, **38**(8), 904–909. https://doi.org/10.1038/ng1847. (2006).

76. Alexander, D. H., & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC bioinformatics*, **12**, 246. https://doi.org/10.1186/1471-2105-12-246. (2011).

77. Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics (Oxford, England)*, **35**(10), 1786–1788. https://doi.org/10.1093/bioinformatics/bty875. (2019).

78. Weir, B. S., & Cockerham, C. C. Estimating F-Statistics for the analysis of population structure. *Evolution; international journal of organic evolution*, **38**(6), 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x. (1984).

79. Sabeti, P. C. *et al*. Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**(7164), 913–918. https://doi.org/10.1038/nature06250. (2007).

80. Maclean, C. A., Chue Hong, N. P., & Prendergast, J. G. Hapbin: An Efficient Program for Performing Haplotype-Based Scans for Positive Selection in Large Genomic Datasets. *Molecular biology and evolution*, **32**(11), 3027–3029. https://doi.org/10.1093/molbev/msv172. (2015).

81. Quinlan, A. R., & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, **26**(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033. (2010).

82. Thomas, P. D. *et al*. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein science: a publication of the Protein Society*, **31**(1), 8–22. https://doi.org/10.1002/pro.4218. (2022).

Acknowledgements

This research was funded in part by the Bill & Melinda Gates Foundation (BMGF) and with UK aid from the UK Foreign, Commonwealth and Development Office (Grant Agreement OPP1127286). The study constituted part of MP-R's Ph.D. research that was funded by Indonesia Endowment Fund for Education (LPDP). This study was carried out 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), and the International Livestock Research Institute. The findings and conclusions contained within were those of the authors and do not necessarily reflect positions or policies of the BMGF nor the United Kingdom Government. This research was conducted as part of the Consultative Group on International Agricultural Research (CGIAR) Research Program on Livestock and was supported by contributors to the CGIAR Trust Fund. We would also like to thank Edinburgh Genomics (Edinburgh, UK) for producing the sequence data used in this study. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Data availability statement

The whole genome sequence data used in this study have been deposited in the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena) under study accession number PRJEB39536.

Ethics statement

All animal works were reviewed and approved by the Institutional Animal Care and Use Committee of the International Livestock Research Institute (IREC2017-26) and were handled strictly in compliance with the guidelines of this committee. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

O.H, A.A.G, M.P.R., and J.S. conceived the research project. O.B, T.D., and O.H. led the collection of samples and population metadata. M.P.R., and A.A.G. performed the analyses and led the writing of the manuscript, but all authors contributed critically to the drafts.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available within this paper: Supplementary Tables and Supplementary Figures.