



Genomic diversity and disease prevalence in Ugandan cattle

Licia Colli¹, Lorenzo Bomba¹, Marco Milanese¹, Elia Vajana¹, Riccardo Negrini^{1,2}, Elisa Eufemi¹, Stefano Capomaccio¹, Raffaele Mazza², Alessandra Stella³, Stephane Joost⁴, Sylvie Stucki⁴, Pierre Taberlet⁵, François Pompanon⁵, Fred Kabi⁶, Vincent Muwanika⁶, Charles Masembe⁶, Paolo Ajmone Marsan¹ and The NEXTGEN Consortium^{7*}

(1) Istituto di Zootecnica e Centro di Ricerca BioDNA, Facoltà di Agraria, Università Cattolica del S. Cuore, Piacenza, Italy, (2) AIA - Associazione Italiana Allevatori, Rome, Italy, (3) IBBA-CNR and FPTP - Fondazione Parco Tecnologico Padano, Lodi, Italy, (4) EPFL, Lausanne, Switzerland, (5) LECA - Lab. d'Ecologie Alpine, UJF-CNRS, Grenoble, France, (6) Institute of Environment & Natural Resources, Makerere University, Kampala, Uganda, (7) EU funded project, <http://nextgen.epfl.ch>

Introduction

Ugandan cattle are represented by three main types: the long-horned Ankole, the short-horned zebu, and the Ankole-zebu crosses called “Nganda”. In the course of the EU-funded project Nextgen, Ugandan cattle have been extensively sampled over the whole country to investigate the association between genotypes and resistance/tolerance to endemic diseases (e.g. tsetse fly- and tick-borne diseases).

Materials & Methods

A total of 917 animals were sampled over the whole country (Figure 1). A subset of samples has been both laboratory-tested for the occurrence of East Coast Fever (PCR test, 564 samples) and Brucellosis (ELISA assay, 729 samples), and genotyped with the Illumina BovineSNP50 BeadChip (ca. 54K markers, 786 samples) SNP panel.

The genotype data were filtered with the following thresholds: SNP call rate ≤ 0.95 , individual call rate ≤ 0.95 , Minor Allele Frequency ≤ 0.01 . The molecular dataset was analyzed with ADMIXTURE software (Novembre *et al.* 2010) to assess population structure through a Bayesian clustering algorithm.

Univariate mixed logistic regression models were run to i) reveal significant environmental variables associated with the spatial occurrence of East Coast Fever and Brucellosis, and ii) model the probability of disease occurrence in function of the selected variables. The set of environmental parameters included monthly values of precipitation, minimum, mean and maximum temperature and 19 derived variables retrieved from the WorldClim dataset (<http://www.worldclim.org>), variables from the Climate Research Unit dataset (<http://www.cru.uea.ac.uk/>) and topographical variables from the Shuttle Radar Topography Mission dataset (<http://www.webgis.com/srtm3.html>). Each variable has been tested by using the function `glm` of the `lme4` R package, by modelling sampling locations as random effects.

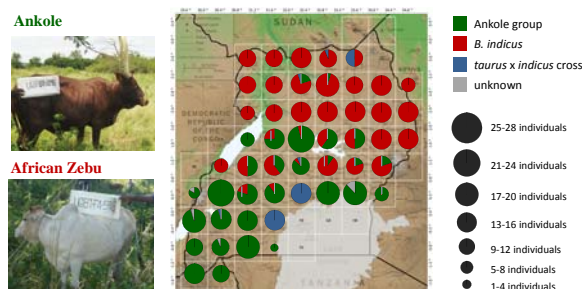


Figure 1: Sampling strategy of Ugandan cattle.

Results

The results obtained from Admixture analysis on 54K genotype data (786 individuals x 38597 markers) showed that most Ugandan individuals had a remarkable level of admixture (Figure 2, panel a), with the two major genomic components roughly corresponding to Ankole and Zebu breed groups (Figure 1).

When analysed in a spatial context, both genomic and disease prevalence data revealed an underlying geographical structure:

- i) despite the high level of admixture, the two major genomic components, corresponding to taurine Ankole and indicine cattle, predominate in southwestern and northeastern Uganda, respectively (Figure 2, panels b and c);
- ii) Brucellosis has a higher prevalence in the northeastern part of the country, while East Coast Fever prevails in central/southern Uganda.

The logistic regression on environmental variables showed that the areas with higher temperature stability throughout the year (e.g. southwestern Uganda) are those where the highest ECF prevalence is recorded, while the northeastern areas of the country, characterized by higher sunshine and lower relative humidity, are those where brucellosis prevalence is higher.

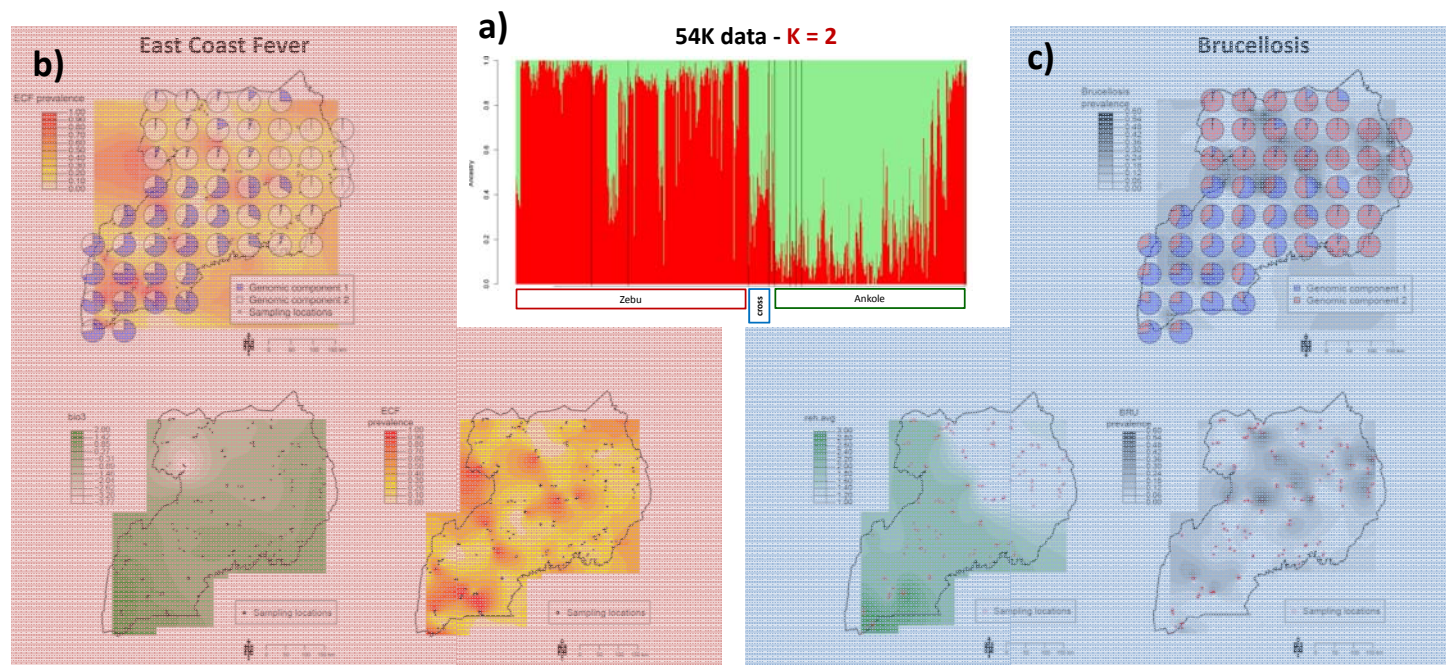


Figure 2: Panel a: Bayesian assignment test obtained with Admixture software at $K=2$ on the 54K genotypes of Ugandan cattle. Panel b, anticlockwise: Geographical representation of individual Q score values obtained with Admixture software averaged per sampling cell and superimposed to the map of East Coast Fever prevalence (Ankole/zebu ancestries are represented in blue/white shades within the piecharts). Geographical trend of the environmental variable “isothermality” (bio3). Map of East Coast Fever prevalence. Panel c, clockwise: Geographical representation of individual Q scores averaged per sampling cell and superimposed to the map of Brucellosis prevalence (Ankole/zebu ancestries are represented in blue/red shades within the piecharts). Geographical trend of the environmental variable “annual mean of relative humidity” (reh.avg). Map of Brucellosis prevalence.

Conclusions

The trends highlighted by our analyses likely reflect the combined effects of general adaptation to the environment (i.e. individuals carrying Ankole/indicine genomes better tolerate differential climatic frameworks) and of adaptation to selective pressure represented by the occurrence of specific diseases.

The significant association of the same environmental feature to both disease prevalence and genotypes, suggests that climatic variables may be used as predictors of the probability of occurrence of both disease and genotypes.

* The NextGen Consortium partners: Centre National de la Recherche Scientifique (CNRS), France; Cardiff University (UNICAR), UK; Università Cattolica del Sacro Cuore Piacenza (UNICATT), Italy; Parco Tecnologico Padano (PTP), Italy; European Molecular Biology Laboratory (EMBL-EBI), UK; Ecole Polytechnique Fédérale de Lausanne (EPFL), France; Università Teramo (UNITE), Italy; Makerere University (MAK), Uganda; Institut National de la Recherche Agronomique (INRA-Mor), Morocco; Gorgan University of Agriculture and Natural Resources (GAU), Iran; Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia.