Combining landscape genomics and ecological modelling to investigate local adaptation of indigenous Ugandan cattle to East Coast fever

Elia Vajana[1,#a*], Mario Barbato[1], Licia Colli[1], Marco Milanesi[1,2,3], Estelle Rochat[4], Enrico Fabrizi[5], Christopher Mukasa[6], Marcello Del Corvo[1], Charles Masembe[7], Vincent Muwanika[8], Fredrick Kabi[9], Tad Stewart Sonstegard[10], Heather Jay Huson[11], Riccardo Negrini[1,12], Stéphane Joost[4¶], Paolo Ajmone-Marsan[1,13¶], on behalf of The NextGen Consortium[^]

[1]Dipartimento di Scienze Animali, della Nutrizione e degli Alimenti, e Centro di Ricerca sulla Biodiversità e sul DNA Antico (BioDNA), Università Cattolica del S. Cuore, Piacenza, Italy

[2]Department of Support, Production and Animal Health, School of Veterinary Medicine, São Paulo State University, Araçatuba, Brazil

[3]International Atomic Energy Agency (IAEA), Colaborating Centre on Animal Genomics and Bioinformatics, Araçatuba, Brazil

[4]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[5]Dipartimento di Scienze Economiche e Sociali, Università Cattolica del Sacro Cuore, Piacenza, Italy

[6]National Animal Genetic Resource Centre and Data Bank, Entebbe, Uganda

[7]Department of Zoology, Entomology and Fisheries, Makerere University, Kampala, Uganda

[8]Department of Environmental Management, Makerere University, Kampala, Uganda

[9]National Livestock Resources Research Institute (NaLIRRI), National Agricultural Research Organisation, Tororo, Uganda

[10]Recombinetics, Inc., St. Paul, Minnesota, United States of America

[11]Department of Animal Science, Cornell University, Ithaca, New York, United States of America

[12]Associazione Italiana Allevatori, Roma

[13]PRONUTRIGEN-Centro di Ricerca Nutrigenomica e Proteomica, Università Cattolica del S. Cuore, Piacenza, Italy

[#a]Current Address: Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[*] Corresponding author

E-mail: elia.vajana@epfl.ch (EV)

[¶]These authors contributed equally to this work.

61 Short title: Local adaptation to East Coast fever

62

Abstract

East Coast fever (ECF) is a fatal sickness affecting cattle populations of eastern, central, and southern Africa. The disease is transmitted by the tick *Rhipicephalus appendiculatus*, and caused by the protozoan *Theileria parva parva*, which invades host lymphocytes and promotes their clonal expansion. Importantly, indigenous cattle show tolerance to infection in ECF-endemically stable areas. Here, the putative genetic bases underlying ECF-tolerance were investigated using molecular data and epidemiological information from 823 indigenous cattle from Uganda. Vector distribution and host infection risk were estimated over the study area and subsequently tested as triggers of local adaptation by means of landscape genomics analysis. We identified 41 and seven candidate adaptive loci for tick resistance and infection tolerance, respectively. Among the genes associated with the candidate adaptive loci are *PRKG1* and *SLA*2. *PRKG1* was already described as associated with tick resistance in indigenous South African cattle, due to its role into inflammatory response. *SLA*2 is part of the regulatory pathways involved into lymphocytes' proliferation. Additionally, local ancestry analysis suggested the zebuine origin of the genomic region candidate for tick resistance.

Author summary

The tick-borne parasite *Theileria parva parva* infects cattle populations of eastern, central and southern Africa, by causing a highly fatal pathology called "East Coast fever". The disease is especially severe for the exotic breeds imported to Africa, as well as outside the endemic areas of East Africa. In these regions, indigenous cattle populations can survive to infection, and this

tolerance might result from unique adaptations evolved to fight the disease. We investigated this hypothesis by using a method named "landscape genomics", with which we compared the genetic characteristics of indigenous Ugandan cattle coming from areas at different infection risk, and located genomic sites potentially attributable to tolerance. In particular, the method pinpointed two genes, one (*PRKG1*) involved into inflammatory response and potentially affecting East Coast fever vector attachment, the other (*SLA2*) involved into lymphocytes proliferation, a process activated by *T. parva parva* infection. Our findings can orientate future research on the genetic basis of East Coast fever-tolerance, and derive from a general method that can be applied to investigate adaptation in analogous host-vector-parasite systems. Characterization of the genetic factors underlying East Coast-fever-tolerance represents an essential step towards enhancing sustainability and productivity of local agroecosystems.

Introduction

East Coast fever (ECF) is an endemic vector-borne disease affecting cattle populations of eastern and central Africa. ECF etiological agent is the protozoan emo-parasite protozoan *Theileria parva* Theiler, 1904, vectored by the hard-bodied tick vector *Rhipicephalus appendiculatus* Neumann, 1901. The disease is reported to cause morbidity in indigenous populations and high mortality rates among exotic breeds and crossbreds, thus undermining the livestock sector development in the affected countries [1–3].

Cape buffalo (*Syncerus caffer* Sparrman, 1779) is *T. parva* native host, being its wild and asymptomatic reservoir [4]. A primordial contact between buffalo-derived *T. parva* and domestic

bovines is likely to have occurred ~4.500 years before present (YBP) [5]. However, it is hard to define if the host jump affected taurine- or indicine-like cattle first, since no consensus can easily be reached to define who among *Bos taurus* and *B. indicus* migrated into ECF endemic regions first [6–8]. Indeed, African taurine cattle might have reached eastern Africa sometime between ~8,000 and 1,500 YBP [7,8], and the most ancient zebuine colonization wave is estimated to have occurred between ~4,000-2,000 YBP from the Asian continent, as suggested by the first certain archaeological record dated 1,750 YBP [6]. Once *T. parva* spread to domestic populations, coevolution between the parasite and the new hosts likely led to the divergence between buffalo- (*T. parva lawracei*) and cattle-specific (*T. parva parva*) parasite strains [9,10], and to the appearance of infection-tolerant indigenous herds [11,12].

Most likely, ECF-tolerance appeared (and is only observable) in areas where environmental conditions guaranteed a constant coexistence between the vector, the parasite and the domestic host. Such a particular situation, together with the evolution of some sort of “innate resistance” [13], plausibly prompted the establishment of an epidemiological state referred to as endemic stability, a condition where hosts become parasite reservoirs with negligible clinical symptoms [14]. However, no clear indication for a genetic control has been provided for ECF-tolerance so far [12], despite host genetic factors were identified for tolerance to tropical theileriosis [15] (a disease caused by the closely related *T. annulata*), and tick resistance [16].

Here, we propose an integrated approach based on ecological modelling and landscape genomics to explore the putative adaptive component sustaining ECF-endemic stability. Furthermore, given the existence of host populations showing differential susceptibility to ECF

124 [13,14,17], we will refer to ECF-tolerance as a potential case of local adaptation [18]. Since
125 endemically stable areas are currently inhabited by the indigenous zebu and the zebu *x* African *B.*
126 *taurus* crosses sanga and zenga [8,19], two basic hypotheses can be associated to the origin of
127 local adaptation to ECF: (i) at first adaptation appeared in local African *B. taurus* populations and
128 was then introgressed into zebu and derived sanga and zenga crossbreds; alternatively, (ii) it
129 appeared in *B. indicus*, and then either evolved independently in zebuine populations of eastern
130 Africa, or was imported from the Indian continent, where similar selective pressures are recorded
131 [20,21].

132 Specific regions in the South-West [14] and in the East [17] of current Uganda are reported
133 to be ECF-endemically stable, thus making this country a candidate for investigating local
134 adaptation to ECF. Moreover, indigenous Ugandan cattle populations are proven to be connected
135 by high rates of gene flow [22], and a strong spatially varying selection is expected on their
136 genomes because of regional climatic differences shaping ECF epidemiology over the country
137 [11]. These requirements are all likely to have promoted local adaptation to the disease [23], even
138 over short time scales (i.e. from thousands of years to few decades), as observed for other plant
139 and animal species [24–26].

140 To test our approach, we exploited genomic data from indigenous Ugandan cattle and
141 spatial information on parasite and vector occurrence. First, we modelled ECF-vector potential
142 distribution and infection risk in cattle to define the spatially varying selective pressure over the
143 host genomes. Then, we searched for Single Nucleotide Polymorphisms (SNPs) potentially
144 involved into local adaptation to ECF through genotype-environment association (GEA) analysis,

and we annotated candidate genes. Finally, we studied the ancestral origin of the identified genomic regions by means of local ancestry analysis to shed light on the possible evolutionary origins of local adaptation to ECF.

Materials and Methods

Ecological modelling

R. appendiculatus occurrence probability (Ψ_R) and *T. parva parva* infection risk in cattle (γ) were modelled and used as environmental predictors into landscape genomics models. Geographical variability in both Ψ_R and γ was assumed to describe the spatially heterogeneous selective pressure on cattle genomes. Further, *S. caffer* occurrence probability (Ψ_S) was estimated and used in combination with Ψ_R to model γ, as the geographical proximity between Cape buffaloes and cattle herds constitutes a factor for explaining ECF incidence. The following three sections will describe data and methods used to estimate Ψ_R, Ψ_S, and γ.

Raster data. Bioclimatic variables (BIO) referring to the time span between 1960 and 1990 were collected from the WorldClim database (v.1.4. release3) [27] at a spatial resolution of 30 arc-seconds and in the un-projected latitude/longitude coordinate reference system (WGS84 datum). Altitude information was collected from the SRTM 90m Digital Elevation Database (v.4.1) [28], which provides tiles covering Earth's land surface in the WGS84 datum, at 90 m resolution at the equator. Altitude was used to compute terrain slope through the function `terrain` implemented in the R package `raster` [29]. The ten-year (2001-2010) averaged Normalized Difference

Vegetation Index (NDVI) was derived for 72 ten-day annual periods from the "eMODIS products" (S1 Text) [30], in the WGS84 datum, and at a resolution of 250 m at the equator. A raster file describing cattle density (number of animals/km^2) was acquired from the Livestock Geo-Wiki database [31], in the WGS84 datum, at a resolution of 1 km^2 at the equator. A raster file describing each pixel distance from the nearest water source was obtained with the function `distance` within the R package `raster`. The "Land and Water Area" dataset from the Gridded Population of the World collection (GPV v.4) [32] was used to define water bodies in Uganda at a resolution of 30 arc-seconds with WGS84 datum.

All raster files were transposed into Africa Albers Equal Area Conic projection to guarantee a constant pixel size and meet the main assumption of the statistical technique used to model Ψ_R and Ψ_S, i.e. that each pixel presents the same probability to be randomly sampled in order to detect the species occurrence [33]. Raster files were standardised to the same resolution (~0.85 km^2), origin, and extent. To avoid the inclusion of potentially misleading background locations while characterizing the occurrence probability of terrestrial species, inland water surfaces were masked prior to Ψ_R and Ψ_S estimation [34]. Quantum GIS (v.2.16.2) [35] and the R package `raster` were used for raster files manipulation.

Species distribution models. The R package `Maxlike` [36] was used to model Ψ_R and Ψ_S over Uganda. `Maxlike` is able to estimate species occurrence probability (Ψ) from presence-only data, by maximizing the likelihood of occurrences under the logit-linear model [33]:

$$ln\left(\frac{\psi_x}{1-\psi_x}\right) = \beta_0 + \boldsymbol{\beta} z(x) \quad (1)$$

where Ψ_x denotes the species occurrence probability in the *x* pixel of the landscape, β_0 the model intercept (i.e. the expected prevalence across the study area), $\boldsymbol{\beta}$ the vector of slope parameters, and *z(x)* the vector of environmental variables for *x*. Occurrence probability in *x* is derived from the inverse logit:

$$\psi_x = \frac{e^{\beta_0+\boldsymbol{\beta}z(x)}}{1+e^{\beta_0+\boldsymbol{\beta}z(x)}} \quad (2)$$

Fifty-one and 61 spatial records of *R. appendiculatus* and *S. caffer* (Figs 1A and 1B) were obtained from a tick occurrence database previously collected [37], and the Global Biodiversity Information Facility [38], respectively.

The most relevant environmental variables affecting tick and Cape buffalo distributions were identified from the literature. Specifically, the BIO variables representing temperature/precipitation interaction in the most extreme periods of the year were used to model *R. appendiculatus* occurrence (Table 1 and S1 Fig.) [39,40], while altitude, terrain slope, NDVI, distance to water sources (Wd), and annual precipitation (BIO_{12}) were used to model the Cape buffalo distribution [41–43]. A `Maxlike` regression analysis was applied to individuate the NDVI values best predicting the available *S. caffer* occurrences, and the period April 6-15 was retained for subsequent analyses (S2 Fig.). No variable depicting the top-down regulatory effect of predators on buffalo populations was considered, as bottom-up ecological mechanisms (like quantity and quality of food resources) are argued to play the main role in determining large herbivores distribution [44].

204 **Table 1. Predictors for *R. appendiculatus* distribution model.**

Bioclim variable	Definition
BIO_8	Mean temperature[a] of the wettest three months (quarter) of the year
BIO_9	Mean temperature of the driest quarter
BIO_{10}	Mean temperature of the warmest quarter
BIO_{11}	Mean temperature of the coldest quarter
BIO_{16}	Precipitation[b] of the wettest quarter
BIO_{17}	Precipitation of the driest quarter
BIO_{18}	Precipitation of the warmest quarter
BIO_{19}	Precipitation of the coldest quarter

205 [a]Temperature was transformed from dC° to C° prior analyses. [b]Precipitation is expressed in
206 millimetres.

207

208 Collinearity was checked prior to analyses by computing pairwise absolute correlations (|r|)
209 between variables, which were considered collinear when |r| exceeded the suggested threshold of
210 0.7 [45]. High collinearity was found among BIO variables, which were then subjected to
211 principal components analysis (PCA) to obtain orthogonal predictors for Ψ_R.

212 Obtained components were tested into univariate and multivariate *R. appendiculatus*
213 distribution models. Particularly, components explaining up to 95% of the original variance [46]
214 were individuated and tested with different combinations into multivariate models, leading to a
215 total of twelve candidate *R. appendiculatus* distribution models. Conversely, all the combinations
216 of environmental variables were tested into univariate up to penta-variate Cape buffalo
217 distribution models, resulting in a total of 31 candidate models for predicting *S. caffer* potential
218 distribution.

219 In both cases, Bayesian Information Criterion (BIC) was used to select the best models

[47]. Bring's standardization [48,49] was applied on predictors before parameters' estimate, and the delta method was implemented to compute the 95% confidence intervals around the fitted Ψ_{Rx} and Ψ_{Sx}.

Infection risk model. In the context of the European Project NextGen (http://nextgen.epfl.ch), 587 blood samples from Ugandan indigenous cattle were tested for the presence/absence of *T. parva parva* p104 antigen DNA sequence [11]. Samples were collected and georeferenced in correspondence of 203 farms distributed over a grid of 51 cells covering the whole Uganda, with an average of 12 (±4 s.d.) animals/cell, and three (±1 s.d.) animals/farm (Fig 1C).

ECF epidemiology is complex and determined by both biotic and abiotic factors [2]. Particularly, *R. appendiculatus* occurrence (Ψ_R) [1,50–52], cattle density (Cd) [3,53], potential proximity with *S. caffer* (Ψ_S) and the maximal temperature in the warmest month of the year (BIO_5) were considered to predict γ. BIO_5 was used to account for the possible limiting effect of high temperatures on the parasite development into the tick [54]. Predictors' values were obtained at the geographical position of each animal (i.e. locations of the farms), checked for the presence of collinearity (as done for the species distribution models) and outliers (S3 Fig.), and subsequently standardized following Bring's procedure prior to parameters' estimation.

Infection risk for any *i-th* animal was modelled using a binary mixed-effects logistic regression, where Ψ_R, BIO_5, Cd, and Ψ_R were specified as fixed effects, and random intercepts were estimated for each farm to account for the possible influence of local environmental conditions and management practices (e.g. differential use of acaricides), as well as unmeasured biological features (e.g. breed- or individual-specific response to tick burden) [13]. Since

241 geographical position of samples was recorded at the farm-level, all the animals coming from a
242 given farm were characterized by equal environmental values. Thus, the model can be written as:

243 $$ln\left(\frac{\gamma_{ij}}{1-\gamma_{ij}}\right) = \left(\beta_0 + b_{0j}\right) + \boldsymbol{\beta} z(j) \quad (3)$$

244 $$b_{0j} \sim \mathrm{N}\left(0, \sigma^2_{b_0}\right) \quad (4)$$

245 where γ_{ij} represents *T. parva parva* infection risk for the *i-th* animal in the *j-th* farm, β_0 is the
246 population intercept [55], β_0+b_{0j} is the *j-th* farm random intercept, $\boldsymbol{\beta}$ the vector of slope
247 parameters, and *z(j)* the vector containing the predictors' values as derived from the pixel where
248 the *j-th* farm is located, equal for all the animals in *j*. In this way, animals in *j* are expected with
249 the same predicted γ, so that infection risk in the *j-th* farm can be calculated using the population
250 model from the previous equation:

251 $$\gamma_j = \frac{e^{\beta_0 + \boldsymbol{\beta} z(j)}}{1 + e^{\beta_0 + \boldsymbol{\beta} z(j)}} \quad (5)$$

252 Estimates of the parameters were obtained through the Maximum Likelihood criterion
253 using the `glmer` function included in the R package `lme4` [56].

254 **Landscape genomics**

255 **Molecular datasets.** The NextGen project genotyped 813 georeferenced indigenous cattle from
256 Uganda using the medium-density BovineSNP50 BeadChip (54,596 SNPs; Illumina Inc., San
257 Diego, CA, USA). Landscape genomics analyses were carried out on this set of animals, which
258 will be referred to as the "landscape genomics dataset" (LGD). Samples were collected according
259 to the spatial scheme represented in Fig 1C, and encompassed 503 of the individuals tested for *T.*

parva parva infection. Quality control (QC) procedures were carried out with the software PLINK v.1.7 [57]. LGD was limited to autosomal chromosomes and pruned for minor allele frequency (MAF) <0.01, genotype call rates <0.95, and individual call rate <0.9. Pairwise genome-wide identity-by-descent (IBD) values were estimated, and one individual per pair showing IBD>0.5 was excluded from analyses to reduce the risk of spurious associations due to unreported kinship [58]. To avoid excluding too many individuals from nearby areas, spatial positions of the highlighted pairs were considered prior to removal.

Population genetic structure of Ugandan cattle was studied on the landscape genomics dataset merged with molecular data from other European taurine, African taurine, zebuine and sanga populations retrieved from various sources and for different geographical areas worldwide (S1 Table). This extended dataset will be referred to as the "population structure dataset" (PSD). PLINK was used to prune PSD for linkage disequilibrium (LD) >0.1 with sliding windows of 50 SNPs and step size of 10 SNPs (option `--indep-pairwise 50 10 0.1`), and to filter for the QC thresholds previously reported.

Population structure analysis. PSD was analysed with ADMIXTURE v.1.3.0 [59] for a dual purpose. Firstly, to provide genotype-environment association tests with population structure predictors in order to reduce the risk of false positive detections [60,61]. To this aim, we decided to use membership coefficients for the four-cluster solution (K=4), as this was reported to be the best partition based on the ADMIXTURE cross-validation index for the same set of Ugandan individuals undergoing landscape genomics in the present study [22]. Due to strong collinearity ($|r|>0.7$) [45] among the membership coefficients of two ancestral components, a PCA was

281 performed trough the R function `prcomp` to obtain synthetic and orthogonal population structure
282 predictors. Secondly, to identify the main gene pools present in Uganda in the context of a
283 worldwide-extended dataset, and therefore guide selection of proper reference populations for
284 local ancestry analysis.

285 **Genotype-environment associations.** We used the software SAMβADA v.0.5.3 [22,62] to test for
286 associations between cattle genotypes and Ψ_R and γ at sampling locations. Given diploid species
287 and biallelic markers, SAMβADA runs three models per locus, one for each possible genotype.
288 Each model estimates the probability π_i for the *i-th* individual to carry a given genotype, as a
289 function of the considered environmental and population structure variables:

290 $$ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \boldsymbol{\beta} z(i) \quad (6)$$

291 and thus:

292 $$\pi_i = \frac{e^{\beta_0+\boldsymbol{\beta} z(i)}}{1+e^{\beta_0+\boldsymbol{\beta} z(i)}} \quad (7)$$

293 Genotype-environment association tests were carried out through a likelihood-ratio test
294 comparing a null and an alternative model for each genotype [22]. Particularly, null models
295 comprised the population structure predictors alone, while alternative ones included population
296 structure predictors plus either Ψ_R or γ. A genotype was considered significantly associated with
297 Ψ_R and/or γ if the resulting *p-value* associated with the likelihood-ratio test statistic (*D*) was lower
298 than the nominal significance threshold of 0.05 after Benjamini-Hochberg (BH) correction for
299 multiple testing (H_0: D=0, α_{BH}=0.05; S2 and S3 Text). The R function `p.adjust` was used to
300 perform *p-values* corrections, and predictors were centred prior to analysis to ease estimation of

model parameters.

Gene annotation. Global linkage disequilibrium (LD) decay was estimated using SNeP v.1.11 [63] to define LD extent around marker loci. A window of ±25 kbp ($r^2 \approx 0.2$) was then selected around those SNPs associated with Ψ_R and/or γ to annotate genes in the Ensembl database release 87 [64]. Annotated genes were investigated for known biological function according to the literature, and candidate genes identified based on their pertinence with ECF local adaptation.

Local ancestry

Molecular dataset. Target population for local ancestry analysis comprised 102 indigenous Ugandan cattle individuals collected during the NextGen sampling campaign (two animals sampled per cell; Fig 1C), and genotyped with the BovineHD BeadChip (777,961 SNPs; Illumina Inc., San Diego, CA, USA). Reference populations (see Results section) were selected in coherence with the major Ugandan gene pools identified by the ADMIXTURE analysis (S4 Text). Target and reference populations were pooled in a "local ancestry dataset" (LAD). Only autosomal SNPs passing the same filtering parameters applied to LGD were retained for analysis.

PCADMIX analysis. Local ancestry investigation allows to assign the ancestral origin of a chromosomal region (window) given two or more reference populations, and have been used to infer the admixture history of closely related groups [65], identify signals of adaptive introgression [66], and highlight target regions of recent selection [67]. Here, PCADMIX v.1.0 [68] was used to infer local genomic ancestry of the Ugandan samples. Given the SNPs density present in LAD (i.e. one SNP every ~3.4 kbp, on average), we used 20 SNPs per window to

obtain a window size comparable to the optimal one suggested in [68].

Beta regression analysis. Genomic windows hosting SNPs in linkage with the candidate genes for local adaptation were identified and their ancestry proportions computed per sampling cell (Fig 1C). Average Ψ_R and γ per cell values (hereafter Ψ_{Rc} and γ_c, respectively) were derived using the `zonal.stats` function included in the R package `spatialEco` [69]. In order to test for significant associations between ancestry proportions and Ψ_{Rc} and γ_c, a beta regression analysis was performed using the R package `betareg` [70], according to the model:

$$ln\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_i \quad (8)$$

$$a_i \sim B(\mu_i, \phi) \quad (9)$$

Where a_i is the ancestry proportion observed in cell i, which is assumed to derive from a beta distribution B(μ_i, □) with mean μ_i=E(a_i) and precision parameter □, x_i is either average Ψ_R or γ in cell i, β_0 and β_1 are intercept and regression coefficient, respectively. Expected ancestry proportion in i was calculated through the inverse logit:

$$\mu_i = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} \quad (10)$$

Ancestry proportions were transformed prior to analysis [71], and the Maximum Likelihood criterion was used to estimate model parameters.

Ethics Statement

The NextGen sampling campaign was carried out during years 2011 and 2012, before

Directive 2010/63/EU came into force (i.e., 1 January 2013). Thus, all experimental procedures were compliant with the former EU Directive 86/609/EEC, according to which no approval from dedicated animal welfare/ethics committee was needed for this study. The permission to carry out the study was obtained from the Uganda National Council for Science and Technology (UNCST) reference number NS 325 [11]. The permission to carry out the sampling at each farm was obtained directly from the owners.

Results

Ecological modelling

Species distribution models. The first three principal components (PC_1, PC_2, and PC_3) accounted for more than 95% of the variance among the BIO predictors, and were subsequently tested into multivariate `Maxlike` models to estimate Ψ_R. Particularly, PC_1 (61%) was mainly correlated with BIO variables linked to temperature (BIO_8, BIO_9, BIO_{10} and BIO_{11}), PC_2 (19%) with precipitation (BIO_{16}, BIO_{17}, BIO_{18} and BIO_{19}), and PC3 (15%) with both temperature and precipitation (BIO_{19} and BIO_8) (Table 1 and S5 Fig.). The model employing PC_1, PC_2, and PC_3 was selected based on the BIC metric (S6 Fig.), with PC_1 and PC_2 showing a significant positive effect on the tick distribution, and PC_3 a significant negative effect (Table 2) (H_0: β_i=0, α=0.05). The model predicts low habitat suitability in the regions North of the Lakes Kwania, Kyoga and Kojwere ($0<\Psi_R<0.1$), and favourable ecological conditions around Lake Victoria ($0.4<\Psi_R<1$) and South-West of Lake Albert ($0.4<\Psi_R<0.8$), these latter separated by a corridor of lower suitability ($0<\Psi_R<0.3$) (Fig 2A and S7 Fig.).

Table 2. `Maxlike` results for *R. appendiculatus* distribution model.

Coefficient	Estimate	SE	*p-value*	OR[a]	OR_{low}[b]	OR_{up}[c]
β_0	−2.905	0.561	2.24E−07***	0.055	0.018	0.164
PC_1	0.796	0.303	8.56E−03**	2.217	1.224	4.014
PC_2	0.822	0.37	2.62E−02*	2.275	1.102	4.698
PC_3	−1.799	0.629	4.27E−03**	0.165	0.048	0.568

Point estimates (Estimate) of the standardized regression coefficients (Coefficient) are reported on the logit scale together with their standard errors (SE), *p-values* and the associated odds ratios (OR). A significant effect is reported with *** when the *p-value* (*p*) associated to a regression coefficient is ≤0.001; ** when 0.001<*p*<0.01; * when 0.01<*p*<0.05; . when 0.01<*p*<0.1.

[a]Odds ratios associated to regression coefficients express the expected change in the ratio $\Psi_R/(1-\Psi_R)$, for a one standard deviation increase of the concerned predictor, holding all the other predictors fixed at a constant value. [b]Odds ratio 95% confidence interval (CI), lower bound. [c]Odds ratio 95% CI, upper bound.

No excessive collinearity was recorded among the predictors for Ψ_S. The best model according to the BIC metric included: altitude, annual precipitation, average NDVI and distance from the nearest water source (Table 3 and S8 Fig.). The model predicts the highest habitat suitability ($0.2<\Psi_S<0.8$) in the near proximity of the water bodies (especially along the White Nile in the North-West, the south-eastern coasts of Lake Édouard, and the northern coasts of Lake George), and in small areas near the Katonga Game Reserve (S9 Fig.).

Table 3. `Maxlike` results for *S. caffer* distribution model.

Coefficient	Estimate	SE	*p-value*	OR[a]	OR_{low}[b]	OR_{up}[c]
β_0	−9.130	0.790	6.46E−31***	0.000	0.000	0.001

Altitude	−1.095	0.293	1.90E−04***	0.335	0.188	0.594
BIO_{12}	−0.800	0.180	9.03E−06***	0.449	0.316	0.639
NDVI	2.862	0.329	3.38E−18***	17.499	9.181	33.343
Wd	−1.996	0.434	4.23E−06***	0.136	0.058	0.318

Point estimates (Estimate) of the standardized regression coefficients (Coefficient) are reported on the logit scale together with their standard errors (SE), *p-values* and the associated odds ratios (OR). Significant regression coefficients are highlighted with *** when their *p-values* (p) are ≤0.001; ** when 0.001<p≤0.01; * when 0.01<p≤0.05; . when 0.05<p≤0.1.

[a]Odds ratios associated to regression coefficients express the expected change in the ratio $\Psi_S/(1-\Psi_S)$, for a one standard deviation increase of the concerned predictor, holding all the other predictors fixed at a constant value. [b]Odds ratio 95% confidence interval (CI), lower bound. [c]Odds ratio 95% CI, upper bound.

Infection risk model. Following outliers inspection, Ψ_R, Cd and Ψ_S were transformed on the $\log_{10}$ scale to reduce the observed skewness in the distributions (S3 Fig.). No excessive collinearity was observed among the model predictors (|r|<0.7). All the explanatory variables except for Cd showed a significant effect (H_0: β_i=0, α=0.05) on infection risk. Particularly, BIO_5 and Ψ_R showed a negative association with γ, while Ψ_S resulted positively associated (Table 4). Overall, northern regions of Uganda present a low probability of infection (0.1<γ<0.3). A similar range is observed southwards, in the region comprised between Lake Kyoga, Lake Victoria, Lake Albert and the eastern borders with Kenya. South-westwards, infection probability increases following a positive gradient from γ≈0.30 to γ≈0.70 in the most southern districts (Fig 2B).

Table 4. *T. parva parva* infection risk model results.

Coefficient	Estimate	SE	*p-value*	OR[a]	OR_{low}[b]	OR_{up}[c]
β_0[d]	−1.128	0.115	1.21E−22***	0.324	0.258	0.406
$\log_{10}(\Psi_R)$	−0.219	0.105	3.72E−02*	0.803	0.654	0.987
BIO_5	−0.432	0.104	3.18E−05***	0.649	0.529	0.796
$\log_{10}$(Cd)	0.015	0.105	8.86E−01	1.015	0.826	1.247
$\log_{10}(\Psi_S)$	0.246	0.111	2.67E−02*	1.279	1.029	1.590

Point estimates (Estimate) of the standardized regression coefficients (Coefficient) are reported on the logit scale together with their standard errors (SE), *p-values* and the associated odds ratios (OR). Significant regression coefficients are highlighted with *** when their *p-values* (p) are ≤0.001; ** when 0.001<p≤0.01; * when 0.01<p≤0.05; . when 0.05<p≤0.1.

[a]Odds ratios associated to regression coefficients express the expected change in the ratio $\gamma/(1-\gamma)$, for a one standard deviation increase of the concerned predictor, holding all the other predictors fixed at a constant value. [b]Odds ratio 95% confidence interval (CI), lower bound. [c]Odds ratio 95% CI, upper bound. [d]Population intercept.

Landscape genomics

Population structure analysis. After pruning for MAF, LD, genotype and individual call rates, PSD counted 12,925 SNPs and 1,355 individuals, among which 743 from Uganda, 131 European taurine, 158 African taurine, 195 sanga from outside Uganda, and 128 zebu cattle.

Sanga and zebuine ancestries were the most represented in Uganda. Particularly, on average the sanga component constituted 76% (±13%) of the individual ancestries, whereas the zebuine counted 18% (±13%), with more than half of the individuals showing a zebuine proportion >20%. Further, ~3% of African and European taurine genomic ancestry components was also observed. Genomic components showed spatial structure, the zebuine gene pool being more

present in the North-East of the country, and the sanga in central and south-western Uganda (S10 Fig.) [22]. The African taurine ancestry component was detectable as background signal especially in the North-West and South-West, whereas European introgression was mostly observed in the South-West.

The first three principal components (PC_1, PC_2 and PC_3, respectively) explained almost the totality of the variance within ADMIXTURE Q-scores for K=4; PC_1 split the dataset between sanga and zebu gene pools, and PC_2 and PC_3 identified the European and African taurine components, respectively. Thus, these three PCs were used as population structure predictors to account for population structure within LGD in the landscape genomics models.

Genotype-environment associations. After QC, LGD counted 40,886 markers and 743 animals (the same in PSD) from 199 farms (4±1 samples/farm), over 51 cells (15±5 samples/cell).

Sixty-three genotypes across 41 putative adaptive loci resulted significantly associated with Ψ_R (Fig 3A, S2 Table, and S11-S12 Figs). Eight genotypes across seven loci resulted significantly associated with γ (Fig 3B, S3 Table, and S11-S12 Figs).

Gene annotation. Among the 41 loci significantly associated with Ψ_R, 18 presented at least one annotated gene in the Ensembl database in close proximity (Table 5A and S12 Fig.). Locus BTA-113604-no-rs (hereafter BTA-113604) is located ~12.5 kbp apart from the Protein kinase, cGMP-dependent, type I (*PRKG1*) gene on chromosome 26. *PRKG1* was already proposed as a candidate gene for tick resistance in South African Nguni cattle [72].

Six out of the seven loci significantly associated with γ presented at least one annotated gene within the selected window size (Table 5B and S12 Fig.). Two SNPs (ARS-BFGL-NGS-

437 110102 and ARS-BFGL-NGS-24867, hereafter ARS-110102 and ARS-24867, respectively) were
438 proximal to the Src-like-adaptor 2 (*SLA*2) gene on chromosome 13. *SLA*2 human orthologue
439 encodes the Src-like-adaptor 2, a member of the SLAP protein family which regulates the T and
440 B cell-mediated immune response [73]. Given *T. parva parva* known ability to promote the
441 proliferation of T and B cells [74,75], we considered *SLA*2 as a second candidate gene for ECF
442 local adaptation.

443

444 **Table 5. Gene annotation for the loci significantly associated with Ψ_R (A) and γ (B).**

A					
SNP ID	**Genotype(s)**	**Chr.**	**Position**	**Annotated gene**	**Biological function**
ARS-BFGL-NGS-110339	AA, AC	1	111,495,891	Uncharacterized	-
Hapmap34409-BES7_Contig244_858	AA	1	120,149,924	Glycogenin-1 (*GYG1*)	Energy metabolism and angiogenesis [76]
Hapmap34056-BES2_Contig421_810	AG, GG	1	138,178,130	DnaJ heat shock protein family (Hsp40) member C13 (*DNAJC13*)	Heat shock proteins [77]
ARS-BFGL-NGS-32909	CC, AC	5	67,846,632	5'-nucleotidase domain containing 3 (*NT5DC3*)	UP-regulated genes for iron content in Nelore cattle [78]
				Uncharacterized	-
ARS-BFGL-NGS-37845	AG, AA	5	48,633,731	Methionine sulfoxide reductase B3 (*MSRB3*)	Affect ear floppiness and morphology in dogs [79]

BTA-46975-no-rs	CG, GG	5	68,220,538	Thioredoxin reductase 1. cytoplasmic (*TXNRD1*)	Milk production and oocyte developmental competence in cattle [80,81]
Hapmap51626-BTA-73514	AA, AG	5	48,834,486	Inner nuclear membrane protein Man1 (*LEMD3*)	Height in pigs and cattle [82]
UA-IFASA-6140	AG, AA	7	102,472,846	ST8 alpha-N-acetyl-neuraminide alpha-2.8-sialyltransferase 4 (*ST8SIA4*)	Metabolism of milk glycoconjugates in mammals [83]
BTB-00292673	AA	7	4,953,801	Phosphodiesterase 4C (*PDE4C*)	Fertility [84]
				Member RAS oncogene family (*RAB3A*)	Calcium exocytosis in neurons [85]
				MPV17 mitochondrial inner membrane protein like 2 (*MPV17L2*)	Immune system [86]
Hapmap31116-BTA-143121	AA	8	7,597,3285	Epoxide hydrolase 2 (*EPHX2*)	In vitro maturation. fertilization and culture on bovine embryos [87]
				L-gulonolactone oxidase (*GULO*)	Involved into vitamin C production in pigs [88]

ARS-BFGL-NGS-104610	AG	11	104,293,559	Surfeit 6 (*SURF6*)	Housekeeping gene [89]
				Mediator complex subunit 22 (*MED22*)	Gestation length in Nelore cattle [90]
				Ribosomal protein L7a (*RPL7A*)	Oocyte developmental competence in cattle [80]
				Uncharacterized	-
				Small nucleolar RNA (*SNORD24*)	May act as methylation guide for RNA targets [91]
				Small nucleolar RNA (*SNORD36*)	2'-O-ribose methylation guide [92]
				Small nucleolar RNA (*snR47*)	2'-O-methylation of large and small subunit rRNA [93]
				Small nucleolar RNA (*SNORD24*)	As above
				Small nucleolar RNA (*SNORD36*)	As above

BTB-00839408	AG. AA	22	18,978,658	Metabotropic glutamate receptor 7 precursor (*GRM7*)	Might be related to parasite resistance [94]
ARS-BFGL-NGS-39898	GG	22	1,319,636	Novel gene	-
ARS-BFGL-BAC-31319	AA	23	4,847,028	3-hydroxymethyl-3-methylglutaryl-CoA lyase like 1 (*HMGCLL1*)	Involved into ketogenesis [95]
Hapmap51155-BTA-11643	AA	24	38,086,180	DLG associated protein 1 (*DLGAP1*)	Role in neurological development and behavioral disorders [96]
Hapmap57868-rs29020458	AA	24	22,746,291	Dystrobrevin alpha (*DTNA*)	Formation and stability of synapses [97]
				U6 spliceosomal RNA (*U6*)	Participate into spliceosome formation [98]
BTA-113604-no-rs	AA	26	8,356,096	Protein kinase. cGMP-dependent. type I (*PRKG1*)	Tick resistance in South African Nguni cattle [72]
ARS-BFGL-NGS-18933	GG	29	34,650,967	Opioid binding protein/cell adhesion molecule like (*OPCML*)	Role in opioid receptor function in humans [99]
B					

BTB-01298953	AA	4	54,930,726	Protein phosphatase 1 regulatory subunit 3A (*PPP1R3A*)	Glycogen synthesis in humans and mice [100]
BTA-33234-no-rs	GG	13	66,291,997	DLG associated protein 4 (*DLGAP4*)	Neuronal membrane protein [101]
				Myosin light chain 9 (*MYL9*)	May participate in regulation of muscle contraction [102]
ARS-BFGL-NGS-112656	AA	13	66,336,246	Myosin light chain 9 (*MYL9*)	As above
				TGFB induced factor homeobox 2 (*TGIF2*)	Transcriptional repressor [103]
ARS-BFGL-NGS-110102	GG	13	66,370,867	TGFB induced factor homeobox 2 (*TGIF2*)	As above
				TGIF2-C20orf24 readthrough (*C13H20orf24* alias *RIP5*)	May promote apoptosis in humans [104]
				Src-like-adaptor 2 (*SLA*2)	Downregulation of T and B cell-mediated responses [73]
ARS-BFGL-NGS-24867	AA	13	66,395,465	Src-like-adaptor 2 (*SLA*2)	As above

				NDRG family member 3 (*NDRG3*)	Linked to prostate cancer cells growth [105]
Hapmap39482-BTA-36746	CC, AC	15	40,279,014	TEA domain transcription factor 1 (*TEAD1*)	Transcription factor promoting apoptosis in mammals [106]

Markers (SNP ID) in linkage disequilibrium with a gene annotated in the Ensembl database are reported together with the associated genotype(s), chromosome (Chr), physical position in base pairs on the chromosome, as well as name and biological function of the annotated gene (as found for a reference species).

Local ancestry

PCAdmix and Beta regression analyses. Based on the gene pools revealed by ADMIXTURE analysis in Ugandan indigenous cattle, we performed PCADMIX analysis using one zebuine (Tharparkar; THA) and one African taurine (Muturu; MUT) reference (S4 Text). After QC, LAD counted 689,339 markers and 128 individuals (102 Ugandan cattle individuals, 13 THA, and 13 MUT).

For the genomic window hosting BTA-113604 (i.e. window 13 on chromosome 26; S13 Fig.), 79 out of the 204 haploid individuals targeted showed MUT ancestry, while 125 THA ancestry (S14 Fig.). For the genomic window hosting ARS-110102 and ARS-24867 (i.e. window 145 on chromosome 13; S13 Fig.), 63 haploid individuals were assigned to MUT, while 141 to THA (S14 Fig.).

Tharparkar ancestry at window 13 of chromosome 26 showed a positive and significant association with Ψ_{Rc} (H_0: β_i=0, α=0.05) (Table 6 and Fig 4), while no significant association was found between the Muturu/Tharparkar ancestries at window 145 of chromosome 13 and γ_c (S5 Text).

Table 6. Beta regression results.

Coefficient	Estimate	SE	*p-value*	OR	OR_{low} [a]	OR_{up} [b]
β_0	0.144	0.194	4.56E−01	1.155	0.790	1.689
Ψ_{Rc}	1.663	0.768	3.04E−02*	5.275	1.171	23.767
ϕ	2.029	0.346				

Association between the inferred proportion of THA ancestry at window 13 (chromosome 26) with average *R. appendiculatus* occurrence probability per sampling cell (Ψ_{Rc}). Point estimates (Estimate) of the intercept (β_0), the regression coefficient associated to Ψ_{Rc} and the precision parameter ϕ are reported on the logit scale together with their standard errors (SE). *P-values* and odds ratios (OR) are shown for β_0 and Ψ_{Rc}. Significant regression coefficients are highlighted with *** when their *p-values* (p) are ≤0.001; ** when 0.001<p≤0.01; * when 0.01<p≤0.05; .

when $0.05<p\leq0.1$.

[a]Odds ratio 95% confidence interval (CI), lower bound. [b]Odds ratio 95% CI, upper bound.

Discussion

East Coast fever represents a major issue for livestock health in sub-Saharan countries [107], with over one million cattle deceased every year, and an annual economic damage of 168-300 million USD [2,108].

ECF incidence is highly correlated with the geographical distribution of the tick vector *R. appendiculatus*, whose occurrence is an essential precondition for *T. parva parva* infection in cattle [3]. However, with the present study we show that areas with predicted poor habitat suitability for the tick can present higher infection rates when compared with regions highly suitable for the tick (Fig 2 and Table 4). Such observation suggests that *T. parva parva* occurrence cannot be explained by the sole presence of its vector. Here, we suggest three possible hypotheses to explain such a counterintuitive pattern.

First, environmental temperature may play a pivotal role in defining *T. parva parva* infection risk in cattle. Piroplasm development within the tick vector appears to be hindered by temperatures >28°C persisting even for short time periods (as less as 15 days) [54]. Therefore, areas exceeding this temperature threshold might present a reduced infection risk due to the low success in parasite development and transmission. The presence of such a temperature constraint might concur in explaining the low infection risk predicted in the regions such as North-East of Lake Victoria, where a highly suitable habitat is predicted for *R. appendiculatus*, but where temperature can reach 30°C in the warmest month of the year (January) [27]. Coherently, in the south-western area, environmental

489 temperature ranges between ~8-28°C during the whole year [27], and the predicted risk of infection
490 increases despite the decrease in habitat suitability for the tick.

491 Second, the most suitable areas for the vector overlap those regions where the highest levels of
492 zebuine ancestry were recorded (S10 Fig.). *B. indicus* is known to be more effective in counteracting
493 tick infestation than *B. taurus* [109–112], and is consequently less affected by tick-borne micro-
494 organisms [111], including *T. parva parva*, whose effects are known to be dose-dependent [107,113].
495 The core adaptive response to tick burden was identified as the inflammatory reaction triggered by the
496 tick bite at the cutaneous level [111], which activates a strong white cells-mediated cutaneous reaction
497 [114] affecting attachment, salivation, engorgement, and ultimately limiting the inoculation of tick-
498 borne microorganisms [115]. Therefore, the low infection risk observed in the most suitable areas for
499 *R. appendiculatus* (e.g. north-eastern districts) might be explained by the coexistence of putative tick-
500 resistant zebuine-like populations [116], along with a sub-optimal environmental niche for the parasite.
501 Further, we speculate that cattle populations living in regions suitable for *T. parva parva* development,
502 but with reduced predicted tick presence (e.g. the southern districts), could have not underwent a tick-
503 specific adaptation, and therefore show higher infection rates.

504 Third, the *R. appendiculatus* distribution model does not explicitly consider anthropogenic
505 factors like tick-control campaigns on a local and temporal basis. However, adequate tick-control
506 campaigns are rarely undertaken in Uganda (Ugandan National Drug Authority), and evidence of *R.*
507 *appendiculatus* developing drug resistance has been recorded [117].

508 Despite infections being observed in the northern farms of Uganda, an almost null occurrence of
509 the ECF-vector is predicted for the same regions ($0<\Psi_R<0.1$; Fig 2A). A possible explanation is the
510 lack of *R. appendiculatus* records, and the consequent bias in the tick distribution model [37,118].

511 Moreover, predicted infection risk in the North ($0<\gamma<0.3$; Fig 2B) may be inflated by the global inverse
512 relationship between γ and Ψ_R as estimated by the infection risk model (Table 4), and care is
513 recommended regarding the infection risk predictions for these areas.

514 Here, we suggest that the putative adaptive component sustaining ECF-endemic stability might
515 be due to a synergic mechanism involving specific adaptations to *R. appendiculatus* (the vector) and *T.*
516 *parva parva* (the parasite). Specifically, adaptations to tick burden could be found along the Lake
517 Victoria coasts, where a higher selective pressure linked to *R. appendiculatus* occurrence is predicted
518 (Fig 2A). We identified 41 loci across 18 chromosomes significantly associated with Ψ_R (Fig 3A), with
519 the majority of putative loci under selection found on the chromosomes 5 (nine loci), 1 (seven loci),
520 and 15 (three loci). Interestingly, the large genomic region hosting the associated SNPs on chromosome
521 5 (S2 Table) overlaps a genomic region which has been previously associated with several traits in
522 tropical cattle, including parasite resistance [119]. Among the genes in LD with the associated markers,
523 we found *PRKG1* on chromosome 26 (Table 5A and S13 Fig.), a gene coding for an important
524 mediator of vasodilation, and already reported as possibly involved in tick resistance in the South
525 African Nguni breed (see Table 6 in [72]). Importantly, vasodilatation is a classical feature of the
526 inflammatory response [120,121], the core mechanisms underlying tick resistance, as discussed before.
527 None of the remaining annotated genes was easily attributable to adaptation to tick burden (Table 5A).

528 A specific adaptive response towards *T. parva parva* infection may have evolved in south-
529 western Uganda, possibly due to ecological conditions suitable for the parasite survival, and to the
530 presence of a more tick-susceptible cattle population (S10 Fig.). *Theileria parva parva* pathogenicity is
531 linked to its ability to invade host lymphocytes, and promoting their transformation and clonal
532 expansion through the activation of several host-cell signalling pathways [15,75,122]. Here, we found

533 seven markers significantly associated with γ, two of which (ARS-110102 and ARS-24867) included
534 within *SLA*2 genic region on chromosome 13 (S13 Fig.). *SLA*2 is known to be involved with signal
535 transduction in B and T cells. Further, *SLA*2 downregulates humoral and cell-mediated immune
536 responses, by contributing to a correct lymphocytes' activation and proliferation [73,123,124]. *SLA*2
537 antagonistic effect on lymphocytes proliferation would suggest its putative involvement in opposing the
538 diffusion of *T. parva parva* in the organism. However, further molecular and immunological
539 investigation will be required for validating this hypothesis.

540 Despite the genetic proximity between Muturu and some tick resistant indigenous *B. taurus*
541 breeds of western Africa (i.e. N'Dama) [111,125], local ancestry of the genomic region candidate for
542 tick resistance was predominantly assigned to Tharparkar (Fig 4, Table 6, and S14 Fig.). This result is
543 in agreement with the known resistance of zebuine cattle to ticks, and suggests the origin of tick
544 resistance in eastern Africa either from imported Indian populations or within local zebuine-like
545 populations after migration from India. Conversely, no easily-interpretable indication was obtained for
546 the genomic region candidate for tolerance to *T. parva parva* infection. Indeed, neither Tharparkar nor
547 Muturu ancestries displayed a significant association with infection risk, while an additional local
548 ancestry analysis revealed a positive correlation with the European taurine Hereford ancestry when
549 tested versus Tharparkar (S5 Text). Although surprising, this result would rather point towards a
550 taurine origin of infection tolerance. However, local ancestry results are inherently reference-dependent
551 [66], and further analyses with different African taurine and zebuine references will be required to
552 disentangle the evolutionary origin of the genomic regions under scrutiny.

553 Besides the identification of candidate regions for ECF local adaptation, our results revealed
554 allochthonous introgression from Europe within the local gene pools of Ugandan cattle (S4 Text and

555 S10 Fig.). This finding is consistent with the generalized loss of agro-biodiversity reported worldwide
556 [8,126], and stresses the importance of monitoring local genetic resources to conserve unique
557 adaptations, including tolerance to tropical endemic diseases.

558 Despite limitations in both epidemiological and species occurrence data, the proposed models
559 allowed the identification of two candidate genes for ECF-tolerance. In general, the combination of
560 ecological modelling (e.g. species distribution models) and landscape genomics showed the potential of
561 revealing candidate genomic regions for local adaptation, and could be considered in any evolutionary
562 study involving interacting species, like symbiotic relationships (i.e. mutualism, parasitism and
563 commensalism), and competition.

564 Acknowledgments

565 The authors are grateful to the members of the NextGen Consortium (http://nextgen.epfl.ch), and
566 to Graeme S. Cumming, who kindly provided the *R. appendiculatus* occurrence dataset used in the
567 present study.

568 Data Accessibility

569 Raster data are available from the public sources mentioned in the references and in S1 Text.
570 Datasets and source code are available from the Dryad Digital Repository (doi:10.5061/dryad.sf5j2bf).

571 Author Contributions

572 Conceptualization: EV, MB, LC, MM, ER, EF, MDC, RN, SJ, PAM. Data Curation: EV, MB,

LC, MM. Formal Analysis: EV, MB. Methodology: EV, MB, ER, EF, SJ, PAM. Software: EV, MB, MM, MDC. Visualization: EV. Original Draft Preparation: EV. Review and Editing: MB, LC, MM, ER, SJ, TSS, PAM. Investigation: LC, MM, FK, RN. Funding Acquisition: LC, RN, SJ, PAM. Supervision: LC, SJ, PAM. Project Administration: SJ, PAM. Resources: CMK, CMS, VM, FK, TSS, HJH.

References

1. Muhanguzi D, Picozzi K, Hatendorf J, Thrusfield M, Welburn SC, Kabasa JD, et al. Prevalence and spatial distribution of Theileria parva in cattle under crop-livestock farming systems in Tororo District, Eastern Uganda. Parasit Vectors. 2014;7: 91.

2. Norval RAI, Perry BD, Young AS. The epidemiology of theileriosis in Africa [Internet]. ILRI (aka ILCA and ILRAD); 1992. Available: http://books.google.it/books?hl=it&lr=&id=_eIeTH63SzIC&oi=fnd&pg=PP9&dq=norval+et+al+theileriosis&ots=7BPcXR0DUT&sig=NuMl28IJ76o5WJsTJYQents3ZbM

3. Olwoch JM, Reyers B, Engelbrecht FA, Erasmus BFN. Climate change and the tick-borne disease, Theileriosis (East Coast fever) in sub-Saharan Africa. J Arid Environ. 2008;72: 108–120. doi:10.1016/j.jaridenv.2007.04.003

4. Oura CAL, Tait A, Asiimwe B, Lubega GW, Weir W. Haemoparasite prevalence and Theileria parva strain diversity in Cape buffalo (Syncerus caffer) in Uganda. Vet Parasitol. 2011;175: 212–219. doi:10.1016/j.vetpar.2010.10.032

5. Epstein H. On the Classification of Cattle. The Origin of Domestic Animals of Africa. New York: African Publishing Corporation; 1971.

6. Freeman AR. Assessing the Relative Ages of Admixture in the Bovine Hybrid Zones of Africa and the Near East Using X Chromosome Haplotype Mosaicism. Genetics. 2006;173: 1503–1510. doi:10.1534/genetics.105.053280

7. Magee DA, MacHugh DE, Edwards CJ. Interrogation of modern and ancient genomes reveals the complex domestic history of cattle. Anim Front. 2014;4: 7–22. doi:10.2527/af.2014-0017

8. Mwai O, Hanotte O, Kwon Y-J, Cho S. Invited Review - African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. Asian-Australas J Anim Sci. 2015;28: 911–921. doi:10.5713/ajas.15.0002R

9. Hayashida K, Abe T, Weir W, Nakao R, Ito K, Kajino K, et al. Whole-Genome Sequencing of Theileria parva Strains Provides Insight into Parasite Migration and Diversification in the African Continent. DNA Res. 2013;20: 209–220. doi:10.1093/dnares/dst003

10. Sivakumar T, Hayashida K, Sugimoto C, Yokoyama N. Evolution and genetic diversity of Theileria. Infect Genet Evol. 2014;27: 250–263. doi:10.1016/j.meegid.2014.07.013

11. Kabi F, Masembe C, Muwanika V, Kirunda H, Negrini R. Geographic distribution of non-clinical Theileria parva infection among indigenous cattle populations in contrasting agro-ecological zones of Uganda: implications for control strategies. Parasit Vectors. 2014;7. doi:10.1186/1756-3305-7-414

12. Bahbahani H, Hanotte O. Genetic resistance: tolerance to vector-borne diseases and the prospects and challenges of genomics. Rev Sci Tech Int Epiz. 2015;34: 185–197.

13. Gachohi J, Skilton R, Hansen F, Ngumi P, Kitala P, others. Epidemiology of East Coast fever (Theileria parva infection) in Kenya: past, present and the future. Parasit Vectors. 2012;5: 194.

14. Kivaria FM, Heuer C, Jongejan F, Okello-Onen J, Rutagwenda T, Unger F, et al. Endemic stability for Theileria parva infections in Ankole calves of the Ankole ranching scheme, Uganda. Onderstepoort J Vet Res. 2004;71: p–189.

15. Chaussepied M, Janski N, Baumgartner M, Lizundia R, Jensen K, Weir W, et al. TGF-b2 Induction Regulates Invasiveness of Theileria-Transformed Leukocytes and Disease Susceptibility. Heussler VT, editor. PLoS Pathog. 2010;6: e1001197. doi:10.1371/journal.ppat.1001197

16. Kim J, Hanotte O, Mwai OA, Dessie T, Bashir S, Diallo B, et al. The genome landscape of indigenous African cattle. Genome Biol. 2017;18. doi:10.1186/s13059-017-1153-y

17. Rubaire-Akiiki CM, Okello-Onen J, Musunga D, Kabagambe EK, Vaarst M, Okello D, et al. Effect of agro-ecological zone and grazing system on incidence of East Coast Fever in calves in Mbale and Sironko Districts of Eastern Uganda. Prev Vet Med. 2006;75: 251–266. doi:10.1016/j.prevetmed.2006.04.015

18. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev Genet. 2013;14: 807–820. doi:10.1038/nrg3522

19. Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO. African Pastoralism: Genetic Imprints of Origins and Migrations. Science. 2002;296: 336–339. doi:10.1126/science.1069878

20. Singh J, Gill JS, Kwatra MS, Sharma KK. Treatment of theileriosis in crossbred cattle in the Punjab. Trop Anim Health Prod. 1993;25: 75–78.

635 21. Boulter N, Hall R. Immunity and vaccine development in the bovine theilerioses. Adv Parasitol.
636 1999;44: 41–97.

637 22. Stucki S, Orozco-terWengel P, Forester BR, Duruz S, Colli L, Masembe C, et al. High
638 performance computation of landscape genomic models including local indicators of spatial
639 association. Mol Ecol Resour. 2017;17: 1072–1089. doi:10.1111/1755-0998.12629

640 23. Kawecki TJ, Ebert D. Conceptual issues in local adaptation. Ecol Lett. 2004;7: 1225–1241.
641 doi:10.1111/j.1461-0248.2004.00684.x

642 24. Stockwell CA, Hendry AP, Kinnison MT. Contemporary evolution meets conservation biology.
643 Trends Ecol Evol. 2003;18: 94–101.

644 25. Crispo E, DiBattista JD, Correa C, Thibert-Plante X, McKellar AE, Schwartz AK, et al. The
645 evolution of phenotypic plasticity in response to anthropogenic disturbance. Evol Ecol Res.
646 2010;12: 47–66.

647 26. Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB. Extent and scale of local adaptation
648 in salmonid fishes: review and meta-analysis. Heredity. 2011;106: 404–420.

649 27. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated
650 climate surfaces for global land areas. Int J Climatol. 2005;25: 1965–1978. doi:10.1002/joc.1276

651 28. Jarvis A, Reuter HI, Nelson A, Guevara E. Hole-filled SRTM for the globe Version 4, available
652 from the CGIAR-CSI SRTM 90m Database. 2008; Available: http://srtm.csi.cgiar.org

653 29. Hijmans RJ. raster: Geographic Data Analysis and Modeling [Internet]. 2016. Available:
654 https://CRAN.R-project.org/package=raster

655 30. US Geological Survey (USGS). Earth Resources Observation and Science (EROS) Center -
656 Famine Early Warning Systems Network (FEWS NET) [Internet]. Available:
657 https://earlywarning.usgs.gov/fews

658 31. Robinson TP, Wint GRW, Conchedda G, Van Boeckel TP, Ercoli V, Palamara E, et al. Mapping
659 the Global Distribution of Livestock. Baylis M, editor. PLoS ONE. 2014;9: e96084.
660 doi:10.1371/journal.pone.0096084

661 32. CIESIN (Center for International Earth Science Information Network) Columbia University.
662 Gridded Population of the World, Version 4 (GPWv4): Land and Water Area. Palisades, NY:
663 NASA Socioeconomic Data and Applications Center (SEDAC) [Internet]. 2016. Available:
664 http://dx.doi.org/10.7927/H45M63M9

665 33. Merow C, Silander JA. A comparison of Maxlike and Maxent for modelling species
666 distributions. Warton D, editor. Methods Ecol Evol. 2014;5: 215–225. doi:10.1111/2041-
667 210X.12152

668 34. Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, et al. The
669 crucial role of the accessible area in ecological niche modeling and species distribution
670 modeling. Ecol Model. 2011;222: 1810–1819. doi:10.1016/j.ecolmodel.2011.02.011

671 35. QGIS Development Team. QGIS Geographic Information System [Internet]. Open Source
672 Geospatial Foundation Project; 2016. Available: http://qgis.osgeo.org

673 36. Royle JA, Chandler RB, Yackulic C, Nichols JD. Likelihood analysis of species occurrence
674 probability from presence-only data for modelling species distributions: *Likelihood analysis of*
675 *presence-only data*. Methods Ecol Evol. 2012;3: 545–554. doi:10.1111/j.2041-
676 210X.2011.00182.x

677 37. Cumming GS. The evolutionary ecology of African ticks. Unpublished DPhil Thesis. University
678 of Oxford. 1999.

679 38. GBIF. Recommended practices for citation of the data published through the GBIF Network.
680 Version 1.0 (Authored by Vishwas Chavan), Copenhagen: Global Biodiversity Information
681 Facility. Pp.12, ISBN: 87-92020-36-4. Accessible at
682 http://links.gbif.org/gbif_best_practice_data_citation_en_v1. 2012;

683 39. Cumming GS. Host distributions do not limit the species ranges of most African ticks (Acari:
684 Ixodida). Bull Entomol Res. 1999;89: 303–327.

685 40. Cumming GS. Comparing Climate and Vegetation as Limiting Factors for Species Ranges of
686 African Ticks. Ecology. 2002;83: 255. doi:10.2307/2680136

687 41. Matawa F, Murwira A, Schmidt KS. Explaining elephant (Loxodonta africana) and buffalo
688 (Syncerus caffer) spatial distribution in the Zambezi Valley using maximum entropy modelling.
689 Ecol Model. 2012;242: 189–197. doi:10.1016/j.ecolmodel.2012.05.010

690 42. Naidoo R, Du Preez P, Stuart-Hill G, Jago M, Wegmann M. Home on the Range: Factors
691 Explaining Partial Migration of African Buffalo in a Tropical Environment. Boyce MS, editor.
692 PLoS ONE. 2012;7: e36527. doi:10.1371/journal.pone.0036527

693 43. Pettorelli N, Ryan S, Mueller T, Bunnefeld N, Jedrzejewska B, Lima M, et al. The Normalized
694 Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. Clim Res.
695 2011;46: 15–27. doi:10.3354/cr00936

696 44. Winnie JA, Cross P, Getz W. Habitat quality and heterogeneity influence distribution and
697 behavior in African buffalo (Syncerus caffer). Ecology. 2008;89: 1457–1468.

698 45. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of
699 methods to deal with it and a simulation study evaluating their performance. Ecography.
700 2013;36: 27–46. doi:10.1111/j.1600-0587.2012.07348.x

46. Jolliffe I. Principal Component Analysis. New York: Springer; 2002.

47. Aho K, Derryberry D, Peterson T. Model selection for ecologists: the worldviews of AIC and BIC. Ecology. 2014;95: 631–636.

48. Bring J. How to Standardize Regression Coefficients. Am Stat. 1994;48: 209–213. doi:10.2307/2684719

49. Cade BS. Model averaging and muddled multimodel inferences. Ecology. 2015;96: 2370–2382.

50. Gachohi JM, Kitala PM, Ngumi PN, Skilton RA. Environment and farm factors associated with exposure to Theileria parva infection in cattle under traditional mixed farming system in Mbeere District, Kenya. Trop Anim Health Prod. 2011;43: 271–277. doi:10.1007/s11250-010-9688-x

51. Magona JW, Walubengo J, Olaho-Mukani W, Jonsson NN, Welburn SC, Eisler MC. Clinical features associated with seroconversion to Anaplasma marginale, Babesia bigemina and Theileria parva infections in African cattle under natural tick challenge. Vet Parasitol. 2008;155: 273–280. doi:10.1016/j.vetpar.2008.05.022

52. Magona JW, Walubengo J, Olaho-Mukani W, Jonsson NN, Welburn SW, Eisler MC. Spatial variation of tick abundance and seroconversion rates of indigenous cattle to Anaplasma marginale, Babesia bigemina and Theileria parva infections in Uganda. Exp Appl Acarol. 2011;55: 203–213. doi:10.1007/s10493-011-9456-2

53. Billiouw M, Vercruysse J, Marcotty T, Speybroeck N, Chaka G, Berkvens D. Theileria parva epidemics: a case study in eastern Zambia. Vet Parasitol. 2002;107: 51–63.

54. Young AS, Leitch BL. Epidemiology of East Coast fever: some effects of temperature on the development of Theileria parva in the tick vector Rhipicephalus appendiculatus. Parasitology. 1981;83: 199–211.

55. Zuur AF, Ieno EN, Walker NJ, Savaliev AA, Smith GM. Mixed Effects Models ans Extentions in Ecology with R [Internet]. Springer Science+Business Media, LCC; 2009. Available: DOI 10.1007/978-0-387-87458-6_1

56. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015;67: 1–48. doi:10.18637/jss.v067.i01

57. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81: 559–575. doi:10.1086/519795

58. Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, et al. Quality Control Procedures for Genome-Wide Association Studies. In: Haines JL, Korf BR, Morton CC, Seidman CE, Seidman JG, Smith DR, editors. Current Protocols in Human Genetics. Hoboken,

NJ, USA: John Wiley & Sons, Inc.; 2011. Available: http://doi.wiley.com/10.1002/0471142905.hg0119s68

59. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19: 1655–1664. doi:10.1101/gr.094052.109

60. Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. Mol Ecol. 2015;24: 4348–4370. doi:10.1111/mec.13322

61. Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S. Adaptive Genetic Variation on the Landscape: Methods and Cases. Annu Rev Ecol Evol Syst. 2012;43: 23–43. doi:10.1146/annurev-ecolsys-110411-160248

62. Joost S, Bonin A, Bruford MW, DespréS L, Conord C, Erhardt G, et al. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. Mol Ecol. 2007;16: 3955–3969. doi:10.1111/j.1365-294X.2007.03442.x

63. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet. 2015;6. doi:10.3389/fgene.2015.00109

64. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. Database. 2016;2016: baw093. doi:10.1093/database/baw093

65. Pa□aniuc B, Sankararaman S, Kimmel G, Halperin E. Inference of locus-specific ancestry in closely related populations. Bioinformatics. 2009;25: i213–i221. doi:10.1093/bioinformatics/btp197

66. Barbato M, Hailer F, Orozco-terWengel P, Kijas J, Mereu P, Cabras P, et al. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. Sci Rep. 2017;7. doi:10.1038/s41598-017-07382-7

67. Tang H, Choudhry S, Mei R, Morgan M, Rodriguez-Cintron W, Burchard EG, et al. Recent Genetic Selection in the Ancestral Admixture of Puerto Ricans. Am J Hum Genet. 2007;81: 626–633. doi:10.1086/520769

68. Brisbin A, Bryc K, Byrnes J, Zakharia F, Omberg L, Degenhardt J, et al. PCAdmix: Principal Components-Based Assignment of Ancestry Along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations. Hum Biol. 2012;84: 343–364. doi:10.3378/027.084.0401

69. Evans JS. spatialEco [Internet]. 2017. Available: https://CRAN.R-project.org/package=spatialEco

70. Cribari-Neto F, Zeileis A. Beta Regression in R. J Stat Softw. 2010;34.

71. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol Methods. 2006;11: 54–71. doi:10.1037/1082-989X.11.1.54

72. Mapholi NO, Maiwashe A, Matika O, Riggio V, Bishop SC, MacNeil MD, et al. Genome-wide association study of tick resistance in South African Nguni cattle. Ticks Tick-Borne Dis. 2016;7: 487–497. doi:10.1016/j.ttbdis.2016.02.005

73. Holland SJ, Liao XC, Mendenhall MK, Zhou X, Pardo J, Chu P, et al. Functional cloning of Src-like adapter protein-2 (SLAP-2), a novel inhibitor of antigen receptor signaling. J Exp Med. 2001;194: 1263–1276.

74. Baldwin CL, Black SJ, Brown WC, Conrad PA, Goddeeris BM, Kinuthia SW, et al. Bovine T cells, B cells, and null cells are transformed by the protozoan parasite Theileria parva. Infect Immun. 1988;56: 462–467.

75. Dobbelaere DA, Küenzi P. The strategies of the Theileria parasite: a new twist in host–pathogen interactions. Curr Opin Immunol. 2004;16: 524–530. doi:10.1016/j.coi.2004.05.009

76. Lancaster PA, Sharman ED, Horn GW, Krehbiel CR, Starkey JD. Effect of rate of weight gain of steers during the stocker phase. III. Gene expression of adipose tissues and skeletal muscle in growing–finishing beef cattle. J Anim Sci. 2014;92: 1462–1472.

77. Kodiha M, Frohlich M, Stochaj U. Spatial Proteomics Sheds Light on the Biology of Nucleolar Chaperones. Curr Proteomics. 2012;9: 186-216(31). doi:https://doi.org/10.2174/157016412803251824

78. Wellison Jarles da Silva D. Expressão gênica diferencial relacionada ao contenúdo de ferro no músculo em animais Nelore. Universidade Federal De São Carlos, Centro de Ciências Biológicas e da Saude, Programa de pós-graduação em genética evolutiva e biologia molecular. 2015.

79. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A Simple Genetic Architecture Underlies Morphological Variation in Dogs. Hoekstra HE, editor. PLoS Biol. 2010;8: e1000451. doi:10.1371/journal.pbio.1000451

80. Gilbert I, Robert C, Vigneault C, Blondin P, Sirard M-A. Impact of the LH surge on granulosa cell transcript levels as markers of oocyte developmental competence in cattle. Reproduction. 2012;143: 735–747. doi:10.1530/REP-11-0460

81. Ghorbani S, Tahmoorespur M, Masoudi-Nejad A, Nasiri MR, Asgari Y. Analysis of the enzyme network involved in cattle milk production using graph theory. Mol Biol Res Commun. 2015;4: 93–103.

82. Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat Genet. 2015;47: 1141–1148. doi:10.1038/ng.3394

83. Song S, Ou-Yang Y, Huo J, Zhang Y, Yu C, Liu M, et al. Molecular cloning, sequence characterization, and tissue expression analysis of three water buffalo (Bubalus bubalis) genes – ST6GAL1, ST8SIA4 and SLC35C1. Arch Anim Breed. 2016;59: 363–372. doi:10.5194/aab-59-363-2016

84. Glick G, Shirak A, Seroussi E, Zeron Y, Ezra E, Weller JI, et al. Fine Mapping of a QTL for Fertility on BTA7 and Its Association With a CNV in the Israeli Holsteins. G3 GenesGenomesGenetics. 2011;1: 65–74. doi:10.1534/g3.111.000299

85. Brondyk WH, McKiernan CJ, Fortner KA, Stabila P, Holz RW, Macara IG. Interaction cloning of Rabin3, a novel protein that associates with the Ras-like GTPase Rab3A. Mol Cell Biol. 1995;15: 1137–1143.

86. Brütting C, Emmer A, Kornhuber M, Staege M. A survey of endogenous retrovirus (ERV) sequences in the vicinity of multiple sclerosis (MS)-associated single nucleotide polymorphisms (SNPs). Mol Biol Rep. 2016;43: 827–836.

87. Smith SL, Everts RE, Sung L-Y, Du F, Page RL, Henderson B, et al. Gene expression profiling of single bovine embryos uncovers significant effects of in vitro maturation, fertilization and culture. Mol Reprod Dev. 2009;76: 38–47.

88. Hasan L, Vögeli P, Stoll P, KramerStranzinger ŠŠG, Neuenschwander S. Intragenic deletion in the gene encoding L-gulonolactone oxidase causes vitamin C deficiency in pigs. Mamm Genome. 2004;15: 323–333. doi:10.1007/s00335-003-2324-6

89. Magoulas C, Zatsepina OV, Jordan PW, et al. The SURF-6 protein is a component of the nucleolar matrix and has a high binding capacity for nucleic acids in vitro. Eur J Cell Biol. 1998;75: 174–83. doi:10.1016/s0171-9335(98)80059-9

90. Matos MC, Utsunomiya YT, Santana do Carmo A, Russiano, Vicente WRR, Grisolia AB, et al. Meta-analysis of genome-wide association studies for gestation length in Nellore cattle - Preliminary results. X Simpósio Brasileiro de Melhoramento Animal, Uberaba. Uberaba; 2013.

91. Kiss-László Z, Henry Y, Bachellerie J-P, Caizergues-Ferrer M, Kiss T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell. 1996;85: 1077–1088.

92. Galardi S, Fatica A, Bachi A, Scaloni A, Presutti C, Bozzoni I. Purified Box C/D snoRNPs Are Able To Reproduce Site-Specific 2'-O-Methylation of Target RNA In Vitro. Mol Cell Biol. 2002;22: 6663–6668. doi:10.1128/MCB.22.19.6663-6668.2002

834 93. Samarsky DA, Fournier MJ. A comprehensive database for the small nucleolar RNAs from
835 Saccharomyces cerevisiae. Nucleic Acids Res. 1999;27: 161–164.

836 94. Xu L, Hou Y, Bickhart DM, Zhou Y, Hay EH abdel, Song J, et al. Population-genetic properties
837 of differentiated copy number variations in cattle. Sci Rep. 2016;6: 23161.
838 doi:10.1038/srep23161

839 95. Tetens J, Heuer C, Heyer I, Klein MS, Gronwald W, Junge W, et al. Polymorphisms within the
840 APOBR gene are highly associated with milk levels of prognostic ketosis biomarkers in dairy
841 cows. Physiol Genomics. 2015;47: 129–137. doi:10.1152/physiolgenomics.00126.2014

842 96. Sorbolini S, Marras G, Gaspa G, Dimauro C, Cellesi M, Valentini A, et al. Detection of selection
843 signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes
844 but different selection histories. Genet Sel Evol. 2015;47. doi:10.1186/s12711-015-0128-2

845 97. Sjö A, Magnusson K, Peterson KJ. Association of α-Dystrobrevin with Reorganizing Tight
846 Junctions. J Membr Biol. 2005;203: 21–30. doi:10.1007/s00232-004-0728-1

847 98. Marz M, Kirsten T, Stadler PF. Evolution of Spliceosomal snRNA Genes in Metazoan Animals.
848 J Mol Evol. 2008;67: 594–607.

849 99. Smith MW, Clark SP, Hutchinson JS, Wei YH, Churukian AC, Daniels LB, et al. A sequence-
850 tagged site map of human chromosome 11. Genomics. 1993;17: 699–725.
851 doi:10.1006/geno.1993.1392

852 100. Savage DB, Zhai L, Ravikumar B, Choi CS, Snaar JE, McGuire AC, et al. A Prevalent Variant
853 in PPP1R3A Impairs Glycogen Synthesis and Reduces Muscle Glycogen Content in Humans
854 and Mice. Groop LC, editor. PLoS Med. 2008;5: e27. doi:10.1371/journal.pmed.0050027

855 101. Takeuchi M, Hata Y, Hirao K, Toyoda A, Irie M, Takai Y. A family of PSD-95/SAP90-
856 associated proteins localized at postsynaptic density. J Biol Chem. 1997;272: 11943–51.

857 102. Kumar CC, Mohan SR, Zavodny PJ, Narula SK, Leibowitz PJ. Characterization and differential
858 expression of human vascular smooth muscle myosin light chain 2 isoform in nonmuscle cells.
859 Biochemistry (Mosc). 1989;28: 4027–35.

860 103. Imoto I, Pimkhaokham A, Watanabe T, Saito-Ohara F, Soeda E, Inazawa J. Amplification and
861 overexpression of TGIF2, a novel homeobox gene of the TALE superclass, in ovarian cancer cell
862 lines. Biochem Biophys Res Commun. 2000;276: 264–70.

863 104. Zha J, Zhou Q, Xu L-G, Chen D, Li L, Zhai Z, et al. RIP5 is a RIP-homologous inducer of cell
864 death. Biochem Biophys Res Commun. 2004;319: 298–303. doi:10.1016/j.bbrc.2004.04.194

865 105. Lee J-C, Chiang K-C, Feng T-H, Chen Y-J, Chuang S-T, Tsui K-H, et al. The Iron Chelator,
866 Dp44mT, Effectively Inhibits Human Oral Squamous Cell Carcinoma Cell Growth in Vitro and

in Vivo. Int J Mol Sci. 2016;17: 1435. doi:10.3390/ijms17091435

106. Landin Malt A, Cagliero J, Legent K, Silber J, Zider A, Flagiello D. Alteration of TEAD1 Expression Levels Confers Apoptotic Resistance through the Transcriptional Up-Regulation of Livin. Hong W, editor. PLoS ONE. 2012;7: e45498. doi:10.1371/journal.pone.0045498

107. Nene V, Kiara H, Lacasta A, Pelle R, Svitek N, Steinaa L. The biology of Theileria parva and control of East Coast fever – Current status and future trends. Ticks Tick-Borne Dis. 2016;7: 549–564. doi:10.1016/j.ttbdis.2016.02.001

108. McLeod A, Kristjanson R. Impact of ticks and associated diseases on cattlein Asia, Australia and Africa. ILRI and eSYS report to ACIAR. Nairobi, Kenya: International Livestock Research Institute; 1999.

109. Brizuela CM, Ortellado CA, Sanchez TI, Osorio O, Walker AR. Formulation of integrated control of Boophilus microplus in Paraguay: analysis of natural infestations. Vet Parasitol. 1996;63: 95–108.

110. Jonsson NN, Piper EK, Constantinoiu CC. Host resistance in cattle to infestation with the cattle tick Rhipicephalus microplus. Parasite Immunol. 2014;36: 553–559. doi:10.1111/pim.12140

111. Mattioli RC, Pandey VS, Murray M, Fitzpatrick JL. Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (Bos taurus) and trypanosusceptible Gobra zebu (Bos indicus) cattle. Acta Trop. 2000;75: 263–277.

112. Wambura PN, Gwakisa PS, Silayo RS, Rugaimukamu EA. Breed-associated resistance to tick infestation in Bos indicus and their crosses with Bos taurus. Vet Parasitol. 1998;77: 63–70. doi:https://doi.org/10.1016/S0304-4017(97)00229-X

113. Brossard M, Wikel SK. Immunology of interactions between ticks and hosts. Med Vet Entomol. 1997;11: 270–276.

114. Willadsen P. Immediate hypersensitivity to Boophilus microplus: factors affecting hypersensitivity, and their relevance in the resistance of cattle to ticks. Proceedings of a Symposium held at the 56th Annual Conference of Australian Veterinary Association. Townsville, 14–18 May 1976, Sydney, Australia, 60-62: In: Johnston, L.A.Y., Cooper, M.G. (Eds.); 1980.

115. Wikel SK, Bergrnan D. Tick-Host Immunology: Significant Advances and Challenging Opportunities. Parasitol Today. 1997;13: 383–389.

116. Bahbahani H, Tijjani A, Mukasa C, Wragg D, Almathen F, Nash O, et al. Signatures of Selection for Environmental Adaptation and Zebu × Taurine Hybrid Fitness in East African Shorthorn Zebu. Front Genet. 2017;8. doi:10.3389/fgene.2017.00068

900 117. Vudriko P, Okwee-Acai J, Tayebwa DS, Byaruhanga J, Kakooza S, Wampande E, et al.
901 Emergence of multi-acaricide resistant Rhipicephalus ticks and its implication on chemical tick
902 control in Uganda. Parasit Vectors. 2016;9. doi:10.1186/s13071-015-1278-3

903 118. Olwoch JM, Rautenbach C de W, Erasmus BFN, Engelbrecht FA, Van Jaarsveld AS. Simulating
904 tick distributions over sub-Saharan Africa: the use of observed and simulated climate surfaces. J
905 Biogeogr. 2003;30: 1221–1232.

906 119. Porto-Neto LR, Reverter A, Prayaga KC, Chan EKF, Johnston DJ, Hawken RJ, et al. The
907 Genetic Architecture of Climatic Adaptation of Tropical Cattle. Rueppell O, editor. PLoS ONE.
908 2014;9: e113284. doi:10.1371/journal.pone.0113284

909 120. Sherwood ER, Toliver-Kinsky T. Mechanisms of the inflammatory response. Best Pract Res
910 Clin Anaesthesiol. 2004;18: 385–405. doi:10.1016/j.bpa.2003.12.002

911 121. Surks HK. cGMP-Dependent Protein Kinase I and Smooth Muscle Relaxation: A Tale of Two
912 Isoforms. Circ Res. 2007;101: 1078–1080. doi:10.1161/CIRCRESAHA.107.165779

913 122. McKeever DJ, Morrison WI. Theileria parva: the nature of the immune response and its
914 significance for immunoprophylaxis. Sci Tech Rev Off Int Epizoot Paris. 1990;9: 405–421.

915 123. Kazi JU, Kabir NN, Rönnstrand L. Role of SRC-like adaptor protein (SLAP) in immune and
916 malignant cell signaling. Cell Mol Life Sci. 2015;72: 2535–2544. doi:10.1007/s00018-015-1882-
917 6

918 124. Marton N, Baricza E, Érsek B, Buzás EI, Nagy G. The Emerging and Diverse Roles of Src-Like
919 Adaptor Proteins in Health and Disease. Mediators Inflamm. 2015;2015: 1–9.
920 doi:10.1155/2015/952536

921 125. Ibeagha-Awemu EM, Jann OC, Weimann C, Erhardt G. Genetic diversity, introgression and
922 relationships among West/Central African cattle breeds. Genet Sel Evol. 2004;36: 673–690.
923 doi:10.1051/gse:2004024

924 126. FAO. The Second Report on the State of the World's Animal Genetic Resources for Food and
925 Agriculture [Internet]. Rome: B.D. Scherf & D. Pilling. FAO Commission on Genetic Resources
926 for Food and Agriculture Assessments; 2015. Available: http://www.fao.org/3/a-
927 i4787e/index.html

928

Figures

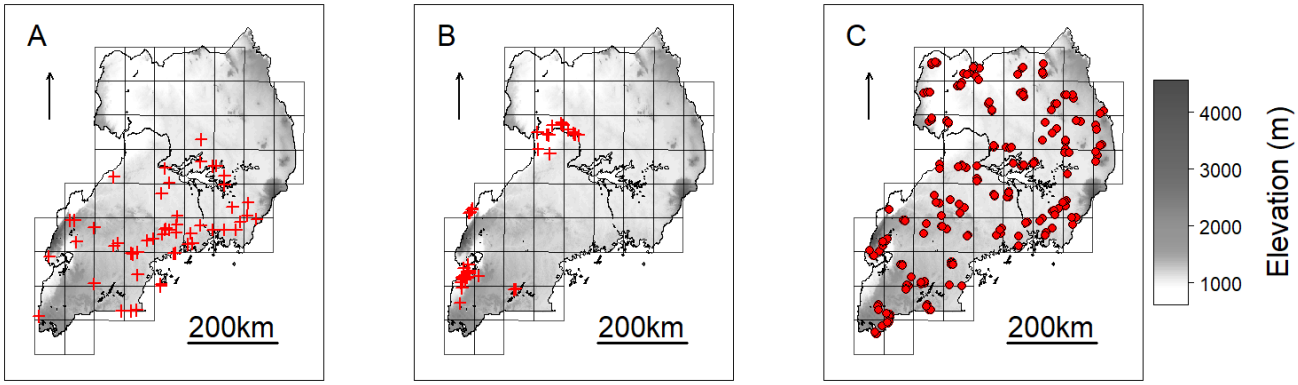

Fig 1. Occurrence records for species distribution modelling and NextGen sampling scheme. Spatial records (red crosses) used to estimate *R. appendiculatus* (A) and *S caffer* (B) distributions over Uganda, as derived from [37] and [38], respectively. Farms where cattle have been sampled to be genotyped and tested for *T. parva parva* infection are represented with red circles (C). The grid scheme used to sample farms during the NextGen project is shown on the background of each map (see main text), together with elevation.

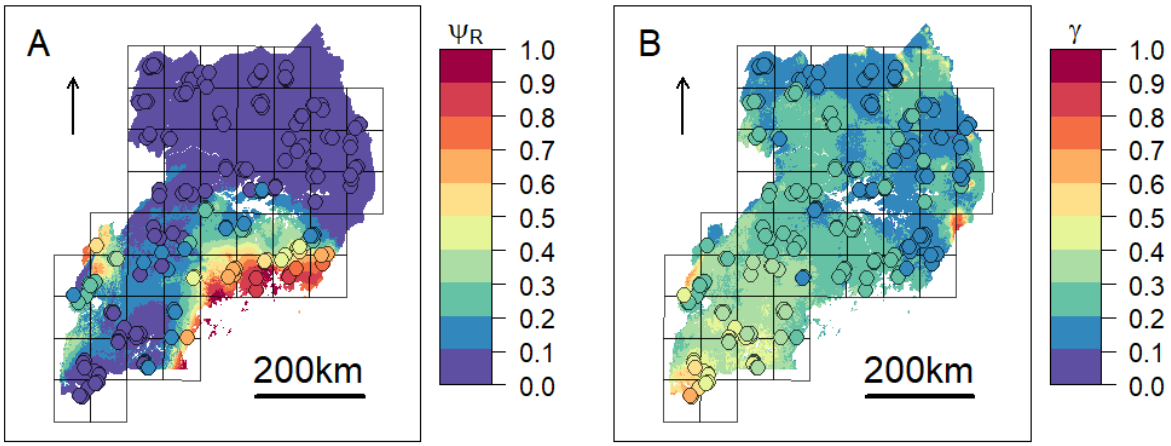

Fig 2. Predicted distributions ECF vector and infection risk. (A) *R. appendiculatus* occurrence probability (Ψ_R) as predicted by the selected distribution model. (B) Predicted *T. parva parva* infection risk (γ). Colour from blue to red tones corresponds to increasing values of Ψ_R and γ. Sampled farms are represented with circles, and coloured according to Ψ_R and γ values estimated at their geographical location.

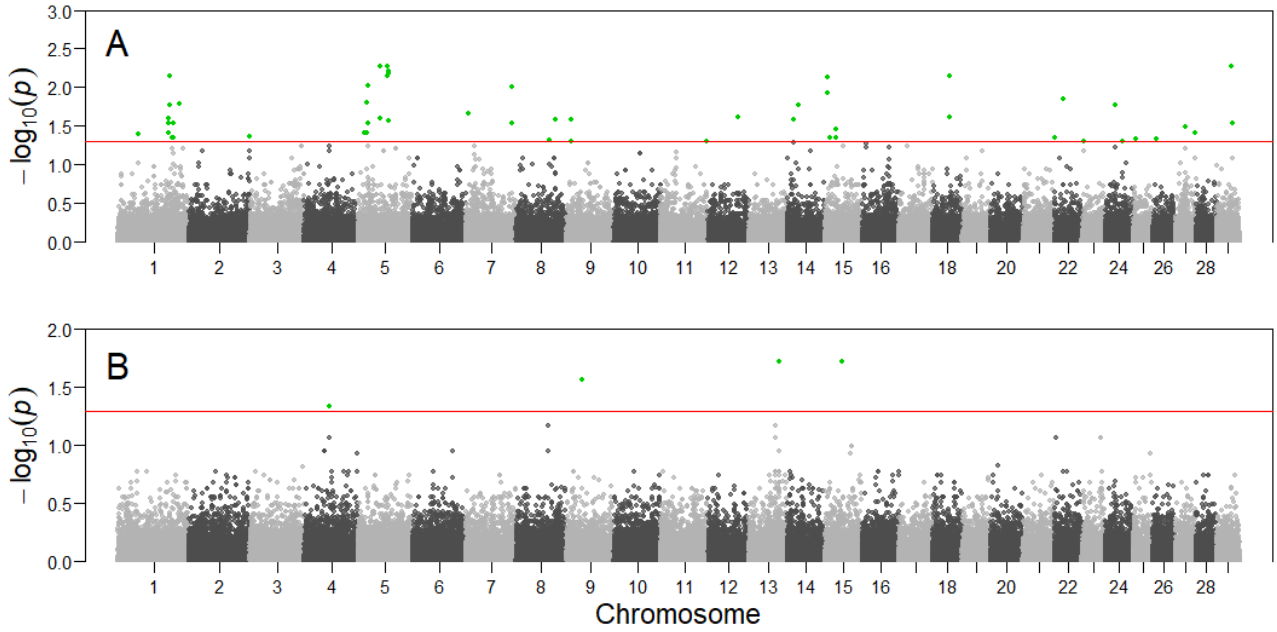

Fig 3. Manhattan plots of the genotype-environment associations. *X*-axis reports chromosomal position of the tested SNPs on cattle chromosomes. *Y*-axis reports the test statistic *p-values* (*p*) for the associations with *R. appendiculatus* occurrence probability (A), and with *T. parva parva* infection risk (B). *P-values* are displayed for each genotype after the Benjamini-Hochberg (BH) correction, and on the $-\log_{10}$ scale. Nominal significance threshold (α_{BH}=0.05) is displayed as a red line, and significant *p-values* are represented in green.

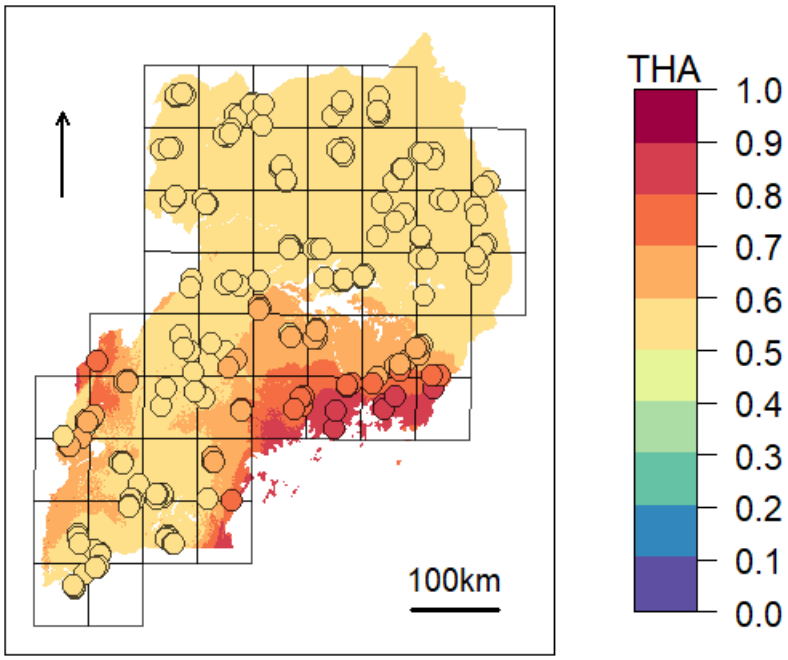

Fig 4. Expected zebuine proportion for the genomic region candidate for tick resistance over the study area. The association inferred through beta regression between Tharparkar ancestry (THA) and average *R. appendiculatus* occurrence probability per cell (Table 6) was used to generalize expected zebuine ancestry over the entire study area. Colour key corresponds to predicted THA proportion, with increasing values from blue to red tones. Sampled farms are represented with circles, and coloured according to the predicted THA proportion at their geographical location.

Supporting information

S1 Text. MODIS collection 5 data.

S2 Text. Statistical significance of genotype-environment associations.

S3 Text. Additional SAMβADA analysis (*K*=3 and *K*=16 corrections).

S4 Text. Selection of reference populations for local ancestry analyses.

S5 Text. Beta regression results on additional local ancestry analyses.

S1 Fig. Bioclimatic variables used in *R. appendiculatus* distribution model.

S2 Fig. Selection of the annual period with NDVI values best explaining *S. caffer* records.

S3 Fig. Outlier detection in the infection model predictors.

S4 Fig. Additional ADMIXTURE analysis: results.

S5 Fig. Correlations between bioclimatic variables and principal components.

S6 Fig. Candidate *R. appendiculatus* distribution models and model selection.

S7 Fig. Selected *R. appendiculatus* distribution model with confidence intervals.

S8 Fig. Candidate *S. caffer* distribution models and model selection.

S9 Fig. Selected *S. caffer* distribution model with confidence intervals.

S10 Fig. Map of the Ugandan ancestry components as derived by ADMIXTURE analysis (*K*=4).

S11 Fig. Quantile-Quantile plots of the genotype-environment association studies (*K*=4 correction).

S12 Fig. Additional SAMβADA analysis (*K*=3 and *K*=16 corrections): results.

S13 Fig. *PRKG1* and *SLA2* genomic regions.

S14 Fig. Spatial representation of PCADMIX assignments (Tharparkar/Muturu comparison).

S1 Table. Composition of the population structure dataset.

S2 Table. SAMβADA analysis (*K*=4 correction): significant associations with *R. appendiculatus* occurrence probability.

S3 Table. SAMβADA analysis (*K*=4 correction): significant associations with *T. parva parva* infection risk.