

# SCIENTIFIC REPORTS



OPEN

## Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts

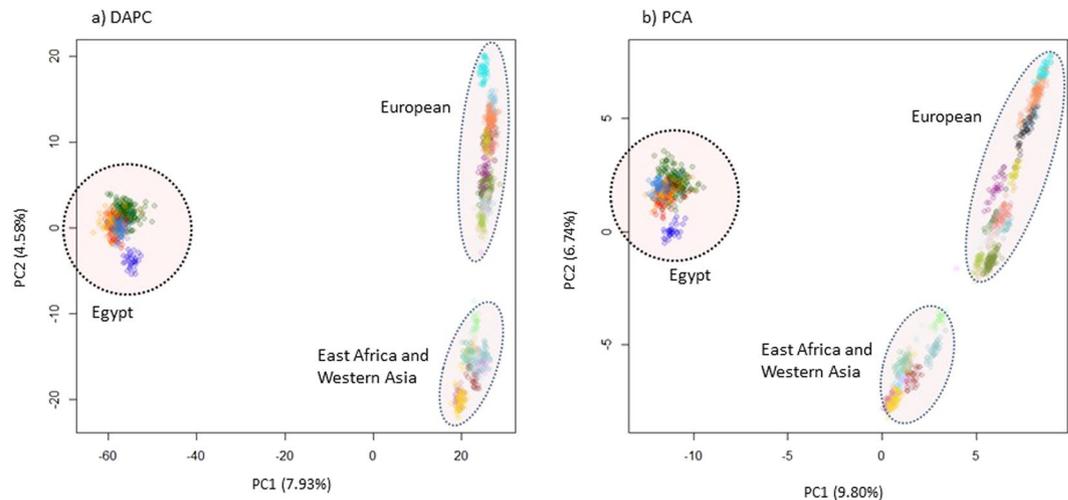
Joram M. Mwacharo<sup>1</sup>, Eui-Soo Kim<sup>2</sup>, Ahmed R. Elbeltagy<sup>3</sup>, Adel M. Aboul-Naga<sup>3</sup>, Barbara A. Rischkowsky<sup>1</sup> & Max F. Rothschild<sup>2</sup>

African indigenous sheep are classified as fat-tail, thin-tail and fat-rump hair sheep. The fat-tail are well adapted to dryland environments, but little is known on their genome profiles. We analyzed patterns of genomic variation by genotyping, with the Ovine SNP50K microarray, 394 individuals from five populations of fat-tail sheep from a desert environment in Egypt. Comparative inferences with other East African and western Asia fat-tail and European sheep, reveal at least two phylogeographically distinct genepools of fat-tail sheep in Africa that differ from the European genepool, suggesting separate evolutionary and breeding history. We identified 24 candidate selection sweep regions, spanning 172 potentially novel and known genes, which are enriched with genes underpinning dryland adaptation physiology. In particular, we found selection sweeps spanning genes and/or pathways associated with metabolism; response to stress, ultraviolet radiation, oxidative stress and DNA damage repair; activation of immune response; regulation of reproduction, organ function and development, body size and morphology, skin and hair pigmentation, and keratinization. Our findings provide insights on the complexity of genome architecture regarding dryland stress adaptation in the fat-tail sheep and showcase the indigenous stocks as appropriate genotypes for adaptation planning to sustain livestock production and human livelihoods, under future climates.

The development of high throughput genome-wide assays and associated computational tools, have made domestic livestock attractive for investigating how an organism's genome is influenced by its production and natural environment. The existence of independent livestock populations/breeds within a species, presents a natural experimental design that can be used to study the genetic mechanisms of adaptive divergence arising from bio-physical and/or anthropological selection. For instance, genome-wide SNP and sequence data has been used to explore genetic mechanisms of adaptation to contrasting environments<sup>1-4</sup> and investigate evidence for genomic selection relating to domestication, breed formation and improvement<sup>5</sup> in livestock species.

Although sheep were domesticated in the Fertile Crescent, Africa is endowed with a diverse repository of the species represented by 179 breeds/populations<sup>6</sup> that have been classified into three broad groups; thin-tail, fat-tail and fat-rump hair sheep. The thin-tails occur in Sudan and in the sub-humid and humid regions of West Africa. The fat-tails occur across the deserts of northern Africa, and in the highlands, semi-arid and arid environments of eastern and southern Africa. The fat-rumps are found exclusively in the semi-arid and arid zones of the Horn of Africa. The thin-tails are the most ancient in the continent and were introduced, via the Isthmus of Suez and/or the southern Sinai Peninsula, whereas the fat-tail's arrived much later, initially via northeastern Africa (Egypt) and later via the Horn of Africa (Ethiopia)<sup>7</sup>. The origin of the fat-rumps remains unknown.

<sup>1</sup>Small Ruminant Genomics Group, International Center for Agricultural Research in the Dry Areas (ICARDA), P. O. Box 5689, Addis Ababa, Ethiopia. <sup>2</sup>Department of Animal Science, Iowa State University, 2255 Kildee Hall, Ames, IA, 50011-3150, USA. <sup>3</sup>Animal Production Research Institute (APRI), Agriculture Research Centre (ARC), Ministry of Agriculture, Nadi Elsaid Street, Dokki, Cairo, Egypt. Correspondence and requests for materials should be addressed to J.M.M. (email: [j.mwacharo@cgiar.org](mailto:j.mwacharo@cgiar.org))



**Figure 1.** Population structure of the study populations revealed by (a) DAPC and (b) PCA analyses.

The genomes of African indigenous sheep have been subjected mainly to natural selection driven by tropical and sub-tropical climates, diseases and parasites. Although their productivity is much lower than that of commercial breeds under intensive production systems, indigenous sheep are often the only option available to millions of resource-poor farmers in agro-pastoral and pastoral production systems, where exotic improved genotypes under-perform under limited (quality and quantity) feed and water resources, and high ecto- and endo-parasite and disease challenges. This is evident in Egypt, a country within the Sahara desert, where sheep of the fat-tail type, are preferred by livestock keepers because of their excellent adaptation to desert-like conditions<sup>8</sup>. This adaptation and the historical significance of Egypt as one of the centers of dispersion of domesticates into Africa, makes Egyptian indigenous sheep of interest in understanding the genetic history of indigenous fat-tail sheep in the continent and the genomic mechanisms underlying their adaptation to dryland environments, which remain poorly investigated. Here, we generated genotype data using the Ovine SNP50K BeadChip from five populations of Egyptian indigenous sheep representative of the fat-tail hair sheep diversity found across the dry belts of Africa, the Middle East and Asia to investigate their genome profiles. Comparative genome analysis with fat-tail sheep from East Africa and western Asia provided insights on the history of the genotype in northeastern and East Africa regions, and the subsequent inclusion of European commercial breeds in selection sweep analysis, allowed us to identify unique genome profiles of fat-tail hair sheep to dryland adaptation.

## Results

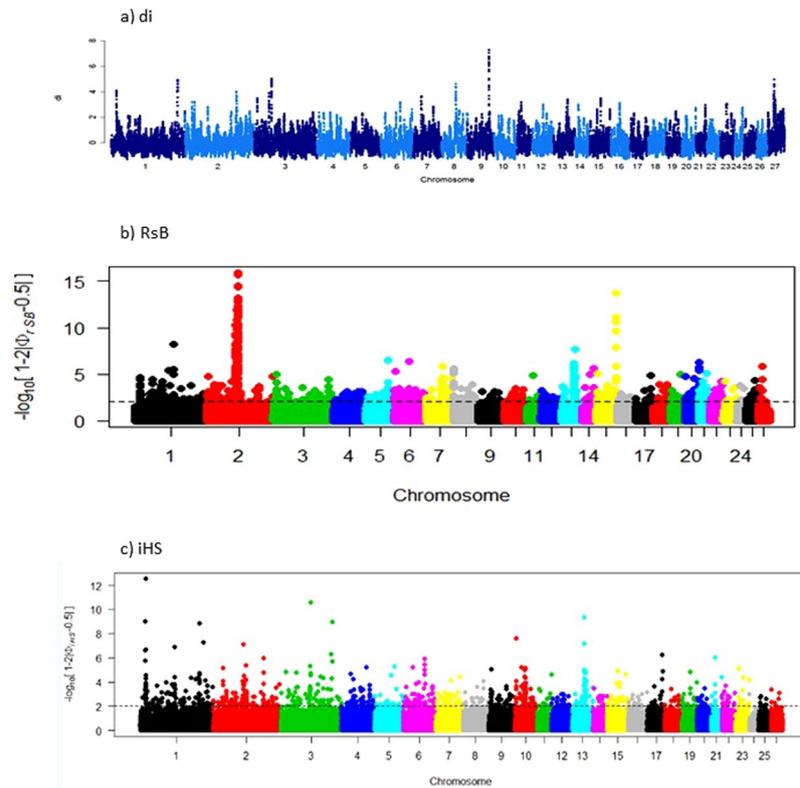
**Population genetic analysis.** Population genetic relationships were assessed with PCA and DAPC (Fig. 1) using 5,140 SNPs that were selected to be unlinked. The first two principal components of the DAPC and PCA accounted for 7.93% and 9.80% (PC1) and 4.58% and 6.74% (PC2), respectively of the total genetic variation. The PC1 separated Egyptian and non-Egyptian populations. The PC2 separated East African and western Asia fat-tail sheep, which seem to cluster together, from the European breeds. The five Egyptian populations clustered very close together, but the non-Egyptian ones disperse along the vertical plane of the two plots. This suggests a lower level of genetic variation between the Egyptian populations but a comparatively higher one between the non-Egyptian populations. Generally, higher PCs (>2) did not result in observable genetic clusters.

**Detection of candidate selection sweep regions.** Based on the results of the DAPC and PCA, SNP genotypes were used to estimate allele frequency differentiation, measured as  $d_i$ , in a pairwise comparison between the Egyptian and non-Egyptian populations. The genome wide distribution of  $d_i$  values for each SNP is shown in Fig. 2a. A total of 109 significant SNPs ( $d_i \geq 4.0$ ) defined seven candidate regions across six chromosomes (Oar1, Oar2, Oar3, Oar8, Oar9, Oar27; Fig. 2a; Supplementary Table S1a). The strongest candidate region was on Oar9 spanning 24 significant SNPs and 15 genes.

The  $R_sB$  analysis revealed 154 significant SNPs ( $p_{RsB} \geq 4.0$ ) that defined 10 candidate selection sweep regions across nine chromosomes (Oar1, Oar2, Oar5, Oar7, Oar8, Oar13, Oar15, Oar20, Oar26; Fig. 2b; Supplementary Table S1b). It identified two candidate regions as the strongest, one on Oar2 and the other on Oar15, spanning 107 and 9 significant SNPs, and, 28 and 0 genes, respectively.

The intra-population  $iHS$  analysis was performed for the five Egyptian populations grouped based on the PCA and DAPC. It identified 47 significant SNPs ( $p_{iHS} \geq 4.0$ ) that defined 14 candidate regions across eight chromosomes (Oar1, Oar2, Oar3, Oar6, Oar10, Oar13, Oar15, Oar17; Fig. 2c; Supplementary Table S1c). Two candidate regions, that were each defined by seven significant SNPs, on Oar1 and Oar13 and spanning 46 and 17 genes, respectively were the strongest.

**Overlap between candidate selection sweep regions.** The three approaches ( $d_i$ ,  $R_sB$ ,  $iHS$ ) used here to detect selection sweeps revealed 31 candidate regions across 15 chromosomes. Selection signatures were identified on Oar1, Oar2, Oar3, Oar13 and Oar15 by more than one approach. The signatures identified by



**Figure 2.** Manhattan plots for results of selection signature analysis undertaken using the (a) *di*, (b) *RsB* and (c) *iHS* approaches for the study populations.

the three approaches on Oar1 had an overlapping segment (*di* = 19,517,811–20,118,195 bp; *RsB* = 19,651,513–19,761,666 bp; *iHS* = 19,409,931–21,607,699 bp) (Supplementary Table S1a, S1b, S1c) spanning two genes (*TOE1*, *TESK2*). The signatures identified by the three approaches on Oar2, and by *di* and *iHS* on Oar3, had no overlaps. Similarly, the selection sweeps that were identified by the three approaches on Oar13 and Oar15 had no overlapping segments. Reducing the significance threshold for *di* to  $\geq 3.0$ , resulted in overlapping segments with *iHS* on Oar13 (*di* = 57,868,286–58,232,687 bp; *iHS* = 57,771,173–58,251,812 bp) that spanned seven genes (*PCK1*, *CTCFL*, *ENSOARG00000017883*, *RAE1*, *SPO11*, *BMP7*, *ZBP1*), and Oar15 (*di* = 43,962,190–44,595,229 bp; *iHS* = 44,112,180–44,332,191 bp) that spanned three genes (*ENSOARG00000015306*, *ENSOARG00000017164*, *ENSOARG00000017173*).

**Gene content and functional annotation of the candidate regions.** From the 31 candidate selection sweep regions, seven (*di* = 2, *RsB* = 3, *iHS* = 2) spanned no genes (Supplementary Table S1a, S1b, S1c) on the OARv4.0 genome assembly. Such regions have also been reported in cattle<sup>5,9,10</sup>. We investigated this further by checking the gene content of the seven regions against the Bovine UMD3.1 and *Capra hircus* V1 (ARS1 (GCF\_001704415.1)) genome assemblies. Interestingly they spanned 83 and 18 genes, respectively on the bovine and caprine genomes, suggesting incomplete annotation of the ovine genome assembly. Genome-wide, we identified 172 genes mapping to 24 (*di* = 5, *RsB* = 7, *iHS* = 12) candidate regions that were defined by 218 significant SNPs across 12 chromosomes (Supplementary Table S1a, S1b, S1c).

We performed functional enrichment for the 172 genes using the Enrichr web tool (See Supplementary Table S2). The rank based and combined score ranking gave similar results and revealed ten GO terms (Table 1) as the most significant ( $P \leq 0.01$ ). The genes were associated with diverse biological functions and some had roles in multiple functions. Relevant to this study was that majority of the functions were associated with adaptation to dryland environment stress (Supplementary Table S3a and S3b). They included response to feed stress; lipid, protein and carbohydrate metabolism; response to heat/temperature stimulus and oxidative stress; protection from ultraviolet radiation; regulation of immune response, DNA damage repair, transcription and translation, protein modification and RNA processing. Other functions included regulation of body size, growth and development; muscle structure, function and adaptation; kidney function and development; and reproduction and nervous system development and function.

Comparisons against the KEGG pathway and WikiPathway databases (Supplementary Table S2), revealed two and five pathways (Table 1), respectively as the most significant ( $P < 0.05$ ). In general, the analysis of GO terms shows an over-representation of GO categories in pathways relating to stress response and which may underlie dryland stress adaptation in the fat-tail sheep (Supplementary Table S4a, S4b, S4c).

Description	GO Term	P-value	Associated genes
<b>(a) Enrichment analysis</b>			
Regulation of stem cell maintenance	GO:2000036	P = 0.005938196	TAL1, BMP7
Granulocyte differentiation	GO:0030851	P = 0.005938196	L3MBTL3, TAL1
Vitamin biosynthetic process	GO:0009110	P = 0.007613258	AKR1A1, MMACHC
Microtubule cytoskeleton organization involved in mitosis	GO:1902850	P = 0.005172798	KIF3B, CHEK2
Spindle assembly involved in mitosis	GO:0090307	P = 0.003174766	KIF3B, CHEK2
Negative regulation of embryonic development	GO:0045992	P = 0.014909626	COL5A2, BMP7
Erythrocyte maturation	GO:0043249	P = 0.003790269	L3MBTL3, TAL1
Benzene containing compound metallic process	GO:0042537	P = 0.013736768	CMPK1, CYP4B1
Hair cell differentiation	GO:0035315	P = 0.005938196	ERCC3, MYO6
Protein-lipid complex assembly	GO:0065005	P = 0.008521348	BIN1, PLAGL2
<b>(b) KEGG Pathway</b>			
Pyrimidine metabolism	hsa00240	P = 0.049048713	POLR2D, CMPK1, DCK
Citrate cycle (TCA Cycle)	hsa00020	P = 0.022794916	IDH3B, PCK1
<b>(c) WikiPathway</b>			
mRNA processing	WP310	P = 0.036816137	ZBP1, ESRP1, ANKAR, GRSF1, KIAA1429, RAE1, SNRPB
TCA Cycle	WP434	P = 0.022794916	PDP1, IDH3B
Oxidation by Cytochrome P450	WP43/ WP1274	P = 0.012581695	CYP27C1, CYP4 × 1, CYP4B1
Splicing factor NOVA regulated synaptic proteins	WP1983	P = 0.042470409	KCNJ6, EPB41L2
miRNA targets in ECM and membrane receptors	WP2911	P = 0.040651645	COL3A1, COL5A2

**Table 1.** Overexpressed GO terms among the 172 candidate genes identified to be under selection in sheep.

## Discussion

Occurring predominantly in the Afro-Asiatic drybelts, the fat-tail sheep account for approximately 25% of the global sheep population. Here, we analysed genotype data generated with the Ovine SNP50K Chip to investigate the genome profiles and the genetic basis of adaptation to dry environments in fat-tail sheep from a desert environment in Egypt. The inclusion of fat-tail sheep from East Africa and western Asia allowed us to gain insights on the history of the fat-tails in Africa. Due to their geographic proximity and ancient history of commercial and religious interactions, we hypothesized that the fat-tail sheep from Egypt and western Asia should show a close genetic relationship. However, the DAPC and PCA showed a genetic divergence between the Egyptian and western Asia and East African populations, respectively and a close genetic relationship between the latter. This confirms the results of a previous analysis with microsatellites in a large sample of African indigenous sheep that showed a clear divergence between Egypt's fat-tail Ossimi and its East and southern African counterparts<sup>11</sup>. These results indicate that Egyptian and East African fat-tail sheep represent different genetic stocks, and suggest one of two possibilities; an independent introduction of at least two gene pools of fat-tail sheep to Africa, or an introduction of one genetic stock that gave rise to two gene pools following reproductive isolation and adaptation to different eco-climates. We favour the first suggestion which is in line with archaeological evidence which supports two separate entry points of fat-tail sheep into Africa, initially via northeast Africa (around 7500 and 7000 years ago) and later (around 5000 years ago) via the Horn of Africa<sup>12</sup>. The close genetic relationship between the East African and western Asia populations is difficult to explain but we suggest that it may be due to their recent common history. On the other hand, long-term reproductive isolation may explain the divergence of the Egyptian from western Asia populations. Intensive anthropological selection and/or genetic drift arising from low effective population sizes in the European breeds may explain their divergence from the African and western Asia populations.

The genetic divergence revealed by DAPC and PCA informed the grouping of populations into Egyptian and non-Egyptian ones for selection sweep analysis. The three approaches (*di*, *iHS*, *Rsb*) detected one overlapping candidate region, and two were detected with *di* and *iHS* after relaxing the stringency of *di*. A modest number of overlapping candidate regions have been reported in several studies<sup>4,13–15</sup> and Bahbahani *et al.*<sup>9</sup> reported none. Although coincident signatures that are detected by multiple approaches may provide strong evidence of selection<sup>16</sup>, their modest occurrence may be due to algorithm differences<sup>17,18</sup>. This may also explain why results from different studies also tend to differ. Therefore, a genomic region that has been identified by only one approach does not exclude the possibility that it could be under selection<sup>19,20</sup>. In total therefore, we detected 24 candidate regions, spanning 172 candidate genes, representing signatures of past and/or on-going selection in Egyptian fat-tail sheep. Unsurprisingly, the regions spanned genes that did not concern production, but rather, adaptation traits. This is because the target populations have mainly been exposed, over a long time period, to complex interacting biophysical stressors (heat, solar radiation, physical exhaustion, resource scarcity, parasites etc.) which impact fitness. This may be the reason why the spanned genes were associated with diverse physiological, molecular and cellular processes and pathways (Supplementary Table S3) emphasizing the importance of multi-functionality for adaptation to dryland environments. The large number of candidate regions and genes detected is also not surprising; similar findings have been reported for livestock species from extreme environments<sup>4,13,15,21</sup>. It reinforces the fact that adaptation is a complex trait that involves many biological processes and quantitative trait loci each having a small and cumulative effect on the overall expression of the phenotype.

Region/Ecology	Type	Country	Breed/Population	Sample size
North Africa (Subtropical dryland)	Indigenous	Egypt	Barki	181
			Saidi	72
			Farafra	62
			Souhagi	49
			AHS	30
		Total	394	
East Africa (Tropical Highland)	Indigenous	Ethiopia	Menz	34
		Kenya	Red Maasai	45
Western Asia (Subtropical)	Indigenous	Cyprus	Cyprus fat tail	30
		Iran	Afshari	37
			Moghani	34
			Qezel	35
		Turkey	Karakas	18
			Norduz	20
			Sakiz	22
		Total	196	
North/Central Europe (Temperate)	Selected	Switzerland	Bundner Oberlander	21
			Swiss White Alpine	21
			Valais Black Nose	21
			Valais Red	21
		Germany	East Friesian Brown	39
			East Friesian White	9
		England	Border Leicester	48
			Dorset Horn	21
			Wiltshire	23
		Ireland	Gallway	49
			Irish Suffolk	55
		Scotland	Scottish Texel	80
		Germany	German Texel	43
		Norway	Old Norwegian Spaelsau	15
Spael coloured	3			
Finland	Finnsheep	96		
		Total	565	

**Table 2.** Sample sizes and information on the goat populations and breeds used in the current study.

Energy and nutrient metabolism is vital for herbivores in food scarce environments. Our candidate regions spanned several genes associated with feeding efficiency and regulation of metabolism and, GO clusters associated with energy metabolic processes (glycolysis/gluconeogenesis, TCA Cycle, insulin signaling pathway, pyruvate metabolism etc.). For instance the gene *PIK3R3*, which has not been reported before, regulates responses to changes in nutritional conditions as well as cellular metabolism and growth<sup>22</sup> and through its association with 5' AMP-activated protein kinase (AMPK), it serves as a metabolic master switch in response to alterations in cellular energy levels<sup>23</sup>. Indeed, sheep under heat stress decrease dry matter intake and rumination time by up to 76%<sup>24</sup> which is related to eating efficiency and metabolic processes<sup>25</sup>. In the drylands most breeds tend to have small body sizes as an adaptation strategy to scarce and poor quality forages and for thermoregulation<sup>25</sup>. This may explain the occurrence, in candidate regions, of genes such as *BMP7*, *MSTN* (*GDF8*) and *STIL* which regulate adult and embryonic size, growth and development<sup>22,26,27</sup>. *BMP7* also plays a crucial role in renal function and development<sup>28,29</sup>. Renal vasodilation, transmembrane transport, water-salt metabolism, bicarbonate absorption, water retention and reabsorption are key functions of the renal cortex and central to desert environment adaptation.

Prolonged exposure to intense solar and ultraviolet radiation, which are key abiotic stressors in arid environments, can result in ophthalmic and skin conditions. Genes controlling pigmentation of coat and skin<sup>30</sup> and eyelids<sup>31</sup> and photoreception and visual protection<sup>32</sup> have been identified and reported to offer protection against solar and UV radiation. None of these reported genes however, occurred in the candidate regions although our study populations had pigmented skins and coats and some such as Barki, Farafra and Souhagi had pigmented eyelids<sup>8</sup>. Instead, we observed *ERCC3* a major nucleotide excision repair (NER) protein. In humans, mutations in *ERCC3* result in skin disorders, such as *xeroderma pigmentosum*, *cockayne syndrome* and *trichothiodystrophy*, which result in sensitivity to UV radiation and oxidative stress<sup>33–35</sup>. Since NER modulates melanocyte stem cell attrition and development of non-pigmented hair<sup>36</sup>, *ERCC3* may be involved in maintaining and regulating the fate and behaviour of melanocyte stem cells and mature melanocytes, and thus the production of melanin which

is responsible for skin, hair and eye colour. Another gene was *TGM3* which is widely expressed in skin cells, specifically keratinocytes and corneocytes. During keratinocyte differentiation, *TGM3* crosslinks structural proteins and lipids in the formation of cornified cell envelope which provides the barrier function of epidermis against harmful environmental stimuli such as UV radiation<sup>37–39</sup>.

Our candidate regions spanned several genes (*PMS1*, *SPO11*, *RAD54L*, *MUTYH*, *CHEK2*, *POLR2D*, *CMPK1*) that maintain cellular functions and DNA repair. For instance, *MUTYH* repairs 8-oxo-G (a mutagenic product of oxidative DNA damage) in the nucleus and mitochondria, via the base excision repair pathway<sup>40,41</sup>. *PMS1* which belongs to the mutL/hexB family of DNA mismatch repair proteins, also forms heterodimers with *MLH1*, a DNA mismatch repair protein<sup>42</sup>. Knockdown of *CMPK1* was observed to delay DNA repair during recovery from UV damage, suggesting it contributes to the efficiency of the DNA repair process<sup>43</sup>. This finding could be associated with the fact that long term exposure to acute and chronic heat stress enhances the production of free radicals which can result in DNA damage and induce oxidative stress leading to mitochondria damage<sup>44–46</sup>, apoptosis and necrosis<sup>47,48</sup>. Therefore the DNA repair system serves to preserve genomic integrity under excessive exposure to UV radiation in dryland environments.

The initiation of stress response involves the activation of the neuroendocrine system to trigger physiological and/or behavioural responses. We found several genes involved in the development of the nervous system and eliciting response to stress, indicating the importance of the neuroendocrine system in activating stress response in the candidate regions. One of the genes, *DMBX1* is involved in brain and sensory organ development<sup>49,50</sup> while *TP53INP1* is a key cell stress response protein with antioxidant function<sup>51,52</sup>. Through its interaction with the ERK1 and p38 mitogen-activated protein kinases, *MKNK1* may play a role in responding to environmental stress and control cytokine production and delayed apoptosis<sup>53,54</sup>. *PLAGL2* is also an oxidative stress responding regulator<sup>55</sup>. The activation of the neuroendocrine system and initiation of stress response results however, in the chronic production of glucocorticoids and catecholamines, which can dysregulate immune functions<sup>56,57</sup>, and corticosteroids which possess potent immunosuppressive properties in lymphocytes<sup>58</sup>. This appears to be counterintuitive. However, we observed several genes such as *ZBP1*, *PRDX1*, *MAST2* and *LURAP* in the candidate regions that enhance immune functions. Indeed some genes such as *MAST2* and *PRDX1* have dual roles. By controlling the activities of TRAF6 and NF- $\kappa$ B; *MAST2* regulates immune response and acts as the first responder to harmful cellular stimuli such as stress, free radicals and UV radiation<sup>59,60</sup>. While *PRDX1* may play an antioxidant protective role in cells, it also contributes to antiviral activity of CD8(+) T-cells<sup>61,62</sup>. The *ZBP1* activates innate immune response by binding foreign DNA, enhances DNA-mediated induction of type I interferons and other genes that activate innate immune responses, as well as, signaling mechanisms underlying DNA-associated antimicrobial immunity and autoimmune disorders<sup>63,64</sup>. *LURAP1* is an activator of the canonical NF- $\kappa$ B pathway and drives the production of proinflammatory cytokines<sup>65</sup>. Taken together, these findings suggest that genes evoking cellular stress and immune responses have been the subject of selection in the course of adaptive evolution to dryland environments.

Thermal stress compromises fertility through a direct effect of hyperthermia on the reproduction axis or through the indirect effect of thermal stress on feed intake to reduce metabolic heat production, leading to changes in energy balance and nutrient availability. Four genes (*CTCFL*, *MAST2*, *TESK2*, *SPO11*) associated with male reproduction physiology were detected in the candidate regions. *CTCFL* is a testis-specific DNA binding protein that forms methylation-sensitive insulators which regulate X-chromosome inactivation, nuclear architecture and transcription<sup>66,67</sup>. *MAST2*<sup>68</sup> and *TESK2*<sup>69</sup> play an important role in spermatid maturation during spermiogenesis and spermatogenesis, respectively. In mouse, knockout of *SPO11* led to meiotic arrest of spermatocytes at zygotene<sup>70,71</sup> resulting in sterility in *SPO11*<sup>-/-</sup> homozygous male mice and atrophied testes. Since thermal stress compromises fertility through a direct effect of hyperthermia on the reproductive axis or indirectly through the effects of thermal stress on feed intake to reduce metabolic heat production resulting in changes in energy balance and nutrient availability, our result suggests that reproductive success is a key determinant of adaptive fitness in the fat-tail sheep.

The observation that some of the genes spanned by the candidate regions are enriched for the GO term “response to hypoxia (GO:0001666; GO:0071456)” and “HIF-1 (hypoxia inducible factor-1) signaling pathway (hsa04066)” are novel findings of our study. The HIF-1 pathway and response to hypoxia plays an important role in cellular response to systemic oxygen levels and has been associated, so far, with adaptation to high altitude environments<sup>1–3</sup>. We suggest that this finding could be related to physical exhaustion arising from long-term trekking of long distances in search of feed and water which results in hypoxia-like conditions and oxygen debt in skeletal muscles. In addition two candidate genes (*MSTN/GDF8*, *BIN1*) occurred in the candidate regions under selection. *MSTN* tightly regulates skeletal muscle homeostasis and promotes the survival of muscle syncytia which make up type I (oxidative/slow) or type II (glycolytic/fast) muscle fibers<sup>72</sup>. Type I fibers are fatigue-resistant and rich in mitochondria and utilize oxidative metabolism to provide a stable and long-lasting supply of ATP<sup>73</sup>. They have been observed in skeletal muscles of endurance athletes and their activation entrains complex pathways that enhance physical<sup>73</sup> and racing performance<sup>74</sup>. Isoforms of *BIN1*, are important in the formation of transverse tubules which play a role in skeletal and cardiac muscle contractility and relation<sup>75,76</sup>. The selection of *MSTN*, *BIN1* and HIF-1 associated genes could have arisen as an adaptive response to endure physical exhaustion arising from long-term long distance walking.

In this study, we generated a catalogue of genetic variants in Egyptian fat-tail sheep. Although the studied populations are only a subset of the fat-tail sheep found in Africa, they illustrate that at least two distinct and phylogeographically structured autosomal gene pools define the genotype in the continent. Genome-wide differentiation and LD based scans of selection sweeps identified several candidate genomic regions under selection that spanned several novel and reported genes with key adaptive physiological functions. These genes especially those associated with pigmentation, muscle function and the HIF pathway would need further investigation and validation using full genome sequences and expression studies in sheep and model species and further test hypotheses arising from our study.

## Material and Methods

**Animals.** The animals used in this study are owned by farmers. Prior to sampling, the objectives of the study were explained to them in local languages so that they could make an informed decision with regard to providing consent to sample their animals. Blood sampling was performed by a licensed veterinarian following the guidelines of the General Organization for Veterinary Service (GOVS), Egypt.

**Sampling, SNP genotyping and data quality control.** Venous blood was collected into EDTA vacutainer tubes from 394 individuals from five fat-tail sheep populations (Barki, Saidi, Farafra, Souhagi, AHS) in Egypt (Table 2). The phenotypic characteristics, and socio-economic and cultural significance of the populations were described by Galal *et al.*<sup>8</sup>. The animals were sampled at random from farmers' flocks where they are managed under transhumant grazing system, and veterinary care and anthropic selection is modest or not practised. The blood samples were transferred onto Whatman FTA™ Classic cards (GE Healthcare UK Ltd) for storage. Genotyping was performed on FTA™ spotted blood at GeneSeek Inc (<http://www.neogen.com/Genomics/>) on the Illumina Ovine SNP50K BeadChip. Similar genotype data from 565 individuals from 15 breeds of European sheep, 196 individuals from seven populations of western Asia fat-tail sheep and 79 individuals from two populations of fat-tail sheep from East Africa (Table 2) were obtained from the Ovine HapMap project (<http://www.sheep-hapmap.org>). The European breeds were used to represent genotypes from a contrasting temperate environment. The East African populations represented fat-tail sheep from a different geographic region within Africa and the western Asia ones represented fat-tails from a region close to the centre of domestication.

PLINK 1.07<sup>77</sup> was used to perform data quality control and processing. A SNP was excluded from the analysis if it failed the following quality criteria; a SNP call rate greater than 95%, SNP genotyping success rate of greater than 90%, a per-SNP minor allele frequency of less than 0.03 and the presence of mapped autosomal loci. Overall, 1,234 individuals genotyped for at least 95% of the SNPs and 51,407 SNPs that passed quality thresholds were available for analysis.

**Population analyses.** We performed the principal component analysis (PCA) using the *adegenet* 1.4–2 package<sup>78</sup> for R<sup>79</sup> to investigate genetic relationships between individuals and populations. To minimize possible confounding effects of linkage disequilibrium on the underlying pattern of population genetic relationship and structure, one in every ten SNPs was sampled and used in PCA. This dataset was also used to perform the discriminant analysis of principal components (DAPC) with *adegenet* 1.4–2 package in R, to assess replication of the population clusters and relationships between individuals and populations. The PCA aims to reduce the total variance in the dataset and identifies the principal components (PCs) that represent population structures based on genetic correlations between individuals. The DAPC reveals genetic differences between groups, as best as possible, while minimizing the variation within the groups.

**Signatures of selection based on genetic differentiation.** We computed locus-specific divergence in allele frequencies between groups of populations revealed by DAPC and PCA, using the *di* statistic<sup>80</sup>. The *di* is a function of unbiased estimates of pairwise values of  $F_{ST}$  calculated for each SNP between one or a group of population(s) against the other(s). For each of the 51,407 SNPs that passed quality control, the expected value and standard deviation of  $F_{ST}$  were calculated. The  $F_{ST}$  values were then averaged across SNPs contained in non-overlapping sliding windows of 20 SNPs. The regions under selection were defined if at least two non-overlapping windows passed the distribution threshold, taking the windows with the highest (top 1%) average values of *di* as the candidate regions<sup>80</sup>.

**Signatures of selection based on linkage disequilibrium (LD).** LD based selection sweeps were identified using the *iHS*<sup>17</sup> and *RsB*<sup>18</sup> approaches. These approaches are based on the extended haplotype homozygosity (EHH) estimates of LD. The *iHS* detects regions with high levels of unexpected EHH within populations relative to neutral expectations and *RsB* detects the same but between populations. The analyses were performed using the *rehh* package<sup>81</sup> in R. SNPs under selection were identified by transforming *iHS* and *RsB* values into  $\text{piHS}$  ( $\text{piHS} = -\log_{10}[1 - 2(\Phi_{(iHS)} - 0.5)]$ ) and  $\text{pRsB}$  ( $\text{pRsB} = -\log_{10}[1 - 2(\Phi_{(RsB)} - 0.5)]$ ), where  $(\Phi_{(x)})$  represents the Gaussian cumulative distribution function. Assuming the values of *iHS* and *RsB* are normally distributed under neutrality, the  $\text{piHS}$  and  $\text{pRsB}$  can be interpreted as  $-\log_{10}(P\text{-value})$ .

For *iHS*, the intra-population Integrated Haplotype Score (*iHS score*) was computed for the five Egyptian populations as a group. For each SNP, we calculated the natural logarithm of the ratio between the integrated EHH for the ancient ( $iHH_A$ ) and derived ( $iHH_D$ ) alleles. We inferred the allelic states of each SNP using two approaches; (i) the SNP with the highest frequency was taken to be the ancestral allele; and (ii) the states were assigned at random following 100 permutations to ensure consistency. The *RsB* scores were computed between the Egyptian and non-Egyptian groups of populations. The integrated EHHS (site-specific EHH) score for each SNP and group of populations (*iES*) were calculated, and the *RsB* statistic between the two groups calculated as the natural logarithm of the ratio between  $iES_{\text{Egyptian}}$  and  $iES_{\text{non-Egyptian}}$ .

Haplotypes for *iHS* and *RsB* were constructed by phasing the genotyped SNPs using Beagle<sup>82</sup>. Haplotype frequencies were then calculated using 20 SNP sliding windows that overlapped by five SNPs. The SNPs having  $\text{piHS}$  and  $\text{pRsB}$  ( $\log_{10}(P\text{-value})$ ) values  $\geq 4.0$  ( $P\text{-value} < 0.0001$ ), were considered significant. Candidate regions were defined if the values for at least two adjacent SNPs were significant. In cases where multiple SNPs were significant, a distance of 0.5 Kb up- and down-stream of the extreme significant SNPs was used to define the candidate regions.

**Functional annotation of candidate regions and genes.** The candidate regions were annotated and the associated genes retrieved from the Ensembl Genome Browser 87 using the Sheep Genome Assembly OARv4.0 (GCF\_000298735.2). Functional enrichment was performed with Enrichr<sup>83</sup> against terms from the

Gene Ontology (GO), KEGG ([www.kanehisa.jp/](http://www.kanehisa.jp/)) and WikiPathways ([www.WikiPathways.org/index.php/WikiPathways](http://www.WikiPathways.org/index.php/WikiPathways)) databases which annotates groups of genes with dedicated terms. To rank the GO terms that are relevant to the candidate genes, we assessed the significance of overlap between the input gene list and the gene sets in each library using the z-score test statistic of the deviation from the expected rank by the Fisher exact test (rank based ranking). This test outperforms the Fisher exact test and is comparable to the combined score ranking test<sup>85</sup>. The NCBI Pubmed (<http://www.ncbi.nlm.nih.gov/pubmed/>), GeneCards<sup>®</sup> (<http://www.genecards.org/>) and UniProt (<http://www.uniprot.org/>) databases were also used to determine the gene functions.

**Data availability.** The data generated in this study is available upon request from Eui-Soo Kim (euisoo.kim@recombinetics.com).

## References

1. Qiu, Q. *et al.* The yak genome and adaptation to life at high altitude. *Nature Genet.* **44**, 946–949 (2012).
2. Dong, K. *et al.* Genomic scan reveals loci under altitude adaptation in Tibetan and Dahe pigs. *PLoS One* **9**, e110520 (2014).
3. Gorkhali, N. A. *et al.* Genomic analysis identified a potential novel molecular mechanism for high-altitude adaptation in sheep at the Himalayas. *Sci. Rep.* **6**, 29963 (2016).
4. Yang, J. *et al.* Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.* **33**, 2576–2592 (2016).
5. Xu, L. *et al.* Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol. Biol. Evol.* **32**, 711–725 (2015).
6. DAGRIS. *Domestic Animal Genetic Resources Information System* (ed. Dessie, T., Hanotte, O. & Kemp, S.). International Livestock Research Institute, Addis Ababa, Ethiopia. <http://www.dagris.info/> (2017).
7. Muigai, A. W. T. & Hanotte, O. The origin of African sheep: archaeological and genetic perspectives. *Afr. Archaeol. Rev.* **30**, 39–50 (2013).
8. Galal, S., Rasoul, F.A., Annous, M.R. & Shoaib, I. *Small Ruminant Breeds of Egypt*. In Iñiguez, L., Characterization of Small Ruminant Breeds in West Asia and North Africa Vol 2: North Africa. International Center for Agricultural Research in the Dry Areas (ICARDA), Aleppo, Syria vi+ 196pp (2005).
9. Bahbahani, H. *et al.* Signatures of positive selection in East African shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis. *Sci. Rep.* **5**, 11729 (2015).
10. Iso-Touru, T. *et al.* Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Anim. Genet.* **47**, 647–657 (2016).
11. Muigai, A. W. T. Characterization and Conservation of Indigenous Animal Genetic Resources: Genetic Diversity and Relationships of Fat-tailed and Thin-tailed Sheep of Africa. PhD thesis, Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya (2003).
12. Gifford-Gonzalez, D. & Hanotte, O. Domesticating animals in Africa: implications of genetic and archeological findings. *J. World Prehist.* **24**, 1–23 (2011).
13. Kim, E.-S. *et al.* Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* **116**, 255–264 (2016).
14. Gouveia, J. J. S. *et al.* Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. *Livest. Sci.* **197**, 36–45 (2017).
15. Kim, J. *et al.* The genome landscape of indigenous African cattle. *Genome Biol.* **18**, 34 (2017).
16. Ramey, H. R. *et al.* Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* **14**, 382 (2013).
17. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, 446–458 (2006).
18. Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* **5**, e171 (2007).
19. Hohenlohe, P. A., Phillips, P. C. & Cresko, W. A. Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *Int. J. Plant Sci.* **171**, 1059–1071 (2010).
20. Oleksyk, T. K., Smith, M. W. & O'Brien, S. J. Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. London [Biol]* **365**, 185–205 (2010).
21. Lv, F.-H. *et al.* Adaptations to climate-mediated selective pressures in sheep. *Mol. Biol. Evol.* **31**, 3324–3343 (2014).
22. Engelman, J. A., Luo, J. & Cantley, L. C. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nature Rev. Genet.* **7**, 606–619 (2006).
23. Winder, W. W. & Hardie, D. G. AMP-activated protein kinase, a metabolic master switch: possible roles in Type 2 diabetes. *Am. J. Physiol. Endocrinol. Metab.* **277**, E1–E10 (1999).
24. Monty, D. E., Kelly, L. M. & Rice, W. R. Acclimatization of St. Croix, Karakul and Rambouillet sheep to intense and dry summer heat. *Small Ruminant Res.* **4**, 379–392 (1991).
25. McManus, C. *et al.* Skin and coat traits in sheep in Brazil and their relation with heat tolerance. *Trop. Anim. Health Prod.* **43**, 121–126 (2011).
26. Wu, M. Y. & Hill, C. S. Tgf- $\beta$  superfamily signaling in embryonic development and homeostasis. *Dev. Cell* **16**, 329–343 (2009).
27. Chen, D., Zhao, M. & Mundy, G. R. Bone morphogenetic proteins. *Growth Factors* **22**, 233–41 (2004).
28. Gould, S. E., Day, M., Jones, S. S. & Dorai, H. BMP-7 regulates chemokine, cytokine, and hemodynamic gene expression in proximal tubule cells. *Kidney Int.* **61**, 51–60 (2002).
29. Zeisberg, M. *et al.* Bone morphogenetic protein-7 inhibits progression of chronic renal fibrosis associated with two genetic mouse models. *Am. J. Physiol. Renal Physiol.* **285**, F1060–F1067 (2003).
30. Cieslak, M., Reissmann, M., Hofreiter, M. & Ludwig, A. Colours of domestication. *Biol. Rev.* **86**, 885–899 (2011).
31. Pausch, H. *et al.* Identification of QTL for UV-protective eye area pigmentation in cattle by progeny phenotyping and genome-wide association analysis. *PLoS One* **7**, e36346 (2012).
32. Wu, H. *et al.* Camelid genomes reveal evolution and adaptation to desert environments. *Nature Comm.* **5**, 5188 (2014).
33. Weeda, G. *et al.* A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell* **62**, 777–791 (1990).
34. Weeda, G. *et al.* A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy. *Am. J. Hum. Genet.* **60**, 320–329 (1997).
35. Riou, L. *et al.* The relative expression of mutated XPB genes results in xeroderma pigmentosum/Cockayne's syndrome or trichothiodystrophy cellular phenotypes. *Hum. Mol. Gen.* **8**, 1125–1133 (1999).
36. Yu, M. *et al.* Deficiency in nucleotide excision repair family gene activity, especially ERCC3, is associated with non-pigmented hair fiber growth. *PLoS One* **7**, e34185 (2012).
37. Kalinin, A. E., Kajava, A. V. & Steinert, P. M. Epithelial barrier function: assembly and structural features of the cornified cell envelope. *Bioessays* **24**, 789–800 (2002).

38. Eckert, R. L., Sturniolo, M. T., Broome, A. M., Ruse, M. & Rorke, E. A. Transglutaminase function in epidermis. *J. Invest. Dermatol.* **124**, 481–492 (2005).
39. Hitomi, K. Transglutaminases in skin epidermis. *Eur. J. Dermatol.* **15**, 313–319 (2005).
40. Pilati, C. *et al.* Mutational signature analysis identifies MUTYH deficiency in colorectal cancers and adrenocortical carcinomas. *J. Pathol.* **242**, 10–15 (2017).
41. Poulsen, M. L. M. & Bisgaard, M. L. MUTYH Associated Polyposis (MAP). *Curr. Genomics* **9**, 420–435 (2008).
42. Smith, C. E. *et al.* Dominant mutations in *S. cerevisiae* PMS1 identify the mlh-Pms1 endonuclease active site and an exonuclease 1-independent mismatch repair pathway. *PLoS Genet.* **9**, e1003869 (2013).
43. Tsao, N., Lee, M. H., Zhang, W., Cheng, Y. C. & Chang, Z. F. The contribution of CMP kinase to the efficiency of DNA repair. *Cell Cycle* **14**, 354–363 (2015).
44. England, K., O'Driscoll, C. & Cotter, T. G. Carbonylation of glycolytic proteins is a key response to drug-induced oxidative stress and apoptosis. *Cell Death Differ.* **11**, 252–260 (2004).
45. Lewandowska, A., Gierszewska, M., Marszalek, J. & Liberek, K. Hsp78 chaperone functions in restoration of mitochondrial network following heat stress. *Bioenergetics* **1763**, 141–151 (2006).
46. Song, X. L., Qian, L. J. & Li, F. Z. Injury of heat-stress to rat cardiomyocytes. *Chin. J. Appl. Physiol.* **16**, 227–230 (2000).
47. Rhoads, R. P., Baumgard, L. H., Suagee, J. K. & Sanders, S. R. Nutritional interventions to alleviate the negative consequences of heat stress. *Adv. Nutr.* **4**, 267–276 (2013).
48. Du, J., Di, H. S., Guo, L., Li, Z. H. & Wang, G. L. Hyperthermia causes bovine mammary epithelial cell death by a mitochondrial-induced pathway. *J. Thermal Biol.* **33**, 37–47 (2008).
49. Kawahara, A., Chien, C. B. & Dawid, I. B. The homeobox gene *mbx* is involved in eye and tectum development. *Dev. Biol.* **248**, 107–117 (2002).
50. Miyamoto, T. *et al.* *Mbx*, a novel mouse homeobox gene. *Dev. Genes Evol.* **212**, 104–106 (2002).
51. Shahbazi, J., Lock, R. & Liu, T. Tumor Protein 53-Induced Nuclear Protein 1 Enhances p53 Function and Represses Tumorigenesis. *Front. Genet.* **4**, 80 (2013).
52. Sándor, N. *et al.* TP53inp1 gene is implicated in early radiation response in human fibroblast cells. *Int. J. Mol. Sci.* **16**, 25450–25465 (2015).
53. Waskiewicz, A. J. *et al.* Phosphorylation of the cap-binding protein eukaryotic translation initiation factor 4E by protein kinase Mnk1 *in vivo*. *Mol. Cell. Biol.* **19**, 1871–1880 (1999).
54. Fortin, C. F., Mayer, T. Z., Cloutier, A. & McDonald, P. P. Translational control of human neutrophil responses by MNK1. *J. Leukoc. Biol.* **94**, 693–703 (2013).
55. Guo, Y., Yang, M. C., Weissler, J. C. & Yang, Y. S. PLAGL2 translocation and SP-C promoter activity – a cellular response of lung cells to hypoxia. *Biochem. Biophys. Res. Commun.* **360**, 659–665 (2007).
56. Padgett, D. A. & Glaser, R. How stress influences the immune response. *Trends Immunol.* **24**, 444–448 (2003).
57. Salak-Johnson, J. L. & McGlone, J. J. Making sense of apparently conflicting data: stress and immunity in swine and cattle. *J. Anim. Sci.* **85**, E81–E88 (2007).
58. Cupps, T. R. & Fauci, A. S. Corticosteroid-mediated immunoregulation in man. *Immunol. Rev.* **65**, 133–155 (1982).
59. Chandel, N. S., Trzyna, W. C., McClintock, D. S. & Schumacker, P. T. Role of oxidants in NF-kappa B activation and TNF-alpha gene transcription induced by hypoxia and endotoxin. *J. Immunol.* **165**, 1013–21 (2000).
60. Fitzgerald, D. C. *et al.* Tumour necrosis factor-alpha (TNF-alpha) increases nuclear factor kappaB (NFkappaB) activity in and interleukin-8 (IL-8) release from bovine mammary epithelial cells. *Vet. Immunol. Immunopathol.* **116**, 59–68 (2007).
61. Sun, C. C. *et al.* Peroxiredoxin 1 (Prx1) is a dual-function enzyme by possessing Cys-independent catalase-like activity. *Biochem. J.* **474**, 1373–1394 (2017).
62. He, T. *et al.* Peroxiredoxin 1 knockdown potentiates  $\beta$ -lapachone cytotoxicity through modulation of reactive oxygen species and mitogen-activated protein kinase signals. *Carcinogenesis* **34**, 760–769 (2013).
63. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).
64. Rathinam, V. A. & Fitzgerald, K. A. Innate immune sensing of DNA viruses. *Virology* **411**, 153–162 (2011).
65. Jing, Z. *et al.* Chromosome 1 open reading frame 190 promotes activation of NF-kB canonical pathway and resistance of dendritic cells to tumor-associated inhibition *in vitro*. *J. Immunol.* **185**, 6719–6727 (2010).
66. Loukinov, D. I. *et al.* BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Nat. Acad. Sci. USA* **99**, 6806–6811 (2002).
67. Ohlsson, R., Renkawitz, R. & Lobanenko, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–527 (2001).
68. Huang, N. *et al.* A screen for genomic disorders of infertility identifies MAST2 duplications associated with non-obstructive azoospermia in humans. *Biol. Reprod.* **93**: 61, 1–10 (2015).
69. Rosok, O., Pedetour, F., Ree, A. H. & Aasheim, H. C. Identification and characterisation of TESK2, a novel member of the LIMK/ TESK family of protein kinases, predominantly expressed in testis. *Genomics* **61**, 44–54 (1999).
70. Baudat, F., Manova, K., Yuen, J. P., Jasin, M. & Keeney, S. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. *Mol. Cell* **6**, 989–998 (2000).
71. Romanienko, P. J. & Camerini-Otero, R. D. The mouse Spo11 gene is required for meiotic chromosome synapsis. *Mol. Cell. Proteomics* **6**, 975–987 (2000).
72. Bradley, L., Yaworsky, P. J. & Walsh, F. S. Myostatin as a therapeutic target for musculoskeletal disease. *Cell. Mol. Life Sci.* **65**, 2119–2124 (2008).
73. Wang, Y.-X. *et al.* Regulation of muscle fiber type and running endurance by PPAR $\delta$ . *PLoS Biol.* **2**, e294 (2004).
74. Mosher, D. S. *et al.* A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* **3**, e79 (2007).
75. Tjondrokoesoemo, A. *et al.* Disrupted membrane structure and intracellular Ca<sup>2+</sup> signaling in adult skeletal muscle with acute knockdown of Bin1. *PLoS One* **6**, e25740 (2011).
76. Zhou, K. & Hong, T. Cardiac BIN1 (cBIN1) is a regulator of cardiac contractile function and an emerging biomarker of heart muscle health. *Sci. China Life Sci.* **60**, 257–263 (2017).
77. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
78. Jombart, T. & Ahmed, I. *adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
79. R-Development Core Team. R: A Language and Environment for Statistical Computing Version 3.0.1. Vienna, Austria. R Foundation for Statistical Computing (2013).
80. Akey, J. M. *et al.* Tracking footprints of artificial selection in the dog genome. *Proc. Nat. Acad. Sci. USA* **107**, 1160–1165 (2010).
81. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
82. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
83. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

## Acknowledgements

We express our gratitude's to all livestock farmers in Egypt whose animals were sampled and the APRI personnel for their assistance in sampling. The International Sheep Genomics Consortium (<http://www.sheephapmap.org>) is acknowledged. This project was supported by funding from the College of Agriculture and Life Sciences of Iowa State University, the State of Iowa, Hatch funds, the Ensminger program and the ARC Egypt-ICARDA Collaborative Research Program. ICARDA also thanks the donors supporting the CGIAR Research Program (CRP) on Livestock (formerly Livestock and Fish CRP). This study forms part of our ongoing efforts to understand the adaptation of local indigenous livestock to improve their productivity.

## Author Contributions

J.M.M., A.M.A., B.A.R., M.F.R. designed the study. A.R.E. collected the samples. J.M.M., E.-S.K., A.R.E. analysed the data. J.M.M. wrote the manuscript. All the authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-17775-3>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as 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 images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017