



Genomic diversity and Population Structure of Ugandan Taurine and Zebuine Cattle Breeds

Licia Colli¹, Marco Milanesi¹, Lorenzo Bomba¹, Elia Vajana¹, Riccardo Negrini², 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)Inst. of Zootechnics, Università Cattolica del S. Cuore, Piacenza, Italy, (2)AIA - Italian Breeders Association, Rome, Italy, (3)FPPT, Lodi, Italy, (4)EPFL, Lausanne, Switzerland, (5)LECA, Université de Grenoble 1, Grenoble, France, (6)Makerere University, Kampala, Uganda, (7)EU funded project, <http://nextgen.epfl.ch>



Introduction

An extensive sampling of Ugandan cattle was carried out in the course of the European project Nextgen to identify possible associations between genotypes, livestock endemic diseases and environmental variables.

Aim of the research: As a prior to the GWAS and selection signatures analyses planned within the project, we analyzed the population structure of Ugandan cattle genotyped with both 54K and 800K HD SNP panels in the context of the worldwide cattle genomic diversity.

Materials & Methods

A total of 917 animals, representative of the Ankole group (crossbred between *Bos indicus* and *Bos taurus*), Zebu and Ankole-Zebu crosses, were sampled over the whole country and genotyped (Figure 1): 815 with Illumina BovineSNP50 BeadChip (ca. 54K SNPs) and 102 with BovineHD Genotyping BeadChip (ca. 800K SNPs). Prior to association studies, both population structure and possible introgression from non-native gene pools (e.g. industrial breeds) have been extensively investigated by comparing the Ugandan dataset to publicly available 54K SNP data from previous research projects or public databases. The genotype data were filtered with the following thresholds: genotype call rate (SNPs) ≤ 0.95 , genotype call rate (Animals) ≤ 0.95 , MAF ≤ 0.01 for the 54K data and MAF ≤ 0.05 for the 800K data. After quality control and merging procedures, the final working set included Ugandan cattle together with 108 reference cattle breeds/species (taurine, indicine, crosses and outgroups) sampled from all over the world: 30 from Africa, 28 from Asia, 40 from Europe and 9 from the Americas (Figure 2), *Bos javanicus* and *Bos gaurus* (outgroup). The different approaches adopted for data analysis included Bayesian clustering procedures (Admixture software; Novembre *et al.* 2010) and multivariate statistics (Multi-Dimensional Scaling plot).

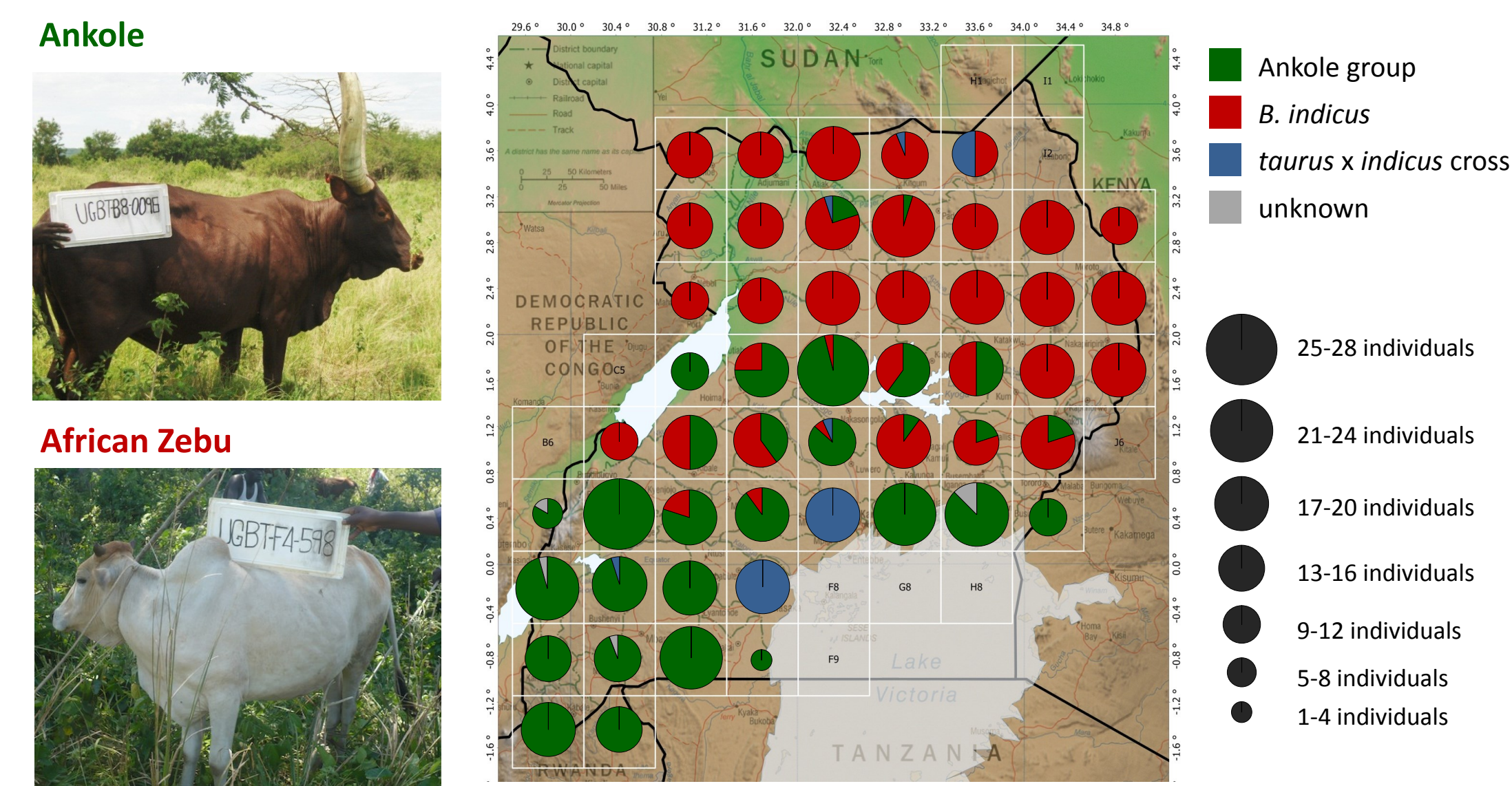


Figure 1: Grid-based sampling strategy of Ugandan cattle.

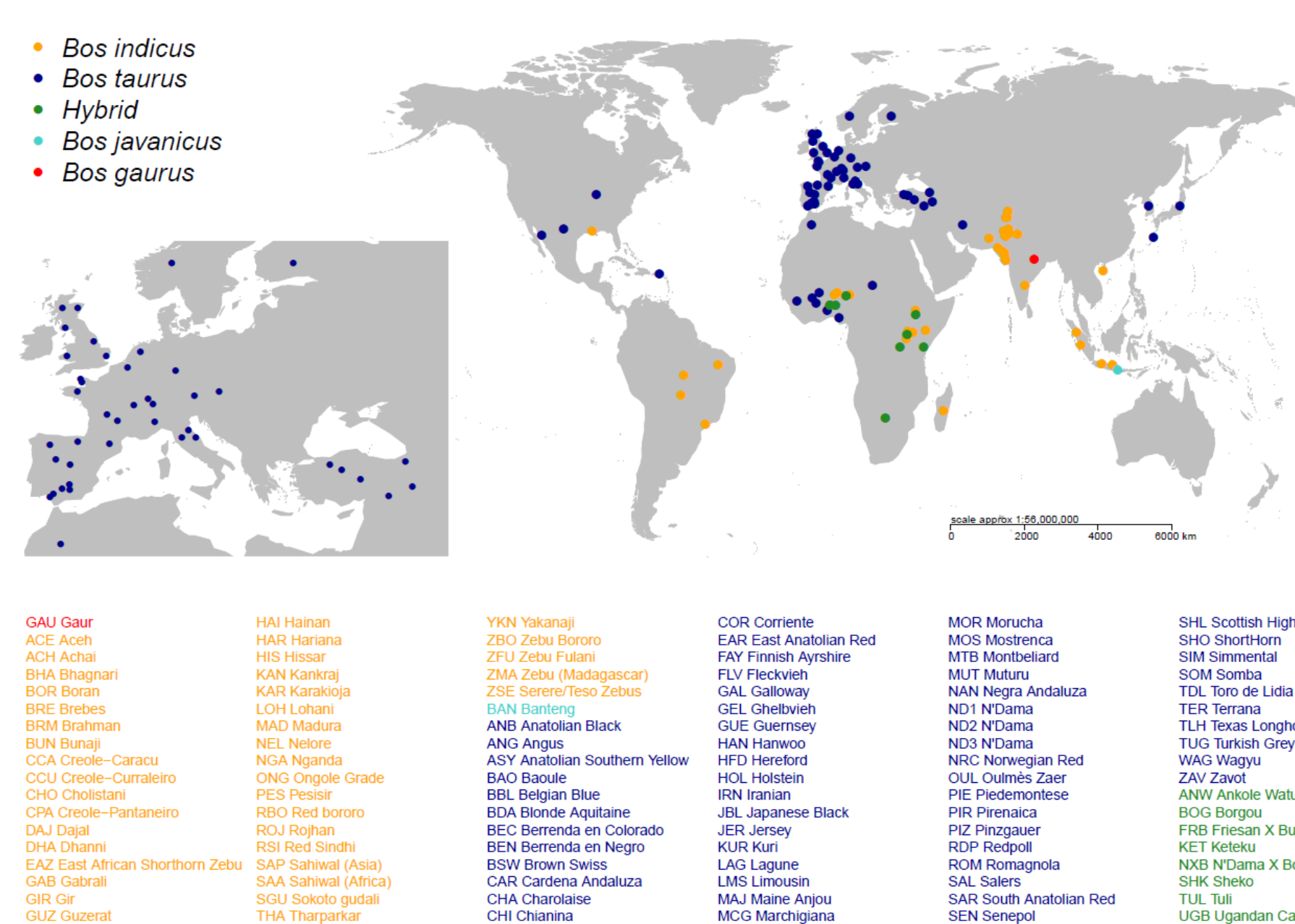


Figure 2: Worldwide distribution of the reference cattle breeds and outgroup species included in the dataset.

Results

The results obtained with Admixture software showed that Ugandan cattle are characterized by two major genomic components roughly corresponding to Ankole and zebu ancestries (Figure 3), and which respectively cluster with Ankole animals sampled outside Nextgen project and Boran, Serere/Teso and Sahiwal indicine breeds from Africa and Asia (Figure 3). Also some minor genomic components were revealed, one of which is partly shared with Boran zebus and Nganda mixed breed (i.e. light orange and fuchsia components in Figure 3). Conversely, the origin of others (i.e. deep purple component in Figure 3) could not be assigned to any of other reference breeds.

The MDS plot (Figure 4) highlighted a clear subdivision of three groups of individuals, namely European taurine cattle, African taurine cattle and Asian zebu. Turkish breeds, the closest to the center of domestication, are positioned near the origin of the axes, while Ugandan cattle are scattered on a wide surface overlapping to Ankole, African taurine zebu crosses and African zebu. Some Ugandan individuals, also, separate from the majority of the samples and spread towards the center of the graph.

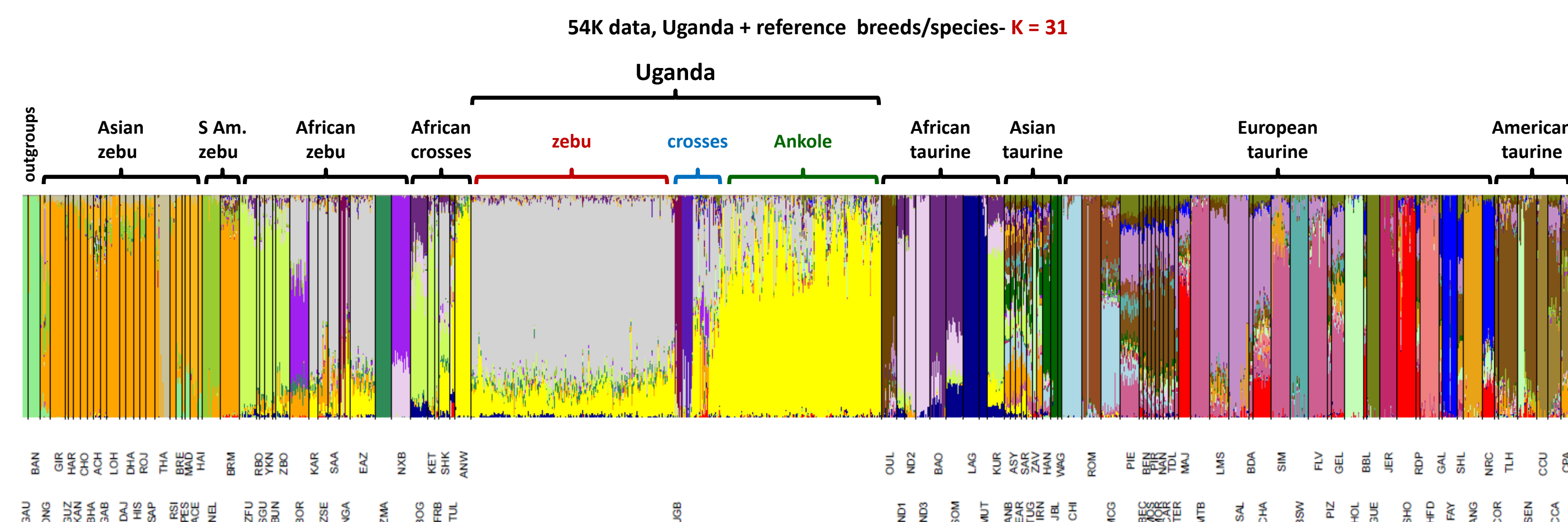


Figure 3: Bayesian assignment test (Admixture software). The picture shows the results of the assignment test at K=31, the number of hypothetical populations identified as the best-fitting through the calculation of Cross-Validation errors. For the correspondence between breed labels and names, see Figure 2.

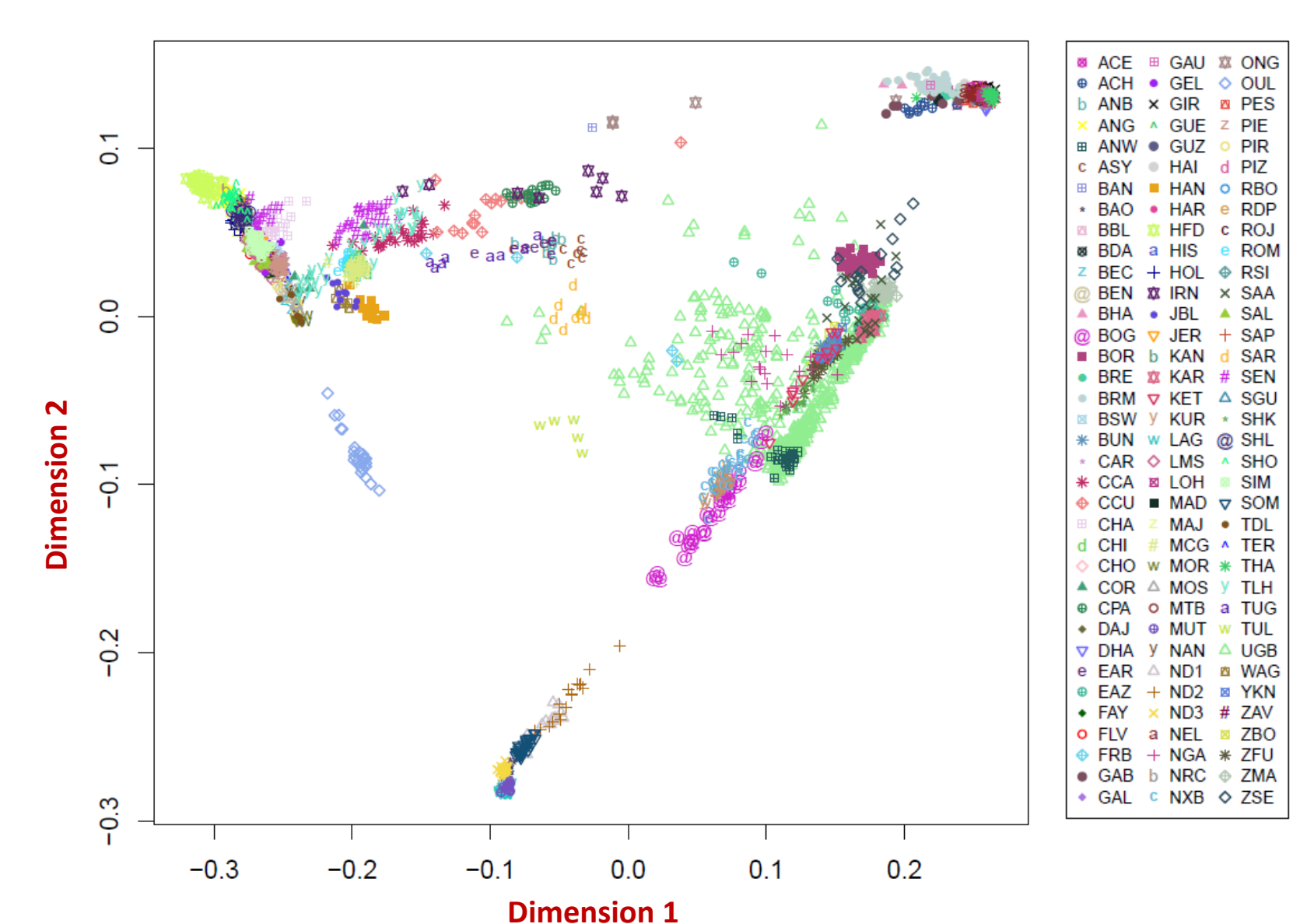


Figure 4: Multi-Dimensional scaling plot. Dimension 1, x axis: 27.82% of explained variance. Dimension 2, y axis: 6.01% of explained variance. For the correspondence between breed labels and names, see Figure 2.

Conclusions

Ugandan cattle genomes are characterized by high levels of admixture between taurine/indicine genomic components, most of which according to the present analyses seem to be of African-native origin. Furthermore, the presence of some minor components that could not be assigned to any of the African and European *B. taurus* and *B. indicus* rises some interesting questions about i) a possible introgression from European cattle populations not included in the present set, or ii) a contribution from other and still unidentified African taurine or indicine breeds. We are currently investigating these hypotheses through the analysis of Whole Genome Sequences of a selected set of Ugandan cattle individuals (see communication C-119 by Bomba *et al.*), which will hopefully also allow to understand if the position of some Ugandan cattle close to the Turkish breeds on the MDS plot mirrors an actual genomic similarity or is caused by introgression from the European taurine gene pool.