

# Spatial areas of genotype probability of cattle genomic variants involved in the resistance to East Coast Fever: a tool to predict future disease-vulnerable geographical regions

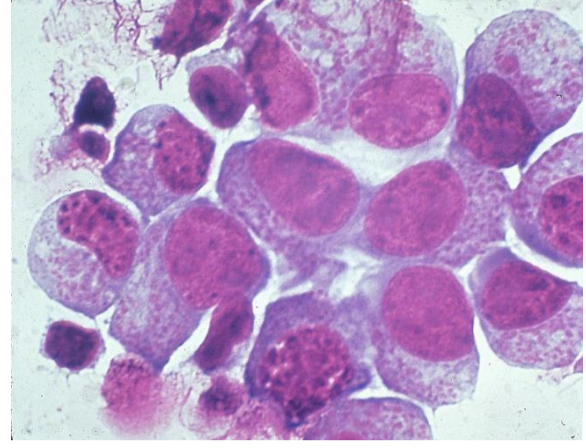
Elia Vajana<sup>1\*</sup>, Estelle Rochat<sup>2\*</sup>, Licia Colli<sup>1</sup>, Charles Mase<sup>3</sup>, Riccardo Negrini<sup>1</sup>, Paolo Ajmone-Marsan<sup>1</sup>, Stéphane Joost<sup>2</sup> and the NEXTGEN Consortium

<sup>1</sup> Institute of Zootechnics and BioDNA Research Centre, Faculty of Agricultural, Food and Environmental Sciences, Università Cattolica del S. Cuore, Piacenza, Italy, <sup>2</sup> Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, <sup>3</sup> Institute of Environment & Natural Resources, Makerere University, Kampala, Uganda

\* These authors contributed equally to this work

## 1. Introduction

- East Coast Fever (ECF) is a major livestock disease caused by the hemo-parasite protozoan *Theileria parva* Theiler, 1904
- It causes high mortality in cattle populations of East and Central Africa, especially in exotic breeds and crossbreeds (Olwoch *et al.*, 2008)
- Main *T. parva* vector is the hard-bodied tick *Rhipicephalus appendiculatus* Neumann, 1901



## 2. Objectives

- Highlight cattle genotypes associated with tick occurrence, and thus likely involved into tolerance/resistance mechanisms against ECF
- Build their **Spatial area of genotype probability (SPAG)** to delimit areas where the concerned genotypes are predicted to be present

## 3. Data

- 803 Ugandan cattle genotyped (54K SNPs) during the NEXTGEN project (2010-2012)
- 532 *R. appendiculatus* occurrences retrieved from a published database (Cumming, 1998)
- 19 bioclimatic variables derived from the *WorldClim* database, both for current conditions and for moderate/severe climate change scenarios for 2070

## 4. Methods

1. Selection of poorly collinear ( $r < 0.7$ ,  $VIF < 5$ ) and ecologically meaningful bioclimatic variables

- 6 bioclimatic variables selected

2. Estimation of *R. appendiculatus* ecological niche/geographical distribution: proxy of the **current** parasite selective pressure

- Probability of presence of the tick estimated using MAXENT models (Muscarella *et al.*, 2014)
- Model selection based on Akaike's Information Criterion (AIC)

3. Detection of genotypes likely involved into host tolerance/resistance

- Correlative approach implemented in *Sambada* (Stucki *et al.*, 2014): detection of genotypes positively associated with the probability of occurrence of the tick

4. SPAGs estimation

- Calculate and map a generalized probability of presence of the concerned genotypes using the regression coefficients as estimated in step 3

6. Comparison between **current SPAGs** and **future** areas of parasite potential presence

- Identification of areas where the genotypes of interest are not present yet but where the parasite is likely to expand its range in the future

5. Estimation of the shift in *R. appendiculatus* niche/geographical distribution as a consequence of climate change: proxy of the **future** parasite selective pressure

- Bioclimatic data corresponding to two climate change scenarios for 2070: one moderate (RCPs 2.6), one severe (RCPs 8.5) (global climate model *GISS-E2-R*)

## 5. Results

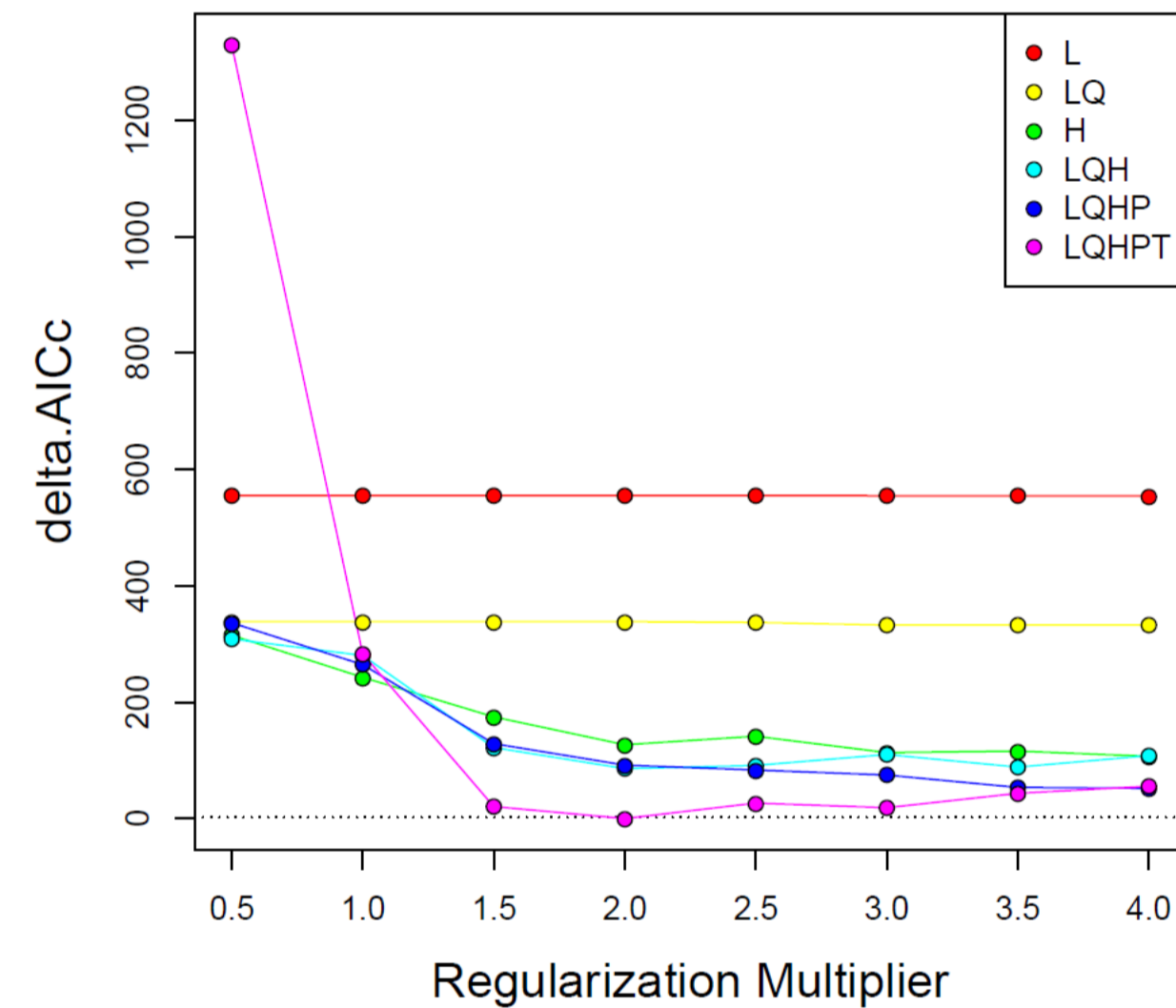
### Ecological niche modelling

#### Current climatic conditions

- Best environment for *R. appendiculatus* South and North - East of Lake Victoria

#### Climatic conditions in 2070

- Contraction of suitable environments for the tick (both with moderate and severe climate change scenarios)
- Expected shift in the tick niche/distribution southwards in the eastern regions of Lake Victoria

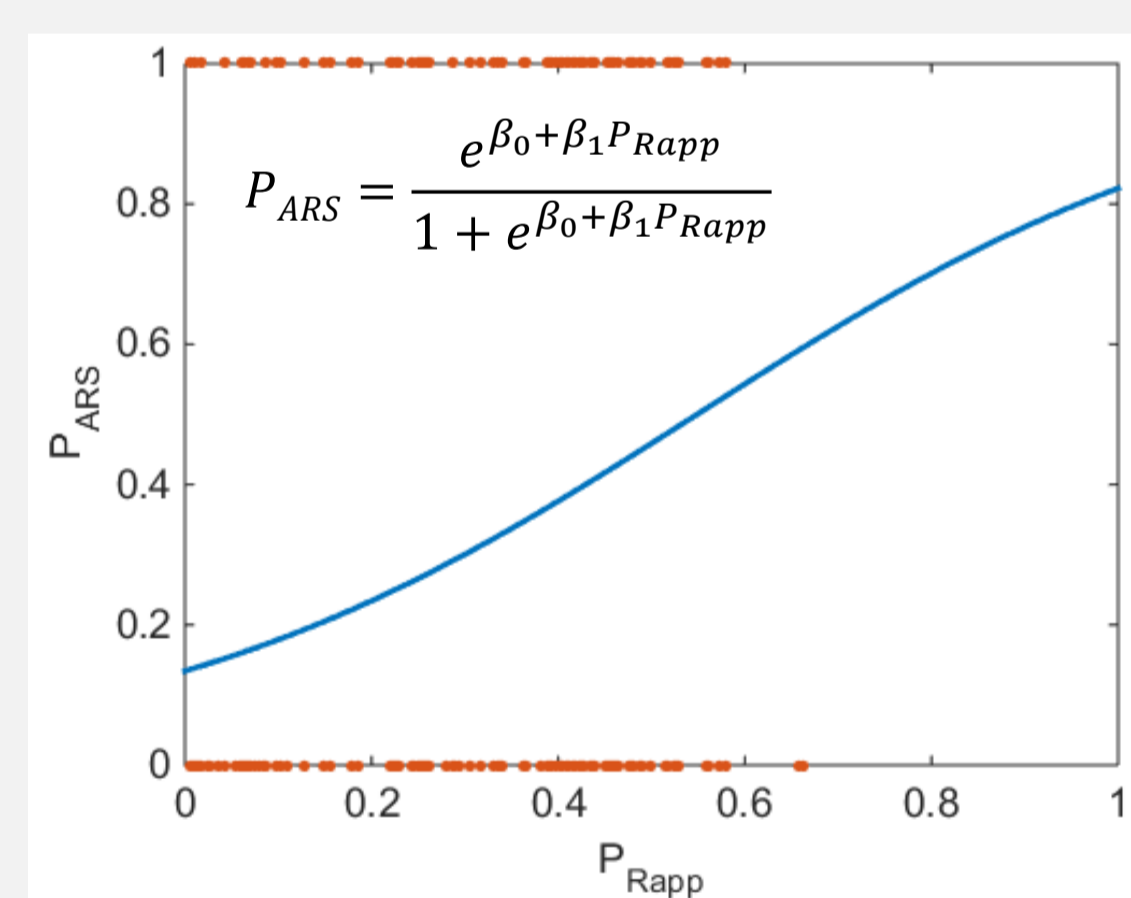


**Fig. 1** - Relative performance of the MAXENT models used to estimate the ecological niche of the tick.

- Regularization Multiplier (RM): the higher the value, the smaller the expected overfitting
- Possible kinds of transformation applicable to the predicting variables: L = Linear, Q = Quadratic, H = Hinge, P = Product, T = Threshold
- Evaluation of the single performance: difference between single AIC value and the AIC value of the best model overall (*i.e.*, best model:  $\Delta AIC = 0$ ). In this case, the best model is obtained by allowing all the transformations (LQHPT) and by setting  $RM = 2$

### Landscape genomics

- 103 genotypes showing a significant ( $\alpha = 0.01$ ) positive association with a high probability of presence of the tick
- Genotype "AG" of the marker "ARS-BFGL-NGS-113888" ("ARS-11") shows a significant positive association and is close to the gene IRAK-M, an essential component of the Toll-like receptors involved in the immune response against pathogens (Kobayashi *et al.*, 2002)



**Fig. 2** - Relationship between the probability of presence of the genotype "AG" of the marker "ARS-11" ( $P_{ARS}$ ) and the current *R. appendiculatus* probability of occurrence ( $P_{Rapp}$ ), as estimated by a logistic regression model

- $\beta_0 = -1.86$  and  $\beta_1 = 3.39$
- Significativity threshold  $\alpha = 0.01$  (Bonferroni correction included)
- Efron pseudo  $R^2 = 0.074$

**Fig. 3** - SPAG and ECF vector niche

- × NEXTGEN sampling farms
- × Presences of *R. appendiculatus*
- Probability of presence of *R. appendiculatus*
- 0.1
- 0.3
- 0.5
- 0.6
- 0.7
- 0.8
- iso-probability 0.7
- iso-probability 0.6

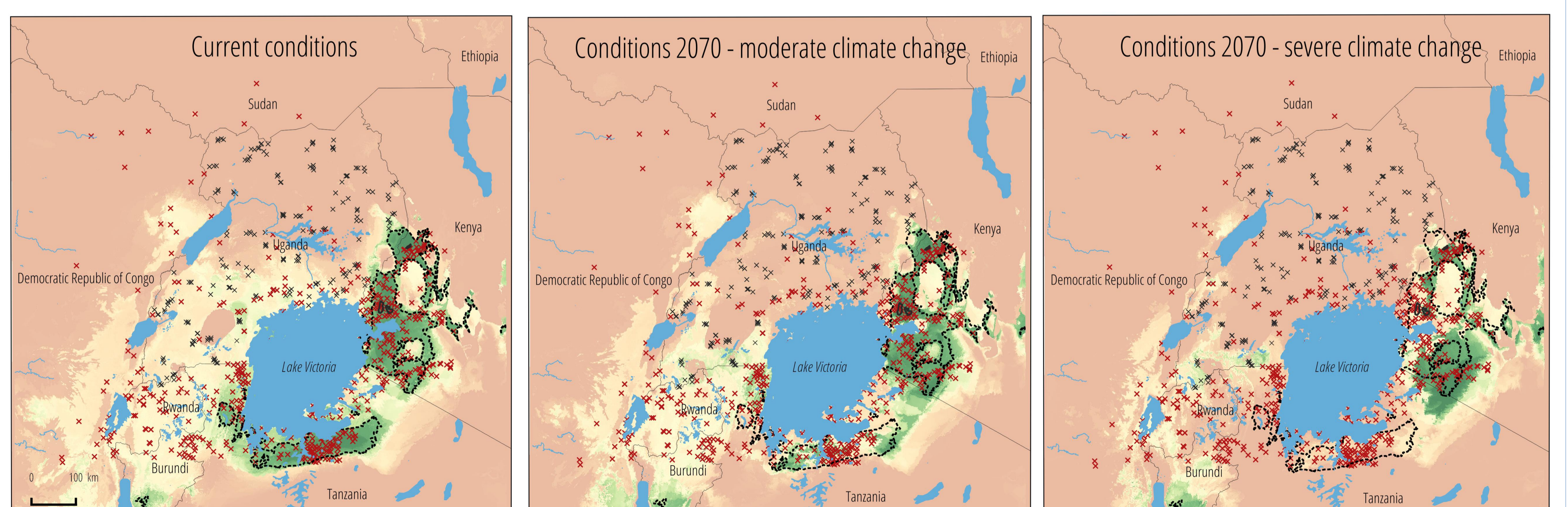
### Spatial area of genotype probability for the genotype "AG" ("ARS-11" marker)

#### Current conditions

- Two main areas of genotype presence, one South and one North - East of Lake Victoria, corresponding to the regions of high ECF vector occurrence

#### Conditions 2070

- New potential disease-vulnerable areas East of Lake Victoria caused by the shift in vector niche as a consequence of climate change



## 6. Conclusion

- Combined use of SPAGs and niche maps could help identifying critical geographical regions where the favourable genetic variant is currently not present, but where a parasite is likely to expand its range in the future
- Valuable tool to support the identification of current resistant populations and to direct future targeted disease prevention, control strategies and breeding schemes

### References

- Olwoch J.M., Reyers B., Engelbrecht F.A., & Erasmus B.F.N. (2008) Climate change and the tick-borne disease, Theileriosis (East Coast fever) in sub-Saharan Africa. *Journal of Arid Environments*, **72**, 108–120
- Cumming G.S. (1998) Host preference in African ticks (Acari: Ixodida): a quantitative data set. *Bulletin of Entomological Research*, **88**, 379.
- Muscarella R., Galante P.J., Soley-Guardia M., Boria R.A., Kass J.M., Uriarte M., & Anderson R.P. (2014) ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods in Ecology and Evolution*, **5**, 1198–1205.
- Stucki S., Orozco-terWengel P., Bruford MW *et al.* (2014) High performance computation of landscape genomic models integrating local indices of spatial association. arXiv:1405.7658
- Kobayashi K., Hernandez L.D., Galán J.E., Janeway C.A., Medzhitov R., & Flavell R.A. (2002) IRAK-M is a negative regulator of Toll-like receptor signaling. *Cell*, **110**, 191–202.