Whole genome sequencing reveals signals of adaptive admixture in Creole cattle

Slim Ben Jemaa[1], Gabrièle Adam[1], Mekki Boussaha[2], Christophe Klopp[3], Nathalie Mandonnet[4], and Michel Naves[4]

[1]INRAE Antilles-Guyane
[2]INRAE Centre Île-de-France Jouy-en-Josas Antony
[3]Plateforme BioInfo Genotoul, INRAE
[4]INRA Antilles Guyane

January 19, 2023

Abstract

Admixture is an evolutionary process that enables short-term adaptation. The Creole cattle from Guadeloupe is a tropically adapted breed. Its three-way admixture and long-term isolation offer a unique opportunity for understanding the genetic determinants of adaptive admixture in livestock. Here, we sequenced 23 Creole cattle from Guadeloupe (GUA) and combined our data with sequenced genomes of 99 cattle from 25 breeds representative of European, African and indicine groups to provide the most detailed exploration, to date, of patterns of genetic variation and to detect selection signatures in this population. We detect 17 228 983 single nucleotide polymorphisms (SNPs) and we confirmed the higher level of African and indicine ancestries, compared to the European ancestry, in the GUA genome. We show that, unlike Criollo cattle, GUA population originates directly from West Africa with indicine ancestry inherited via West African ancestors. We relied on consistency of signals across various methods based on excess of haplotype homozygosity, differences in allele frequencies and excess / deficiency of local ancestry to identify five strong candidate regions showing an excess of indicine ancestry. These encompass immune-, heat-tolerance- and physical exercise-related genes. Moreover, we found that a previously identified horn-related gene, RXFP2 is under strong selective pressure in GUA genome likely owing to human-driven (socio-cultural) pressure. Our study highlights the role played by population admixture for driving rapid adaptive response to local environmental constraints.

Whole genome sequencing reveals signals of adaptive admixture in Creole cattle

Running title : adaptive admixture in Creole cattle

Slim Ben-Jemaa[1,2*], Gabriele Adam[1], Mekki Boussaha[3], Christophe Klopp[4], Nathalie Mandonnet[1] and Michel Naves[1]

[1] INRAE, ASSET, 97170, Petit-Bourg, France.

[2] Laboratoire des Productions Animales et Fourragères, Institut National de la Recherche Agronomique de Tunisie, Université de Carthage, 2049 Ariana, Tunisia.

[3] Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350, Jouy-en-Josas, France.

[4] Plateforme Bio-Informatique GenoToul, MIAT, INRAE, UR875 Mathématiques et Informatique Appliquées Toulouse, 31326 Castanet-Tolosan, France.

[*] Corresponding author : INRAE, ASSET, 97170, Petit-Bourg, France. Email : slim.ben-jemaa@inrae.fr

Abstract

Admixture is an evolutionary process that enables short-term adaptation. The Creole cattle from Guadeloupe is a tropically adapted breed. Its three-way admixture and long-term isolation offer a unique opportunity for understanding the genetic determinants of adaptive admixture in livestock. Here, we sequenced 23 Creole cattle from Guadeloupe (GUA) and combined our data with sequenced genomes of 99 cattle from 25 breeds representative of European, African and indicine groups to provide the most detailed exploration, to date, of patterns of genetic variation and to detect selection signatures in this population. We detect 17 228 983 single nucleotide polymorphisms (SNPs) and we confirmed the higher level of African and indicine ancestries, compared to the European ancestry, in the GUA genome. We show that, unlike Criollo cattle, GUA population originates directly from West Africa with indicine ancestry inherited via West African ancestors. We relied on consistency of signals across various methods based on excess of haplotype homozygosity, differences in allele frequencies and excess / deficiency of local ancestry to identify five strong candidate regions showing an excess of indicine ancestry. These encompass immune-, heat-tolerance- and physical exercise-related genes. Moreover, we found that a previously identified horn-related gene, *RXFP2* is under strong selective pressure in GUA genome likely owing to human-driven (socio-cultural) pressure. Our study highlights the role played by population admixture for driving rapid adaptive response to local environmental constraints.

Keywords

Admixture, adaptation, Creole cattle, whole genome sequence, selection signature

Introduction

One of the major consequences of long-distance human migrations is the displacement of domestic animals into new environments, thus putting a strong selective pressure on the genome of these animals over a brief period of time. One example is the introduction of cattle to the western hemisphere. New World Creole cattle were first brought from the Iberian peninsula by Spanish colonists since the second expedition of Christopher Columbus in the late 15th century. Creole cattle breeds have then undergone a rapid expansion throughout the American continent (Rodero Serrano, Rodero Franganillo, & Delgado-Bermejo, 1992). Subsequently, between the 16th and 18th centuries, West African cattle are thought to have entered the Caribbean and Brazil, presumably as a consequence of slave trade routes (Primo 1992; Maillard, Palin, Trap, & Bensaid, 1993). Genetic evidence also points to a West African influence on Creole Cattle (Magee et al., 2002 ; Ginja et al., 2019). Around the middle of the 19th century, several other European cattle breeds were brought in large numbers to the Americas. Later on, during the beginning of the 20th century, *Bos indicus* were imported from India to improve the adaptability of local populations in tropical areas of the Americas through extensive crossbreeding (Maillard, Palin, Trap, & Bensaid, 1993; Pitt et al., 2019).

Creole cattle in the Guadeloupe island (GUA) is an admixed breed. Published estimates of ancestry proportions of this population indicate 26%, 36% and 38% of European taurine (EUT), African taurine (AFT) and indicine (IND) ancestries in the GUA genome, respectively (Gautier & Naves 2011). GUA population is well adapted to tropical environment. This is exemplified by resistance to high temperatures as well as to local parasitic and infectious diseases (Gauthier et al., 1984). GUA individuals are also able to endure undernourishment during dry period and to valorize to the poor-quality feed (Naves 2003).

Elucidating the genetic architecture of adaptation is becoming an increasingly relevant topic in animal genetics. This is because it is expected that innovative breeding schemes will make use of heritable resilience biomarkers to overcome the foreseeable adverse impacts of climate change characterized by increasing temperatures, the expansion into new areas of invasive parasites and possibly the degradation of forage production and quality. An increasingly number of studies have addressed the effect of admixture and introgression in cattle adaptation to new environmental challenges (Chen et al., 2018; Barbato et al., 2020; Kim et al., 2020). These studies benefited from ongoing advancements in genomic technology and the development of improved statistical and computational methods to identify signatures of selection in cattle genome, that is genomic regions that appear to be shaped by selection.

Owing to its three-way admixture and its long-term isolation, the genome of Creole cattle from Guadeloupe

offers a unique opportunity to study admixture-enabled adaptation.

There are only a few studies that aimed at detecting selection signatures in Creole cattle (e.g. Pitt et al., 2019; Toro-Ospina et al., 2022). A previous study provided a first insight into footprints of selection in GUA cattle using the the Illumina BovineSNP50 chip assay (Gautier & Naves 2011). To expand our understanding of the genomic architecture of local adaptation in this population, we leverage whole-genome sequence data from 23 GUA animals and genomes of 95 cattle individuals from various origins. Our objective was to use high-resolution genomic data and consistency of signals among different methods based on excess of haplotype homozygosity, differences in allele frequencies and excess / deficiency of local ancestry to identify new, strong candidate regions under selection in the GUA genome.

Materials and Methods

Samples selection and genome sequencing

Twenty three Creole bulls representative of the INRA nucleus in Guadeloupe were selected for the purpose of this study. Most of these animals have limited relationships (based on pedigree information) and are among the most frequently used in artificial insemination. Genomic DNA was extracted from whole-blood and semen samples collected between 1995 and 2015. Four Colombian individuals belonging to the IMAGE (« Innovative Management of Animal Genetic Ressources ») project were also included in the study. Paired-end libraries with insert size of 500 bp were constructed for each individual and sequenced using the HiSeq 3000 platform (Illumina) in the Genome et Transcriptome (GeT) GenoToul platform (Toulouse, France), following the manufacturer's protocol.

Sequence alignment and genotype calling

Genotype data were generated following the 1000 Bull Genomes Project Run 8 guideline (*http://www.1000bullgenomes.com/*). Quality control of raw sequence reads was performed using the fastQC software v.0.11.7 (*http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/*). Trimmomatic-0.36 (Bolger, Lohse, & Usadel, 2014) was used to remove Illumina adapter sequences, low-quality bases and artefact sequences. Filtered sequences were then mapped against the bovine reference genome (ARS-UCD1.2) using the Burrows-Wheeler Alignment tool (bwa mem v.0.7.17) (Li et al., 2009) with default parameters. The resulting SAM files were then converted to BAM format, sorted, and indexed using SAMtools (Li et al., 2009). Potential PCR duplicates were removed using the MarkDuplicates tool from Picard version 1.88 (http://broadinstitute.github.io/picard). Only properly reads with a mapping quality of at least 30 were kept. Prior to variant discovery, local realignment was performed using two GATK (Genome Analysis Toolkit) version 4.0.0.0 modules, RealignerTargetCreator and IndelRealigner. GVCF files were then created for each sample using the GATK HaplotypeCaller module. Single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) were subsequently called for all samples using the GATK Genotype-GVCFs tool. VCFtools v.0.1.17 software was used to filter out indels, multiallelic SNPs and variants that i) displayed a minor allele frequency (MAF) < 0.1; ii) have missing genotype rates > 0.1; iii) did not pass Hardy–Weinberg equilibrium (HWE) test (p<0.01) and iv) have a quality score (–minQ option) below 100. The selected variants were subsequently annotated using the SnpEff annotation software (Cingolani et al. 2012).

Data merging and relatedness check

We downloaded publicly available genomes of of 41 individuals including 12 Africain indicine (AFI), 10 Africain taurine (AFT), 7 Australian-American indicine (AMI), 4 Asian indicine (ASI), 4 European taurine breeds from Spain (IBER) and 4 Limonero (LIMO) individuals (representative of Criollo (CRIO) cattle) (Table S1). We followed the same procedure aforementioned in sequence alignment and variant calling. Finally, we included in the data set, genotyping data belonging to 50 individuals from six French breeds provided by the 1000 Bull Genomes Project (Table S1). We applied the same marker selection thresholds (maf = 0.1, missing genotype rate= 0.1, P-value HWE = 0.01 and minQ=100) to build these two databases which resulted in 16,360,962 and 8,955,346 high-quality SNPs for the public and the 1000 Bull Genomes

genotyping data, respectively. All three databases were combined into a single one containing 5,316,956 common snps. Nucleotide diversity (Π) at a whole-genome scale was estimated in sliding windows of 1 Mb using VCFtools considering the following breed groupings : AFI (N=11) and AFT (N=7) and the each of the following breeds separately : AUB (N=8), CHA (N=7), HOL (N=6), LIM (N=7), MON (N=6). In Π computation, we selected 10 GUA individuals among the least related. This was done to have sample groups with similar size (for a matter of comparison).

We further performed an LD-based SNP pruning using PLINK with the "-indep-pairwise 20 4 0.6" option which resulted in 1,064,342 SNPs spread over all autosomal chromosomes used in population structure analyses. Average inter-marked distance was 2.3 Kb ± 4.3 Kb (Table S2). To overcome the effect of closely related animals, we performed a relatedness test between individuals within each population using PLINK (Purcell et al., 2007). The software calculates a variable called PI-HAT reflecting extended haplotypes shared between distantly related individuals. We excluded one individual from each pair of individuals with a PI-HAT value > 0.05 which is a value roughly corresponding to 3rd-degree relationships. In total, after relatedness filtering, 99 individuals including 19 GUA animals, were available for the different analyses (Table S1).

Population structure and genetic relationship analyses

We performed a principle component analysis (PCA) using the adegenet R package (Jombart, 2008). Furthermore, genetic structure was inferred from SNP data in ADMIXTURE 1.3 software (Alexander, Novembre, & Lange, 2009). We ran ADMIXTURE for values of K from 3 through 5. DISTRUCT software (Rosenberg, 2004) was used to graphically display ancestry within each individual. Global averages of pairwise population differentiation measured by fixation index (Fst) was estimated using Genepop 4.6 software (Rousset, 2008) for populations represented by more than one animal. The four Spanish individuals were considered as one population as well as the four Colombian animals. Next, we performed local ancestry assignment across GUA genome using the Efficient Local Ancestry Inference (ELAI) algorithm (Guan, 2014) under a three-way admixture model using American-Australian and Asian Indicine (AMASI) and AFI populations as proxies for indicine ancestry, Muturu and Ndama as proxies for AFT ancestry, the French and the Spanish breeds as proxies for EUT ancestry. ELAI uses a two-layer hidden Markov model to detect the structure of haplotypes for unrelated individuals. The algorithm models two scales of linkage disequilibrium (one within a group of haplotypes and one between groups) and provides a map, for each admixed individual, showing the probability, for each SNP, to descend from each one of the ancesral populations. ELAI analysis was conducted across 19 GUA genomes by setting the parameters -mg (number of generations) to 70, -s (EM steps) to 30, -C (upper clusters) to 3, and -c (lower clusters) to 15.

The patterns of population splits and mixtures were inferred using TreeMix (Pickrell & Pritchard, 2012). To run TreeMix, we considered the 17 populations with more than two individuals. We built a maximum likelihood tree using blocks of 30000 SNPs with no migration events allowed. Then, we built a phylogenetic tree of these populations and started adding migration events (modeled as edges) sequentially to the phylogenetic model. The migration edges were added until 99.79% of the variance in ancestry between populations was explained by the model.

Identification of runs of homozygosity

Runs of homozygosity (ROHs) were identified in sliding windows of 100 SNPs using PLINK and the 1,064,342 SNPs that passed quality control filtering. ROH were detected within the following five breed groupings : Creole cattle from Guadeloupe (GUA), African Taurine (AFT) : Muturu and Ndama, Africain Indicine (AFI) : BOR, BUT and BARK, European Taurine (EUT) : IBER (represented by the four Spanish breeds), AUB, CHA, BAQ, LIM, MON, HOL, LIMO, COL (represented by the four Colombian breeds) and American-Australian and Asian Indicine (AMASI) : BRA, GIR, NEL, HAR and SAHW. The following parameters were used to define a ROH : (i) homozyg-snp (minimum number of SNPs that a ROH is required to have) : 100. (ii) homozyg-density (required minimum density to consider a ROH) : 50. (iii) homozyg-gap (length in Kb between two SNPs in order to be considered in two different segments) : 1000 (iv) homozyg-window-het

(number of heterozygous SNP allowed in a window) : 3 (v) homozyg-window-missing (number of missing calls allowed in a window) : 5. The –homozyg-group option implemented in PLINK was used to assess ROH islands shared among GUA individuals. These were defined as the homozygous segments shared by at least 30% of the samples.

Identification of selection signatures

Prior to selection signature analysis, we performed a more stringent LD-based pruning with the '-indep-pairwise 20 4 0.4' option of PLINK which resulted in 433,408 SNPs. Integrated haplotype score (iHS) (Voight, Kudaravalli, Wen, & Pritchard, 2006), Rsb (Tang, Thornton, & Stoneking, 2007) and cross population extended haplotype homozygosity ($XP\text{-}EHH$) (Sabeti et al., 2007) scans were performed using the $rehh$ package (Gautier and Vitalis, 2012). In iHS computation, the information on the ancestral and derived allele state is needed for each SNP because this statistic is based on the ratio of the extended haplotype homozyosity (EHH) associated to each allele. In our analysis, the ancestral allele was inferred as the most common allele within our dataset. iHS scores for each SNP were transformed into two-sided p-values : $piHS$ =-log10[1-$2|\Phi(iHS)$-0.5|]. Rsb and $XP\text{-}EHH$ analyses were performed for each of the three pairwise comparisons: GUA Vs AFT, GUA Vs EUT (French, Spanish and Criollo breeds) and GUA Vs IND (AMASI and AFI populations). Haplotype phasing was performed using fastPHASE 1.4 (Scheet & Stephens, 2006). Since fastPHASE is based on haplotype clusters, whose size should be set a priori, we used the toolkit implemented in imputeqc R package (Khvorykh & Khrunin, 2020) to estimate the optimal number of haplotype clusters (K) needed for haplotype phasing. Imputeqc package has been designed to assess the imputation quality and/or to choose the model parameters for imputation. In the present study, we found that K = 20 provided the best imputation quality (for 5% of masked data). Therefore, we used this value to run fastPHASE. Considering that Rsb and $XP\text{-}EHH$ values are normally distributed, a Z-test was applied to identify significant SNPs under selection. Two-sided p -values were derived as $pRsb$ =-log10[1-$2|\Phi(Rsb)$-0.5|] and $pXP\text{-}EHH$ =-log10[1-$2|\Phi(XP\text{-}EHH)$-0.5|] where Φ (x) represents the Gaussian cumulative distribution function. For all three EHH-based tests, the maximum allowed gap between two SNPs was set to 500 Kb. We used sliding nonoverlapping 500-Kb windows to perform selection signature detection. A window is classified as putatively under selection when it contains at least 5 markers exceeding the significance threshold of -log10 (p-value) = 6. Positive $XP\text{-}EHH$ and Rsb values indicate longer haplotypes in the target population (i.e GUA) therefore suggesting that selection occurred in GUA population.

Confirming the relevant candidate regions putatively under selection in GUA

Regions putatively under selection in GUA genome, identified by at least two EHH-based tests and experiencing a sudden increase in one of the three ancestries of GUA compared to the average level on their respective chromosomes were considered as relevant. We employed additional approaches to further confirm these relevant candidate regions. First, Fst and nucleotide diversity were calculated for nonoverlapping 50-Kb windows across the genome using VCFtools. Highly differentiated windows with at least 4 SNPs between GUA and each of its three ancestries were identified. We checked if our candidate regions were among the top 1% windows and if these regions present specific patterns of nucleotide diversity compared to the neighboring regions. Next, we used SweeD v4.0.0 software (Pavlidis, Živkovic, Stamatakis, & Alachiotis, 2013) to calculate the composite likelihood ratio (CLR) in nonoverlapping 500-kb windows along the chromosomes containing the relevant regions. The software detects Site Frequency Spectrum (SFS) patterns generated by complete selective sweeps.

We collected information through a literature search to discuss the biological implications of our findings with regard to the candidate genes located in the relevant genomic regions known to be involved in phenotypic variation of adaptive traits.

Results

Sequencing and detection of variation

A total of 7,543,644,154 reads were generated after sequencing the complete genome of the 23 Creole cattle

samples. The reads were aligned to the latest Bos taurus reference genome (ARS-UCD1.2) with an average alignment rate of 95,7 % (ranging between 94.93 % and 96.51%) and an average depth of 16.35 (min depth = 9.3 ; max depth = 23.77) (Table S3). We identified a total of 17,228,983 filtered SNPs in the 23 GUA individuals (Table S4). Functional annotation revealed that the vast majority of SNPs are located within intronic (48%) and intergenic regions (41,5%) while exons accounted for 0.89% of total SNPs including 928 nonsense and 87,621 missense mutations (Table S5).

Genetic diversity and ROH detection

To gauge the level of within-population genetic diversity, we computed nucleotide diversity in windows of 1 Mb across the cattle genome. Nucleotide diversity is defined as the average number of nucleotide differences per site between two randomly chosen DNA sequences in a population (Nei & Li, 1979). Creole cattle from Guadeloupe population has the highest nucleotide diversity (median = $8,16 \times 10^{-4}$) which is consistent with their three-way admixture. African taurines have the lowest diversity (median = $6,28 \times 10^{-4}$) (Fig. 1a) which can mainly be explained by a higher level of inbreeding. This is reflected by both a higher number of ROH and a larger cumulative ROH length compared to the other populations. Conversely, ROH detection in windows of minimum amount of 100 homozygous SNPs revealed that GUA has the lowest number of ROH (average number of 153 ± 34 ROHs) (Fig. S1a). Additionally, together with Africain indicine populations, GUA has most of its individuals with a total ROH length per individual, below 200 Mb (Fig. S1b).

A total of 12 ROH islands located on chromosomes 1, 5, 6, 7, 10, 11, 12 and 19 were identified in the GUA genome. Among the detected ROH islands, the strongest pattern was observed on BTA11 (34,764,780 – 34,919,599 bp), BTA05 (48,438,356 – 49,020,572 bp) and BTA12 (28,635,496 – 28,916,516 bp) with an overlapping ROH region present in 47%, 41% and 41 % of the samples, respectively (Table S6).

Population structure and genetic relationship analyses

Principle component analysis (PCA) grouped individuals in clusters according to their populations of origin (Fig. 1b). The first principal component (PC1) explained approximately 5.72 % of the global variation and formed a gradient between American-Australian and Asian indicine cattle on one hand and European taurines on the other hand. The second principal component (PC2) explained approximately 2.63 % of the global variation and defines the differences between African and European taurines. The three AFI populations, BARK, BOR and BUT are in the same genetic cluster and maintain a certain distance from the AMASI group. GUA fell at an intermediate position between AFI and EUT groups. It is worth noting that the two Criollo (CRIO) breeds (the Colombian and Limonero individuals) are very close to the Iberian group with the latter being very close to the other European breeds. *ADMIXTURE* analysis also recapitulated these findings (Fig. 1c). When K = 3, European breeds (blue) were separated from American-Australian and Asian indicine (red) and African taurines (green). GUA individuals have, on average, 29%, 35% and 36% of EUT, AFT and IND ancestries, respectively. When K was set to 4, the European ancestry of GUA is mainly from southern Europe while the Criollo cattle have a higher proportion of Northern European ancestry (represented by Holstein (HOL)) and very little proportion of AFT and IND ancestries. Increasing K to 5, separated Limonero from the Colombian breeds. From K=3 through K = 5, the three African zebu populations showed a similar genetic structure composed exclusively from indicine (75% on average) and African taurine (25% on average) ancestries. To have a more accurate picture of AFT, EUT and indicine contribution across the mixed ancestry of GUA genome, we ran the ELAI algorithm for each GUA individual. The results show that the contributions of each of the three ancestries is not homogeneous across individuals. For instance, individuals GUA14 and GUA2 have the highest proportions of EUT introgression (42.2 and 41.6%, respectively). This contrasts with individuals GUA15 and GUA16 which have only 16% of EUT ancestry (Fig. S2). Likewise, the proportions of the three ancestries vary widely across chromosomes. For instance, AFT ancestry in GUA8 (individual with the highest global AFT ancestry (48,7%)) varied between 36,2% on the BTA04 and 63,3% on the BTA10 (Fig. S2).

Pairwise Fst values obtained with Genepop corroborated the genetic proximity within AFI populations (0.016<Fst<0.044) and between these latter and GUA (0.049 <Fst<0.062) while Fst estimates between GUA

and AMASI breeds were almost two times greater (Fst > 0.1). Moderate values of Fst are observed between GUA and EUT breeds with a South European origin, IBER, BAQ, LIM and CHA (0.0691<Fst<0.0799) (Table S7).

We have examined our data with the TreeMix software that allows to model both population splits and gene flow between populations. Ten migration events were sequentially added to the phylogenetic tree which explained 99.79% of the model's variance in terms of relatedness between populations. The resulting phylogenetic network structure shows GUA as a sister population to Ndama. Both populations are in clade with the East African Zebu, Butana. The other two African zebu populations (BOR and BARK) were in clade with the Asian zebu (BRA and SAHW). TreeMix shows GUA strongly introgressed with the African taurine Muturu and the East African zebu, Barka. A third edge originating more basally in the phylogenetic network was also placed towards GUA. High levels of admixture are shown between African taurines (MUTU and Ndama) and African zebu (BOR and BUT) (Fig. S3).

Selection signature detection in Creole cattle from Guadeloupe

To detect genomic regions putatively contributing to local adaptation in GUA cattle, we used 3 EHH-derived statistics (*iHS* , *Rsb* and *XP-EHH*) based on the decay of haplotype homozygosity as a function of recombination distance. Candidate regions were defined by groups of at least 5 outlier SNPs exceeding the significance threshold of -log10 (p-value) = 6. The rationale is that selective sweeps tend to produce clusters of extreme scores across the sweep region, while under a neutral model, extreme scores are scattered more uniformly (Voight, Kudaravalli, Wen, & Pritchard, 2006). *Rsb* and Cross-population Extended Haplotype Homozygosity (*XP-EHH*) statistics were computed at each SNP. Haplotypes estimated in each population were pooled, for each autosome, according to their group of origin (AFT, EUT and IND). In total, 14, 102 and 44 haplotypes were thus considered representative of AFT, EUT and IND ancestry, respectively. *Rsb* detected 6, 11 and 8 regions putatively under selection for GUA/AFT, GUA/EUT and GUA/IND comparisons, respectively (Fig. 2a, 2b, 2c and Table 1). *XP-EHH* identified 3, 13 and 8 regions putatively under selection for GUA/AFT, GUA/EUT and GUA/IND comparisons, respectively (Fig. 2d, 2e, 2f and Table 1). The two significant windows on BTA04 (at position: 113 - 113.5 Mb) and BTA05 (at position: 99 - 99.5 Mb) revealed by the intra-population iHS test (Fig. 2g) were among the candidate regions jointly detected by Rsb and XP-EHH tests (Table 1). The variant under strongest selection on chromosome 4 (-log(P-value) > 11) fell within GIMAP genes. On chromosome 5 the highest signal fell 16 Kb upstream NKG2-A/NKG2-B type II integral membrane protein. Overall, 17 candidate regions were identified by at least two EHH-based tests (Table 1) of which six regions have significantly positive Rsb values suggesting that they are under selection in GUA. Of these, five regions located on BTA02 (at position : 120 - 120.5 Mb), BTA05 (at position : 47 - 47.5 Mb), BTA06 (at position : 69 - 69.5 Mb), BTA12 (at position : 29 - 30 Mb) and BTA13 (at position : 63.5 - 64 Mb) were identified in the EUT Vs GUA comparison. The first four candidate regions overlapped with areas harboring a sudden increase in indicine allele dosage (Fig. S4) which are among the top 1% regions with the highest indicine ancestry in the whole genome. The region on chromosome 13 is ~ 1 Mb away from one of these top 1% regions (Position : 65.175 – 65.775 Mb) (Table S8). On this chromosome, we observed a second 3-Mb region with a high indicine ancestry located between 45 – 48 Mb. All five aforementioned regions showed a reduced level of diversity within GUA and an increased level of genetic differentiation between GUA on one hand and AFT and EUT on the other hand (Fig. 3 and Fig. S5). Furthermore, these five regions overlapped with one of the top 1% windows showing the highest genetic differentiation with both AFT and EUT breeds (Tables S9 and S10).

We also tested for signals of selection using the composite likelihood ratio implemented in SweeD software to detect patterns of site frequency spectrum in GUA population. We found that all the five candidate regions had the highest CLR values in their respective chromosomes (Fig. S6). Genes that fell within the peak area of the aforementioned five candidate regions in addition to the outlier window on the BTA04 (at position 113 - 113.5 Mb) detected with the *iHS* test were considered as relevant candidates (Table 2). On chromosome 2, the highest signal is 73.5 Kb downstream of *EIF4E2* gene. On chromosome 5, the two variants with the highest p-value fell within *GRIP1* , a gene that facilitates the anti-inflammatory effects of

7

glucocorticoids in vivo (Chinenov et al., 2012). We identified another gene in this region : *DNA Helicase B (HELB)* located 250 Kb upstream the highest signal. In candidate region on chromsome 12, the three SNPs of highest significance (8.48<-log(P-value)<10.4) are less than 80 Kb upstream of relaxin family peptide receptor 2 (*RXFP2*). On chromosome 13, the cluster of the four most signicant outlier SNPs is less than 100 Kb upstream the *ASIP* gene. Overall, these genes have functions associated with response to exercise and to DNA damage, immune response, horn size, coat color and behavior (Table 2).

Discussion

A unique population structure of GUA

In this paper, we present a characterization of the first complete genome sequence of the creole cattle from Guadeloupe. In line with previous reports (Gautier & Naves, 2011), our results highlight the three-way, European taurine x African taurine x indicine admixture in the Creole cattle from Guadeloupe. Quantifying the amount of each of the three ancestries across the chromosomes indicate that, despite some variation observed in admixture proportions between GUA individuals genomes, there is a clear trend in favour of a dominance of non European ancestries in the genome of almost all GUA animals. Conversely, we found high proportions of EUT ancestry in the other Creole breeds (Fig. 1c) leading to a clear separation between these and GUA in our data set regardless of the analytical method we used (Fig. 1b,c and Fig. S3). More particularly, the GUA population is not as much as closely related to Iberian cattle as the other Criollo breeds used in the present study do (Fig. 1b,c). Previous studies found that American Criollo cattle originate from Iberia with African ancestry inherited via Iberian ancestors (Decker et al., 2014). Our results suggest that this does not seem to be the case of Creole cattle from Guadeloupe. The various analyses are consistent with a model wherein GUA population originates from a direct introduction of African taurine cattle to Guadeloupe (J.-C. Maillard & Maillard, 1998; Gautier & Naves, 2011). A previous analysis of sequence variation in the hypervariable segment of the mitochondrial DNA control region (mtDNA) similarly reported a high proportion of African mtDNA haplotypes in twenty five GUA indivduals (Magee et al., 2002). Interestingly, the aformentionned study included Creole samples from Antigua and St. Lucia, two Guadeloupe's neighboring regions and found that mtDNA pools of these samples are predominantly European, with very low African mtDNA admixture proportions. Moreover, when using microsatellite data, the authors of the study reported that the Guadeloupe cattle was the only population to exhibit West African–specific alleles (Magee et al., 2002). This further confirms the uniqueness of this population with respect to the other Creole breeds even those found in neighbouring regions and which are expected to have a similar genetic history.

The lower genetic differentiation and a closer position in PCA analysis between GUA and East African zebu compared to indian zebu and the migration edge placed by TreeMix between GUA and the East African Zebu Barka, all suggest that the indicine ancestry in the GUA genome has an African origin. African indicine ancestry in GUA genome could be inherited via west african taurines which also carry an indicine component in their genome stemming from past admixture events between migrating populations of East African zebu and local taurine cattle in West Africa. This admixture was previously reported (Flori et al., 2014) and is recognizable in our phylogenetic network where migration edges are placed between African taurines (MUTU and Ndama) and African zebu (BOR and BUT) (Fig. S3). Another plausible hypothesis is that African indicine ancestry of GUA cattle could be inherited from African zebu populations accompanying nomadic people such as Fulani. Pastoral populations represent a major force for genetic exchange between taurine and indicine cattle all over central and West Africa through transhumance of their livestock along the African Sahel. Written records suggest that nomadic herders spread from West Africa (currently Senegal, Guinea, Mauritania) around 1000 years ago, reaching the Lake Chad Basin 500 years later (Vicente et al., 2019). Clearly, estimating with high accuracy the origin of African indicine ancestry in the GUA genome would require using additional cattle populations from western Africa and the Sahel region.

Detecting selection signature in GUA genome

To identify footprints of selection in GUA cattle, we used pools of individuals haplotypes belonging to

8

contemporary populations sampled from geographical locations remote from each other, as proxies for their assumed ancestral counterparts. In this regard, EUT ancestry was represented by breeds from France, Spain as well as Criollo populations which were shown to be genetically close to the former breeds (See Fig. 1b,c (K = 4)). Likewise, indicine ancestry was represented by zebu breeds from East Africa, America, Australia and the Indian subcontinent. Such a haplotype pooling applied to an extensive comparison of groups of populations would 'smooth out' effects specific to each of the populations from a given ancestry (Gautier & Naves, 2011).

As a first step, we applied two types of complementary EHH-based statistics, the integrated Haplotype Homozygosity Score (*iHS*) reflecting primarily ongoing selection, *Rsb* and *XP-EHH*, having the most power to detect completed selection after fixation of the advantageous allele (Voight et al. 2006; Sabeti et al. 2007, Tang et al., 2007).

Two candidate regions on chromosomes 4 (position : 113 – 113.5 Mb) and 5 (position : 99 – 99.5 Mb) were jointly identified by the two types of approaches. Additionally, we observe that the candidate region on chromosome 4 is also under selection in AFT since it has a negative*Rsb* value. This is further supported by the identification of this region when we computed the *Rsb* statistic for AFT Vs EUT and AFT Vs IND comparaisons (Tables S11 and S12). The strongest selection signal in this region fell within *GIMAP* genes which play a central role in lymphocyte maturation and lymphocyte-associated diseases (Schwefel et al., 2010). *GIMAP5* knockout mice have complete loss of natural killer cells (Schulteis et al., 2008) which are critical to the protective response during *Trypanosoma cruzi* infection, (Duthie & Kahn, 2005) widespread in West Africa. Taken together, these findings lead us to speculate *GIMAP* genes are under ongoing selective pressure in GUA genome owing to the presence of novel pathogens in the Caribbean islands (compared to those present in Africa). Infectious pathogens are among the strongest selective forces that shape the genome of several mammalian species such as human (Karlsson, Kwiatkowski, & Sabeti, 2014).

One of the main drawbacks of selection signature detection methods is their elevated rate of false positives (Thornton & Jensen, 2007). Limiting the number of spurious signals that can arise owing to various confounding factors such as the marker discovery process and/or population demographics is the main challenge in genome-wide scans aiming at the detection of selective sweeps. In the present study every attempt has been made to reduce the number of false-positive signals while focusing on candidate regions under selection in the GUA genome. First, the use of whole genome sequencing should reduce SNP ascertainment bias afflicting commercial genotyping arrays. Avoiding SNP ascertainment bias is critical for accurate population genetic analyses because levels of variability, distribution of allele frequencies, and levels of linkage disequilibrium will all be strongly affected by such ascertainment schemes (Nielsen, Hubisz, & Clark, 2004 ; McTavish & Hillis, 2015). Second, in our EHH-based tests, we relied on consistency of signals over regions, that is taking clustering of highly significative outliers as evidence for selection. Third, and most importantly, we considered that selection signals that are consistently supported across different statistical tests are less likely to be false-positives. Indeed, although these tests are designed to detect selective sweeps that vary in terms of type, age and strength of selection events, selection signals supported across different methodologies can increase power, reduce sensitivity to confounding factors (which are unlikely to affect different methods in a similar manner) and increase precision of the detection of the selective sweep (Grossman et al., 2010). Here, we used various methods based on excess of haplotype homozygosity, deformation of the allele frequency spectrum, excess of differentiation in allelic frequencies between GUA population and proxies of its ancestral populations and reduction in genetic diversity around the selected region to identify reliable selection signatures in GUA genome. Fourth, we relied on the identification of excess / deficiency of local ancestry in GUA genome to further confirm selection signals revealed by the various tests we used. The rationale is that, under a recent admixture scenario followed by a strong selection, we expect a parallel increase in local ancestry proportions in the regions surrounding the beneficial variants. We found evidence of congruent signals between methods for five candidate regions on chromosomes 2, 5, 6, 12 and 13, all of them identified in the GUA Vs EUT comparison. These five regions exhibit a sudden increase in indicine ancestry (Fig. S4), clusters of highly significant SNPs in *Rsb*and *XP-EHH* tests (Fig. 2), low nucleotide diversity and high differentiation levels between GUA on one hand and EUT and AFT on the other hand (Fig. 3 and Tables

9

S9 and S10). These regions also displayed the highest CLR values on their respective chromosomes (Fig. S6). A further result is that the two regions on chromosomes 5 and 12 overlap with ROH islands including at least 35% of the individuals (Table S6). Altogether these findings validate the five genomic regions as the most biologically relevant results and support previous studies suggesting that adaptation of GUA cattle to tropical environment occurred mainly through its indicine ancestry (Gautier & Naves 2011).

Adopting stringent criteria to declare candidate regions in our EHH-based tests (5 SNPs exceeding the significance threshold of 10^{-6} within 500-Kb windows) constitutes a potential limitation of our study since such approach is likely to lead to a large number of missed selection signals. This might partially explain the little congruence with the candidate genomic regions reported by Gautier & Naves 2011. When we remade *Rsb* calculation for EUT Vs GUA comparison using the same criteria reported by these authors (threshold p-value = 10^{-4} in 1-Mb sliding windows with 500-kb overlapping step and one marker exceeding the significance threshold by window), we detected 278 regions of which 13 reported in Gautier & Naves 2011 (out of 16 candidate regions) (Table S13).

Biological function of the most relevant candidate genes located within the relevant candidate regions

The highest selection signals in the five strong candidate regions showing an excess of indicine ancestry fell within or nearby genes involved in stress response to tropical constraints and probably to some human-driven socio-cultural pressure (Table 2) . Many of our candidate genes have been reported in other species. For instance, *ASIP,* a gene linked to skin pigmentation in human (Liu et al., 2015) and mice (Bultman et al., 1994) was also localized in a strong selective sweep in Indian water buffalo (Dutta et al., 2020). *ASIP* was previously reported to be associated with darkness of hair coat in Nellore cattle (Trigo et al., 2021). Coat color is an important potential adaptive function that helps regulate body temperature in mammals (Stuart-Fox, Newton, & Clusella-Trullas, 2017).

Our results demonstrate that the Creole cattle from Guadeloupe which are usually exposed to long periods of direct, intense sunlight either during grazing or during ploughing, possess signatures of putative selection within or around genes associated with thermo-tolerance. Aside from *ASIP* gene,

the strongest candidates within the region on chromosome 5 are *GRIP1* and *HELB.* Both genes were previously identified in a 430-kb selective sweep in Asian indicine cattle (Naval-Sanchez et al., 2020). Importantly, *HELB* is known to be involved in the response to DNA damage and replication stress (Hazeslip, Zafar, Chauhan, & Byrd, 2020) that could be induced by prolonged exposure to solar ultraviolet radiation (Rastogi, Richa, Kumar, Tyagi, & Sinha, 2010). We also confirmed the adaptive role of *Relaxin family peptide receptor 2* (*RXFP2*) gene, previously identified by Gautier & Naves 2011. *RXFP2* is a gene with a pleiotropic effects. It affects both inguinoscrotal testis descent (Yuan et al., 2010) and horn size in wild bighorn (Kardos et al., 2015) and domestic sheep (Luhken et al., 2016; Pan et al., 2018). The use of horns is likely part of a thermoregulatory mechanism in several pecoran species. Indeed, since the core of the horn is part of the sinus, horns may contribute to nasal heat exchange, a mechanism that considerably reduces water loss through cooling of the air during exhalation (Langman, Maloiy, Schmidt-Nielsen, & Schroter, 1979). Another explanation to the selection pressure exerted on *RFLXP2* is that Creole bulls were traditionally used for sugarcane cart pulling and cattle cart race competition, with a yoke attached to the horns. This has probably led to the development of a stronger horn base (Gautier & Naves 2011). Creole cattle have been historically selected for draught works. They are known to have a better endurance in long-term effort than crossbreed or exotic breeds (Versini, 1997). Accordingly, we found that the strongest evidence of selective pressure on chromosomes 2 co-localizes with *EIF4E2.* In human cells under hypoxia, eIF4E2 plays a fundamental role in protein synthesis. This gene substitutes its homologue, eIF4E and forms a complex with the oxygen-regulated hypoxia-inducible factor 2α (Uniacke et al., 2012). *EIF4E2* was among the candidate genes that were shown to be under positive selection in Fuzhong buffalo characterized by strong muscles and able to endure the strength to pull a plough through muddy rice paddies (Sun et al., 2020). We also identified another gene, *OCIAD1,* as a good candidate for mitochondrial adaptation during exercise. *OCIAD1* lies within a 3-Mb region which is among the top 1% regions with the highest indicine ancestry and is located in the vicinity of our candidate region on BTA06. The region encompassing this

gene is detected by *Rsb* and *XP-EHH* tests when relaxing the P-value threshold to 10^{-5}. Almost ~ 11% of the 220 SNPs contained in this region (66.5 – 68 Mb) exceed the significance threshold (data not shown). Importantly, the two variants with the highest P-value in this region are located 16 Kb upstream *OCIAD1* which encodes a mitochondrial inner membrane protein that regulates mitochondrial Complex III assembly in cells (Le Vasseur et al., 2021). The latter is among the complexes that play a key role in electron transport and proton gradient production, precisely across the inner mitochondrial membrane (Bennett et al., 2022). Proton gradient provides the energy necessary for the production of ATP whose demand increases in the muscle with exercise intensity (Calbet et al., 2020). Selection for draught traits in GUA population thus seems to have provoked responses in a diversity of pathways involving at least *EIF4E2* and *OCIAD1* genes. Such selective pressure promoted the fixation of beneficial alleles from an indicine origin that allowed GUA individuals to cope with withstanding hours of high intensities of physical activity.

Conclusion

In this study, we have generated for the first time a catalog of genetic variants found in the Creole cattle from Guadeloupe. We were able to show that GUA adaptation to local environment occurred mainly through its indicine component. We also demonstrate that pathogenic environment, thermo-tolerance and physical stamina are important drivers of local adaptation in Creole cattle. Overall, our results provide clues for understanding the adaptive admixture in the Creole cattle from Guadeloupe thus contributing to the emerging picture of the genes and pathways associated with traits resilience in livestock species. Our study may represent a starting point for a targeted and sustainable genetic breeding improvement of Creole cattle.

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19* (9), 1655–1664. doi: 10.1101/gr.094052.109

Barbato, M., Hailer, F., Upadhyay, M., Del Corvo, M., Colli, L., Negrini, R., ... Ajmone-Marsan, P. (2020). Adaptive introgression from indicine cattle into white cattle breeds from Central Italy. *Scientific Reports*, *10* (1), 1279. doi: 10.1038/s41598-020-57880-4

Bennett, C. F., Latorre-Muro, P., & Puigserver, P. (2022). Mechanisms of mitochondrial respiratory adaptation. *Nature Reviews Molecular Cell Biology*, *23* (12), Article 12. https://doi.org/10.1038/s41580-022-00506-6

Bickhart, D. M., Xu, L., Hutchison, J. L., Cole, J. B., Null, D. J., Schroeder, S. G., ... Liu, G. E. (2016). Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. DNA Research, 23(3), 253–262. doi: 10.1093/dnares/dsw013

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30* (15), 2114–2120. doi: 10.1093/bioinformatics/btu170

Bultman, S. J., Klebig, M. L., Michaud, E. J., Sweet, H. O., Davisson, M. T., & Woychik, R. P. (1994). Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a(t)) and white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. *Genes & Development*, *8* (4), 481–490. doi: 10.1101/gad.8.4.481

Calbet, J. A. L., Martín-Rodríguez, S., Martin-Rincon, M., & Morales-Alamo, D. (2020). An integrative approach to the regulation of mitochondrial respiration during exercise: Focus on high-intensity exercise. Redox Biology, 35, 101478. https://doi.org/10.1016/j.redox.2020.10147

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., ... Lei, C. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nature*

Communications , *9* (1), 2337. doi: 10.1038/s41467-018-04737-0

Chinenov, Y., Gupte, R., Dobrovolna, J., Flammer, J. R., Liu, B., Michelassi, F. E., & Rogatsky, I. (2012). Role of transcriptional coregulator GRIP1 in the anti-inflammatory actions of glucocorticoids. *Proceedings of the National Academy of Sciences* , *109* (29), 11776–11781. doi: 10.1073/pnas.1206059109

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* , *6* (2), 80–92. doi: 10.4161/fly.19695

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., . . . Taylor, J. F. (2014). Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics* , *10* (3). doi: 10.1371/journal.pgen.1004254

Duthie, M. S., & Kahn, S. J. (2005). NK cell activation and protection occur independently of natural killer T cells during Trypanosoma cruzi infection. *International Immunology* , *17* (5), 607–613. doi: 10.1093/intimm/dxh239

Dutta, P., Talenti, A., Young, R., Jayaraman, S., Callaby, R., Jadhav, S. K., . . . Prendergast, J. G. D. (2020). Whole genome analysis of water buffalo and global cattle breeds highlights convergent signatures of domestication. *Nature Communications* , *11* (1), 4739. doi: 10.1038/s41467-020-18550-1

Flori, L., Thevenon, S., Dayo, G.-K., Senou, M., Sylla, S., Berthier, D., . . . Gautier, M. (2014). Adaptive admixture in the West African bovine hybrid zone: Insight from the Borgou population. *Molecular Ecology* , *23* (13), 3241–3257. doi: 10.1111/mec.12816

Gauthier, D., Aumont, G., Barré, N., Berbigier, P., Camus, E., Lafortune, E., . . . Thimonier, J. (1984). Le bovin créole en Guadeloupe: Caractéristiques et performances zootechniques. *Revue d'élevage et de médecine vétérinaire des pays tropicaux* , *37* (2), 212–224. doi: 10.19182/remvt.8439

Gautier, M., & Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology* , *20* (15), 3128–3143. doi: 10.1111/j.1365-294X.2011.05163.x

Gautier, M., & Vitalis, R. (2012). rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics (Oxford, England)* , *28* (8), 1176–1177. doi: 10.1093/bioinformatics/bts115

Ginja, C., Gama, L. T., Cortés, O., Burriel, I. M., Vega-Pla, J. L., Penedo, C., . . . Martinez, A. (2019). The genetic ancestry of American Creole cattle inferred from uniparental and autosomal genetic markers. *Scientific Reports* , *9* (1), 11486. doi: 10.1038/s41598-019-47636-0

Grossman, S. R., Shlyakhter, I., Karlsson, E. K., Byrne, E. H., Morales, S., Frieden, G., . . . Sabeti, P. C. (2010). A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science (New York, N.Y.)* , *327* (5967), 883–886. doi: 10.1126/science.1183863

Guan, Y. (2014). Detecting Structure of Haplotypes and Local Ancestry. *Genetics* , *196* (3), 625–642. doi: 10.1534/genetics.113.160697

Hazeslip, L., Zafar, M. K., Chauhan, M. Z., & Byrd, A. K. (2020). Genome Maintenance by DNA Helicase B. *Genes* , *11* (5), 578. doi: 10.3390/genes11050578

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* , *24* (11), 1403–1405. doi: 10.1093/bioinformatics/btn129

Kanetsky, P. A., Swoyer, J., Panossian, S., Holmes, R., Guerry, D., & Rebbeck, T. R. (2002). A Polymorphism in the Agouti Signaling Protein Gene Is Associated with Human Pigmentation. *The American Journal of Human Genetics* , *70* (3), 770–775. doi: 10.1086/339076

Kardos, M., Luikart, G., Bunch, R., Dewey, S., Edwards, W., McWilliam, S., . . . Kijas, J. (2015). Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Molecular Ecology* , *24* (22), 5616–5632. doi: 10.1111/mec.13415

Karlsson, E. K., Kwiatkowski, D. P., & Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nature Reviews Genetics* , *15* (6), 379–393. doi: 10.1038/nrg3734

Khvorykh, G. V., & Khrunin, A. V. (2020). imputeqc: An R package for assessing imputation quality of genotypes and optimizing imputation parameters. *BMC Bioinformatics* , *21* (Suppl 12), 304. doi: 10.1186/s12859-020-03589-0

Kim, K., Kwon, T., Dessie, T., Yoo, D., Mwai, O. A., Jang, J., . . . Kim, H. (2020). The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nature Genetics* ,*52* (10), 1099–1110. doi: 10.1038/s41588-020-0694-2

Koufariotis, L., Hayes, B. J., Kelly, M., Burns, B. M., Lyons, R., Stothard, P., . . . Moore, S. (2018). Sequencing the mosaic genome of Brahman cattle identifies historic and recent introgression including polled. Scientific Reports, 8(1), 17761. doi: 10.1038/s41598-018-35698-5

Le Vasseur, M., Friedman, J., Jost, M., Xu, J., Yamada, J., Kampmann, M., Horlbeck, M. A., Salemi, M. R., Phinney, B. S., Weissman, J. S., & Nunnari, J. (2021). Genome-wide CRISPRi screening identifies OCIAD1 as a prohibitin client and regulatory determinant of mitochondrial Complex III assembly in human cells. *ELife* , *10* , e67624. https://doi.org/10.7554/eLife.67624

Langman, V. A., Maloiy, G. M. O., Schmidt-Nielsen, K., & Schroter, R. C. (1979). Nasal heat exchange in the giraffe and other large mammals.*Respiration Physiology* , *37* (3), 325–333. doi: 10.1016/0034-5687(79)90079-3

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25* (16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Liu, F., Visser, M., Duffy, D. L., Hysi, P. G., Jacobs, L. C., Lao, O., . . . Kayser, M. (2015). Genetics of skin color variation in Europeans: Genome-wide association studies with functional follow-up.*Human Genetics* , *134* (8), 823–835. doi: 10.1007/s00439-015-1559-0

Lühken, G., Krebs, S., Rothammer, S., Küpper, J., Mioč, B., Russ, I., & Medugorac, I. (2016). The 1.78-kb insertion in the 3'-untranslated region of RXFP2 does not segregate with horn status in sheep breeds with variable horn status. *Genetics Selection Evolution* , *48* (1), 78. doi: 10.1186/s12711-016-0256-3

Magee, D. A., Meghen, C., Harrison, S., Troy, C. S., Cymbron, T., Gaillard, C., . . . Bradley, D. G. (2002). A Partial African Ancestry for the Creole Cattle Populations of the Caribbean. *Journal of Heredity* , *93* (6), 429–432. doi: 10.1093/jhered/93.6.429

Maillard, J. C., Palin, C., Trap, I., & Bensaid, A. (1993). An attempt to identify genetic markers of resistance or susceptibility to dermatophilosis in the zebu Brahman population of Martinique.*Revue D'elevage Et De Medecine Veterinaire Des Pays Tropicaux* ,*46* (1–2), 291–295.

Maillard, J.-C., & Maillard, N. (1998). Historique du peuplement bovin et de l'introduction de la tique Amblyomma variegatum dans les îles françaises des Antilles: Synthèse bibliographique.*Ethnozootechnie* . (Antilles françaises). Retrieved from https://agritrop.cirad.fr/401313/

McTavish, E. J., & Hillis, D. M. (2015). How do SNP ascertainment schemes and population demographics affect inferences about population history? *BMC Genomics* , *16* (1), 266. doi: 10.1186/s12864-015-1469-5

Naval-Sánchez, M., Porto-Neto, L. R., Cardoso, D. F., Hayes, B. J., Daetwyler, H. D., Kijas, J., & Reverter, A. (2020). Selection signatures in tropical cattle are enriched for promoter and coding regions and reveal missense mutations in the damage response gene HELB.*Genetics, Selection, Evolution : GSE* , *52* , 27. doi: 10.1186/s12711-020-00546-6

Naves, M. (2003). Caracterisation et gestion d'une population bovine locale de la zone tropicale: Le bovin Creole de Guadeloupe (INA Paris-Grignon). INA Paris-Grignon, INRA Departement de Genetique Animale. https://hal.inrae.fr/tel-02832443/document. Retrieved from https://hal.inrae.fr/tel-02832443/document

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* ,*76* (10), 5269–5273.

Nielsen, R., Hubisz, M. J., & Clark, A. G. (2004). Reconstituting the Frequency Spectrum of Ascertained Single-Nucleotide Polymorphism Data.*Genetics* , *168* (4), 2373–2382. doi: 10.1534/genetics.104.031039

Pan, Z., Li, S., Liu, Q., Wang, Z., Zhou, Z., Di, R., . . . Li, Y. (2018). Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-feralization. *GigaScience* , *7* (4), giy019. doi: 10.1093/gigascience/giy019

Pavlidis, P., Živkovic, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution* , *30* (9), 2224–2234. doi: 10.1093/molbev/mst112

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics* , *8* (11). doi: 10.1371/journal.pgen.1002967

Pitt, D., Bruford, M. W., Barbato, M., Orozco-terWengel, P., Martínez, R., & Sevane, N. (2019). Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evolutionary Applications* , *12* (1), 105–122. doi: 10.1111/eva.12641

Porto-Neto, L. R., Bickhart, D. M., Landaeta-Hernandez, A. J., Utsunomiya, Y. T., Pagan, M., Jimenez, E., . . . Sonstegard, T. S. (2018). Convergent Evolution of Slick Coat in Cattle through Truncation Mutations in the Prolactin Receptor. Frontiers in Genetics, 9. Retrieved from

https://www.frontiersin.org/articles/10.3389/fgene.2018.00057

Primo, A. T. (1992). The Iberic cattle in the Americas: 500 years later.*Archivos de Zootecnia 41 (suppl.), 421-432* .

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.*American Journal of Human Genetics* , *81* (3), 559–575.

Rastogi, R. P., Richa, Kumar, A., Tyagi, M. B., & Sinha, R. P. (2010). Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. *Journal of Nucleic Acids* , *2010* , 592980. https://doi.org/10.4061/2010/592980

Rodero Serrano, E., Rodero Franganillo, A., & Delgado-Bermejo, J. V. (1992). Primitive andalusian livestock and their implications in the discovery of America. *Archivos de Zootecnia 41 (154), 383-400* . Retrieved from http://helvia.uco.es/xmlui/handle/10396/3294

Rosenberg, N. A. (2004). distruct: A program for the graphical display of population structure. *Molecular Ecology Notes* , *4* (1), 137–138. doi: 10.1046/j.1471-8286.2003.00566.x

Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* , *8* (1), 103–106. doi: 10.1111/j.1471-8286.2007.01931.x

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., . . . Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations.*Nature* , *449* (7164), 913–918. doi: 10.1038/nature06250

Scheet, P., & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human*

Genetics , *78* (4), 629–644. doi: 10.1086/502802

Schulteis, R. D., Chu, H., Dai, X., Chen, Y., Edwards, B., Haribhai, D., . . . Weiler, H. (2008). Impaired survival of peripheral T cells, disrupted NK/NKT cell development, and liver failure in mice lacking Gimap5. *Blood* , *112* (13), 4905–4914. doi: 10.1182/blood-2008-03-146555

Schwefel, D., Frohlich, C., Eichhorst, J., Wiesner, B., Behlke, J., Aravind, L., & Daumke, O. (2010). Structural basis of oligomerization in septin-like GTPase of immunity-associated protein 2 (GIMAP2).*Proceedings of the National Academy of Sciences* , *107* (47), 20299–20304. doi: 10.1073/pnas.1010322107

Stuart-Fox, D., Newton, E., & Clusella-Trullas, S. (2017). Thermal consequences of colour and near-infrared reflectance.*Philosophical Transactions of the Royal Society B: Biological Sciences* , *372* (1724), 20160345. doi: 10.1098/rstb.2016.0345

Sun, T., Huang, G., Wang, Z., Teng, S., Cao, Y., Sun, J., . . . Liao, Y. (2020). Selection signatures of Fuzhong Buffalo based on whole-genome sequences. *BMC Genomics* , *21* (1), 674. doi: 10.1186/s12864-020-07095-8

Talenti, A., Powell, J., Hemmink, J. D., Cook, E. a. J., Wragg, D., Jayaraman, S., . . . Prendergast, J. G. D. (2022). A cattle graph genome incorporating global breed diversity. Nature Communications, 13(1), 910. doi: 10.1038/s41467-022-28605-0

Tang, K., Thornton, K. R., & Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biology* , *5* (7). doi: 10.1371/journal.pbio.0050171

Thornton, K. R., & Jensen, J. D. (2007). Controlling the False-Positive Rate in Multilocus Genome Scans for Selection. *Genetics* ,*175* (2), 737–750. doi: 10.1534/genetics.106.064642

Toro-Ospina, A. M., Herrera Rios, A. C., Bizarria Santos, W., Pimenta Schettini, G., Vallejo Aristizabal, V. H., Tovar Claros, G., & Morea, E. G. O. (2022). Genetic Architecture and Signatures of Selection in the Caqueteno Creole (Colombian Native Cattle). *Diversity* ,*14* (10), 828. doi: 10.3390/d14100828

Trigo, B. B., Utsunomiya, A. T. H., Fortunato, A. A. A. D., Milanesi, M., Torrecilha, R. B. P., Lamb, H., . . . Utsunomiya, Y. T. (2021). Variants at the ASIP locus contribute to coat color darkening in Nellore cattle. *Genetics Selection Evolution* , *53* (1), 40. doi: 10.1186/s12711-021-00633-2

Uniacke, J., Holterman, C. E., Lachance, G., Franovic, A., Jacob, M. D., Fabian, M. R., . . . Lee, S. (2012). An oxygen-regulated switch in the protein synthesis machinery. *Nature* , *486* (7401), 126–129. doi: 10.1038/nature11055

Upadhyay, M., Bortoluzzi, C., Barbato, M., Ajmone-Marsan, P., Colli, L., Ginja, C., . . . Crooijmans, R. P. M. A. (2019). Deciphering the patterns of genetic admixture and diversity in southern European cattle using genome-wide SNPs. Evolutionary Applications, 12(5), 951–963. doi: 10.1111/eva.12770

Versini, F. (1997). De la competition des boeuf-tirants et des charretiers en Guadeloupe. *Ethnozootechnie* , *60* , 75–83.

Vicente, M., Priehodova, E., Diallo, I., Podgorna, E., Poloni, E. S., Černý, V., & Schlebusch, C. M. (2019). Population history and genetic adaptation of the Fulani nomads: Inferences from genome-wide data and the lactase persistence trait. *BMC Genomics* , *20* (1), 915. doi: 10.1186/s12864-019-6296-7

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biology* , *4* (3), e72. doi: 10.1371/journal.pbio.0040072

Young, P. W. (2018). LNX1/LNX2 proteins: Functions in neuronal signalling and beyond. *Neuronal Signaling* , *2* (2), NS20170191. doi: 10.1042/NS20170191

Yuan, F. P., Li, X., Lin, J., Schwabe, C., Büllesbach, E. E., Rao, C. V., & Lei, Z. M. (2010). The role of RXFP2 in mediating androgen-induced inguinoscrotal testis descent in LH receptor knockout mice. *Reproduction (Cambridge, England)* , *139* (4), 759–769. doi: 10.1530/REP-09-0518

Data Accessibility

The generated sequences for the 23 Creole cattle from Guadeloupe and 4 Colombian samples are available from the European Nucleotide Archive (ENA) with the Bioproject accession number PRJEB58555.

Author Contributions

N.M and M.N designed the study ; C. K. generated raw data ; S.B.J, M.B and G.A performed sequence alignment and variant discovery. S.B.J performed population structure and selection signature analyses and wrote the manuscript.

Tables and Figures

Table 1. Genomic regions putatively under selection identified using *iHS* , *Rsb* and *XP-EHH* statistics. Regions jointly identified by at least two methods are in bold.

Test	BTA	Start (Mb)	End (Mb)	Population under selection	Genes
iHS	**4**	**113**	**113,5**		**ENSBTAG00000053452, GIM**
	5	**99**	**99.5**		**ENSBTAG00000052617, ENS**
RsbAFT Vs GUA	3	114,5	115	AFT	SH3BP4
	4	**112,5**	**113,5**	**AFT**	**ZNF746, ENSBTAG00000049**
	5	**99**	**99.5**	**GUA**	**ENSBTAG00000052617, ENS**
	11	100.5	101	AFT	NCS1, HMCN2, ASS1, FUBP3, U
	20	**71,5**	**72**	**AFT**	**CEP72, SLC9A3, EXOC3, EM**
	28	38	38.5	AFT	NRG3
Rsb EUT Vs GUA	**1**	**76,5**	**77**	**EUT**	**IL1RAP, TMEM207, CLDN1**
	2	5,5	6	EUT	ENSBTAG00000017214, NAB1, N
	2	**120**	**120,5**	**GUA**	**ENSBTAG00000053448, ENS**
	5	**47**	**47,5**	**GUA**	**GRIP1, U1, HELB**
	5	**74**	**74,5**	**EUT**	**RBFOX2, ENSBTAG0000004**
	6	**69**	**69,5**	**GUA**	**FIP1L1,LNX1, ENSBTAG000**
	12	**29**	**30**	**GUA**	**RXFP2, ENSBTAG00000053**
	13	**63.5**	**64**	**GUA**	**ASIP, AHCY, ENSBTAG000**
	15	**50**	**50,5**	**EUT**	**ENSBTAG00000051323, ENS**
	18	58	58,5	EUT	ENSBTAG00000052289, PPP2R1
	18	**61**	**61,5**	**EUT**	**MGC157082, ENSBTAG0000**
Rsb IND Vs GUA	**1**	**26**	**26,5**	**IND**	**ROBO1**
	2	12,5	13	IND	ENSBTAG00000042196, ENSBTA
	3	40	40.5	IND	COL11A1
	5	**99**	**99,5**	**GUA**	**ENSBTAG00000052617, ENS**
	7	**105**	**105,5**	**IND**	**No genes**
	13	**83**	**83,5**	**IND**	**CBLN4**
	21	**13**	**13,5**	**IND**	**ENSBTAG00000033335**
	27	**30.5**	**31**	**IND**	**ENSBTAG00000042113, UNC**
XP-EHH AFT Vs GUA	4	112.5	114	AFT	**ZNF746, ENSBTAG00000049**
	5	99	99.5	GUA	**ENSBTAG00000052617, ENS**
	20	71,5	72	AFT	**CEP72, SLC9A3, EXOC3, EM**
XP-EHH EUT Vs GUA	1	76,5	77	EUT	**IL1RAP, TMEM207, CLDN1**
	2	120	120,5	GUA	**ENSBTAG00000053448, ENS**
	5	47	47,5	GUA	**GRIP1, U1, HELB**
	5	48,5	49	GUA	MSRB3, LEMD3, WIF1, U6, TB
	5	74	74,5	EUT	**RBFOX2, ENSBTAG0000004**
	6	69	69,5	GUA	**FIP1L1,LNX1, ENSBTAG000**
	9	8	8,5		ADGRB3, ENSBTAG0000005467

16

	10	99	99,5		No genes
	12	**29**	**30**	**GUA**	**RXFP2, ENSBTAG00000053**
	12	61.5	62		No genes
	13	**63.5**	**64**	**GUA**	**ASIP, AHCY, ENSBTAG000**
	15	**50**	**50,5**	**EUT**	**ENSBTAG00000051323, ENS**
	18	**61**	**61,5**	**EUT**	**MGC157082, ENSBTAG0000**
XP-EHH IND Vs GUA	**1**	**26**	**26,5**	**IND**	**ROBO1**
	5	**99**	**99,5**	**GUA**	**ENSBTAG00000052617, ENS**
	7	**105**	**105,5**	**IND**	**No genes**
	13	**83**	**83,5**	**IND**	**CBLN4**
	14	59,5	60	IND	ENSBTAG00000042347
	15	3,5	4	IND	ENSBTAG00000051613 , ENSBT
	21	**13**	**13,5**	**IND**	**ENSBTAG00000033335**
	27	**30,5**	**31**	**IND**	**ENSBTAG00000042113, UNC**

Table 2. The most relevant candidate genes putatively under selection in the Creole cattle from Guadeloupe.

BTA	Region (Mb)	Gene closest to the most significant SNPs	Phenotype	Reference
4	113-113.5	*GIMAP4, GIMAP5, GIMAP7*	parasite resistance	(C. Y. Kim, Zhang, & Witola, 2018)
2	120-120.5	*EIF4E2*	response to exercise	(Sun et al., 2020)
5	47-47.5	*HELB*	response to DNA damage, reproductive traits, yearling weight	Naval-Sánchez et al., 2020
		GRIP1	immune response (anti-inflammatory actions of glucocorticoids)	(Chinenov et al., 2012)
6	69-69.5	*LNX1*	neuronal signaling and anxiety-related phenotypes	(Young, 2018)
12	29-30	*RXFP2*	horn size thermoregulation	(Kardos et al., 2015) Langman, Maloiy, Schmidt-Nielsen, & Schroter, 1979)
13	63.5-64	*ASIP*	coat color	(Kanetsky et al., 2002)

(a) (b)
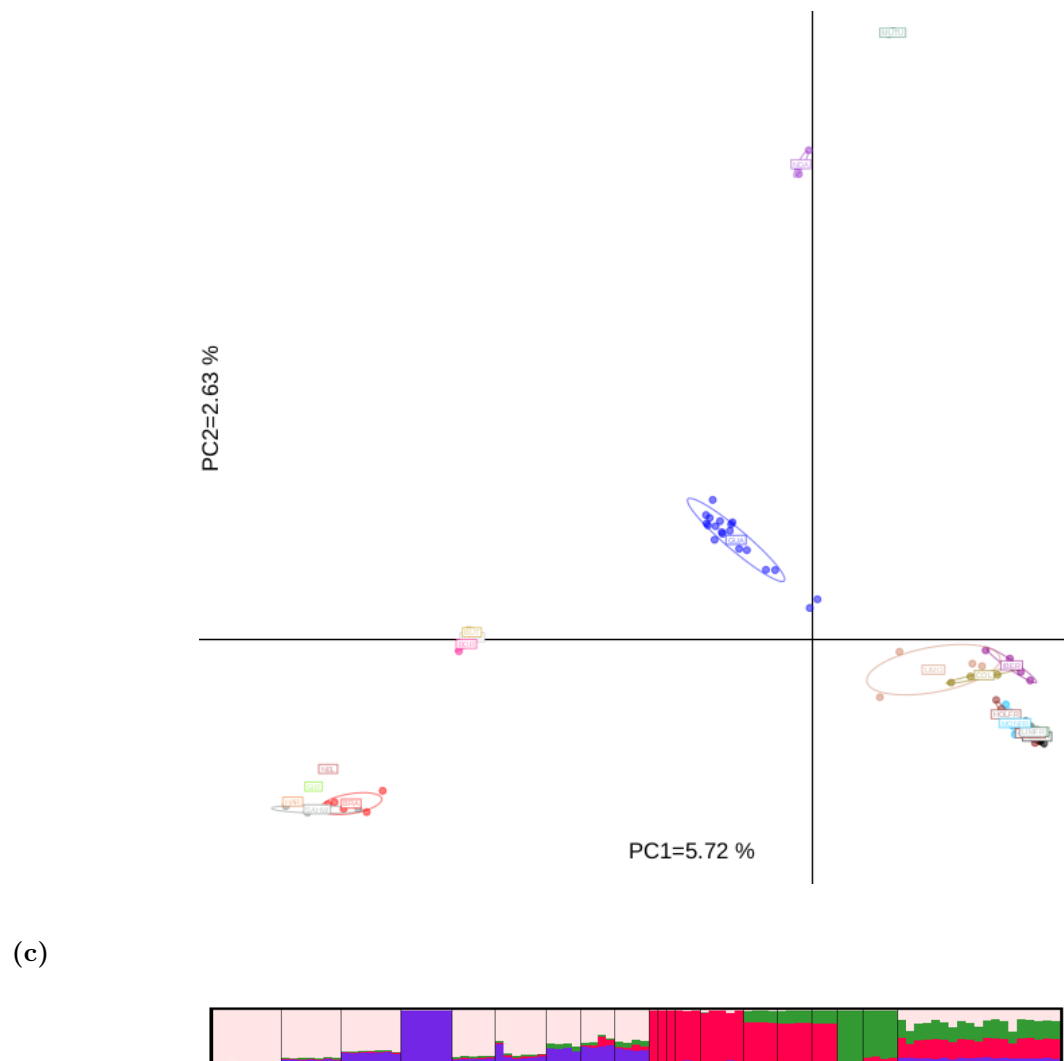
PC2=2.63 %

PC1=5.72 %

(c)
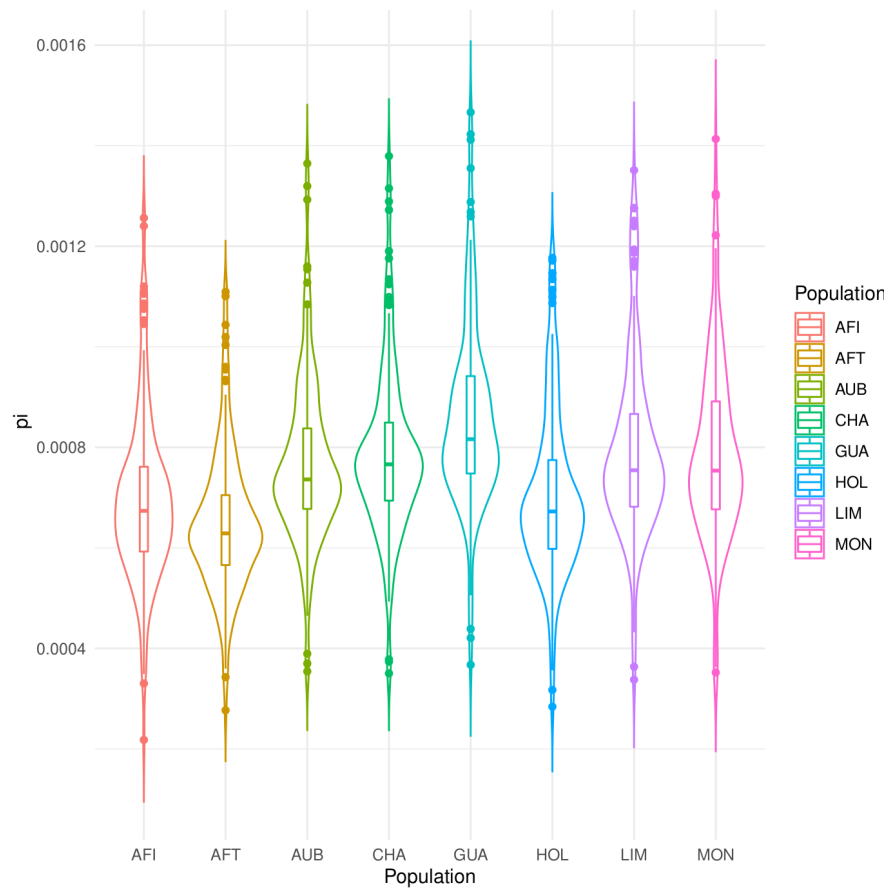
Fig. 1 Genetic diversity and population structure and genetic diversity of Creole cattle from Guadeloupe (GUA). Population names and locations are described in Table S1 (Supporting information).**(a)** Genome-wide distribution of nucleotide diversity (Pi) estimated in 1-Mb nonoverlapping window. AFI : African indicine ; AFT : African taurine **(b)** Principal component analysis, PC 1 against PC 2. CRIO: Criollo (Colombian and venezuelan cattle breeds); IBER : Spanish breeds; AMASI : American-Australian and Asian indicine breeds**(c)** Results of admixture analysis for K 3–5.
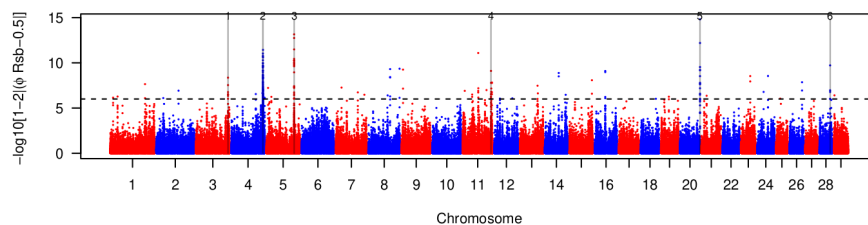

Fig. 2 Manhattan plots showing the results of Extended Haplotype Homozygisty-based tests. **(a)** *Rsb* test

19

AFT Vs GUA cattle. **(b)** Rsb test EUT Vs GUA cattle **(c)** Rsb test IND Vs GUA cattle. **(d)** *XP-EHH* test AFT Vs GUA cattle. **(e)** *XP-EHH* test EUT Vs GUA cattle **(f)***XP-EHH* test IND Vs GUA cattle. **(g)** *iHS* test for GUA cattle. Horizontal dashed lines mark the significance threshold applied to detect the outlier SNPs (–log10 (p-value) = 6.
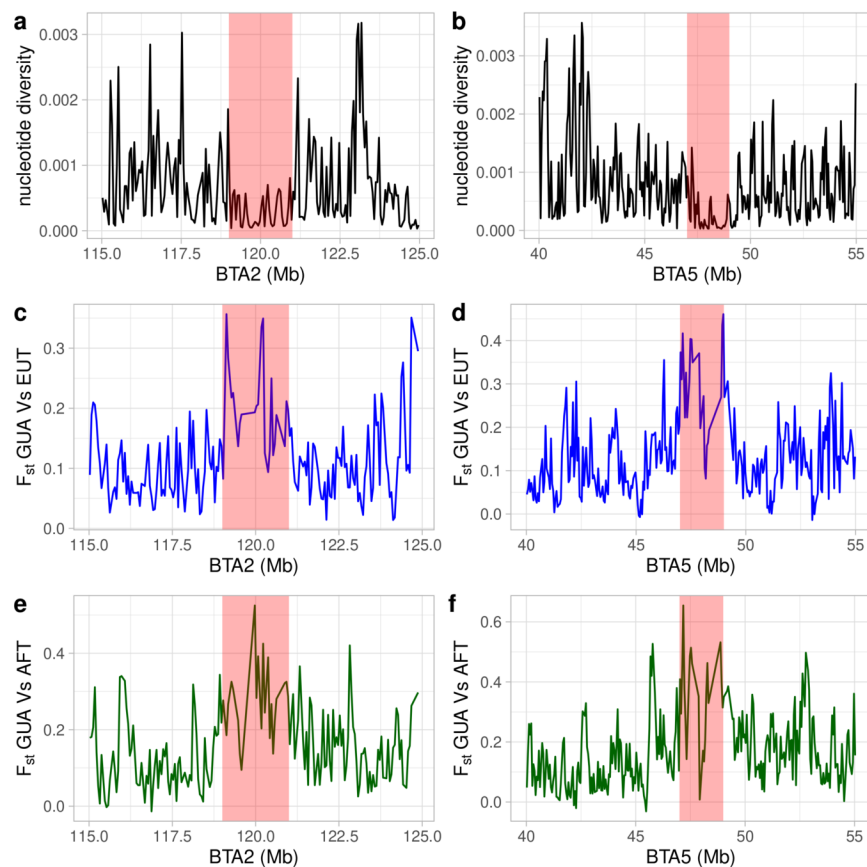

Fig. 3 Nucleotide diversity and Pairwise Fst values (EUT Vs GUA (blue) and AFT Vs GUA (green)) calculated for each 50-kb window around the candidate regions on chromosomes 2 and 5.