SIGNATURES OF SELECTION IN ADMIXED DAIRY CATTLE OF KENYA

H. Aliloo[1], R. Mrode[2,3], A.M. Okeyo[2] and J.P. Gibson[1]

[1]University of New England, Armidale, NSW 2351, Australia
[2]International Livestock Research Institute, Nairobi, Kenya
[3]Scotland's Rural College, Edinburgh, United Kingdom

SUMMARY

Small holder dairy farmers in Kenya rear crossbred cattle to combine the environmental adaptation features of indigenous populations with the high milk yield potential of exotic dairy breeds. The identification of signatures of selection in Kenyan admixed cattle could lead to a better understanding of the genetic structure of adaptation and productivity in challenging environmental conditions. Here, we examined the genome of the admixed cattle populations of Kenya for candidate regions under adaptive selection. We employed a haplotype based method, integrated extended haplotype homozygosity score (iHS), and scanned the genome of 1,475 admixed cattle using 521,362 SNPs. The local ancestry of the admixed cattle were inferred and used to identify the admixed cattle with more than 3 generations of crossing. This improved the power in detection of signatures of selection and after removing recently admixed animals, we identified 16 candidate regions and 8 candidate genes across 7 autosomes. Investigation of the candidate genes showed that several are involved in feed efficiency and disease resistance pathways that are important for adaptation under small-holder production systems. If substantiated, this information could be integrated into breeding programs aiming to improve dairy cattle productivity and adaptation in East Africa.

INTRODUCTION

The crossbred dairy cattle in Kenya consist of an admixed population resulting from around 50 years of crossing and inter-se matings of African indigenous cattle to several exotic dairy breeds, mainly from Friesian, Holstein, Ayrshire and related red dairy breeds, and Jersey. These animals are kept by smallholder dairy farmers, typically in herds of size 1 to 5 cows, and produce about 80% of the total milk in Kenya. The majority of Kenyan crossbred dairy cattle are bred via natural mating and only a small proportion of matings are made by AI to imported and locally bred purebred dairy bulls. Very few animals have pedigree records and there is no systematic genetic evaluation systems or breeding programs to support farmers. The identification of footprints of selection in admixed cattle through the use of molecular markers such as single nucleotide polymorphism (SNP) can lead to a better understanding of the genetic structure underlying adaptation and productivity in challenging environmental conditions. Genomic regions with selection advantage can be incorporated in breeding strategies to select animals that are well suited in such environments and production systems. In this study we scanned the genome of the Kenyan admixed cattle by applying an intra-population haplotype-based method (iHS) for signatures of post-admixture selection. We aimed to detect genomic regions responsible for adaptation and productivity under the challenging environment of East Africa. The local ancestry of individual loci are inferred to find the crossover events across the admixed genome and to assign each crossbred animal to a generation of crossing since the ancestral crossing happened.

MATERIALS AND METHODS

The genotypic data included 1,475 crossbred cattle sampled in Kenya between 2010 and 2014 and genotyped for 777,962 SNP markers using Illumina BovineHD BeadChip (Illumina, San Diego, CA). Routine QC was applied to genotypes and this resulted to 521,362 SNPs on 1,475 crossbred

animals distributed over 29 autosomes based on the UMD3.1 bovine reference genome.

Local ancestry and crossing-overs in crossbred cattle. The local ancestry of the crossbred cattle was inferred at individual SNP level using samples from 3 groups of ancestral populations including *Bos indicus* (IND) African *Bos taurus* (AFT) and European *Bos taurus* (EUT) by LAMD-LD software (Baran *et al.* 2012). The local ancestry inferences were used to calculate the average number of crossover events across each crossbred genome by first counting the number of transitions from either IND or AFT ancestry to EUT ancestry and vice versa, and then standardizing it by chromosome length. A recombination rate of 1 cM = 1 Mb across the whole genome and 1 crossover per Morgan per generation after crossing was assumed to assign each crossbred animal to an approximate generation since the ancestral crossing (indigenous × taurine) happened. A minimum of 4 generations of crossing was used to remove the impact of recent admixture on selection of signature analysis. This was also to keep only animals for which selection has had enough time to leave its footprint on their genome.

Detection of footprints of selection. The integrated extended haplotype homozygosity score (iHS) was used as an intra-population measure of the extent of haplotype homozygosity in crossbreds (Voight *et al.* 2006). We used R software *rehh* package (Gautier *et al.* 2017) to calculate iHS and then we transformed these values into *p-values* according to Gautier and Naves (2011). The *qvalue* package in R software was then used to correct *p-values* for multiple testing by calculation of a false discovery rate and generating the corresponding *q-values*. A candidate region for selection was defined by first identifying SNPs with a *q-value* <0.1 and then searching within the 500 Kb interval downstream and upstream (1 Mb window) of the identified SNP for SNPs with a *p-value* $<10^{-3}$. Genes with at least 1 SNP with a *q-value* <0.1 found within them were deemed as candidate genes under selection.

RESULTS AND DISCUSSION

The haplotypes from the 3 ancestral groups, IND, AFT and EUT, were used to infer the local ancestries of the admixed cattle at individual loci level. The majority of haplotypes in the admixed cattle were found to have originated from EUT ancestor (≈0.73) while IND and AFT ancestral populations contributed smaller proportions of admixed haplotypes (≈0.24 and ≈0.03, respectively). The local ancestry inferences were further used to calculate the genome-wide average number of crossover events on haplotypes carrying the lowest number of crossovers between the two haplotypes of each individual for each chromosome (Figure 1). For most of the admixed cattle, the number of recent crossovers per Morgan was found to be relatively small (<3). This suggested that the admixed cattle in East Africa are mainly recent crosses of indigenous cattle with exotic breeds. Only 55 animals had an average of more than 3 crossovers per Morgan, which is approximately equivalent to 4 or more generations of inter-se mating after the original cross to an exotic or indigenous ancestor (Figure 1).

Selection needs time to leave its footprints on the genome and if there is not enough time since the most recent admixture, the detection analysis is underpowered. Including recently admixed animals in the analysis adds noise to the detection of signatures of selection and potentially masks the footprints that would have otherwise been detected. We found evidence for this in our results (not shown). When we included all admixed cattle in calculation of iHS, no candidate region at a FDR threshold of 0.1 was detected. However, removing crossbreds with a genomic average crossover frequency of less than 3 per Morgan identified 16 candidate regions across 7 autosomes at the same FDR shown in Figure 2.

The details of the 16 identified candidate regions from the iHS analysis of the filtered admixed cattle are in Table 2. The size of these candidate regions ranged from only 112.25 Kb on chromosome 12 up to 683 Kb on chromosome 7 and collectively encompassed 8 candidate genes. Chromosome 7 had the highest number of candidate regions for selection among all chromosomes and it contained 3 candidate genes. Chromosome 3 contained 2 candidate genes while chromosomes 6, 11 and 12 each had one candidate gene. The ancestry of all candidate regions in chromosome 3 was dominated

by EUT while for chromosomes 6, 7 and 12 that had more than 1 candidate region, the dominant ancestry was either IND or EUT (Table 1).

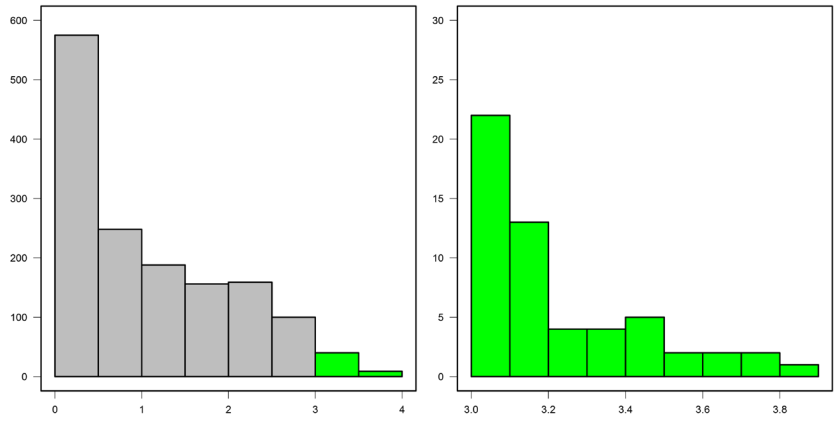

Figure 1. Average number of crossover events per Morgan in all admixed cattle (left) and in those with more than 3 crossovers per Morgan (right and green)

The *S100A10* gene is located on chromosome 3 and encodes a protein which regulates several cellular processes such as cell cycle progression and differentiation. It has been found as a candidate gene for residual feed intake in Angus cattle (Al-Husseini *et al.* 2013) through a single SNP genome-wide association study. Given that feed efficiency is a very important factor in low input smallholder production systems, it could be justified why this gene has been the target of selection in the African environment. Furthermore, the candidate region harbouring *S100A10* shows a dominant EUT ancestry in our study, suggesting possible EUT contribution to feed efficiency in the admixed cattle.

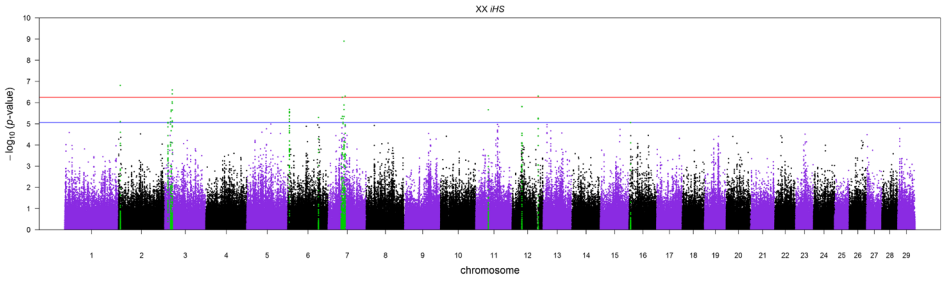

Figure 2. Manhattan plots of *p-values* for genome-wide iHS within the crossbred population. The red and blue horizontal lines correspond to false discovery rates at 5% and 10%, respectively

We identified *NLRP3* gene in a candidate region on chromosome 7 with a dominant IND ancestry. This gene encodes a pyrin-like protein and it plays a role in the regulation of inflammation, the immune response, and apoptosis. *NLRP3* has been found to be a candidate gene for Crohn's disease (Villani *et al*. 2009) and Johne's disease (Scanu *et al.* 2007; Mallikarjunappa *et al.* 2018) in human and livestock populations, respectively. The selection sweep harbouring this gene is of IND ancestry, suggesting that the IND ancestors may have contributed a version of *NLRP3* conferring resistance to local disease or other environmental challenges. Another candidate region on chromosome 7 harbours

the gene *LYPD8*, which has been reported to be differentially expressed between cows with versus without subclinical mastitis (Song *et al.* 2016) and it provides defence against gram negative bacteria in the colon of non-ruminants. This region is of EUT origin, suggesting possible EUT contribution to disease resistance in the crossbred population.

Table 1. The details of the identified candidate regions from iHS analysis

Chromosome	Region (Mb)	Top SNP *q-value*	Dominant ancestry	Candidate genes
2	5.46 – 6.00	0.0378	IND	–
3	9.58 − 9.80	0.0995	EUT	–
3	17.18 − 17.70	0.0861	EUT	–
3	18.80 − 19.29	0.0578	EUT	*S100A10*
3	22.07 − 22.71	0.0390	EUT	*ACP6*
6	4.91 − 5.29	0.0578	IND	–
6	90.70 − 91.12	0.0861	EUT	*MTHFD2L*
7	38.55 − 38.92	0.0861	IND	–
7	41.40 − 42.00	0.0390	IND	*BTNL9*, *NLRP3*
7	43.84 − 44.16	0.0861	EUT	*LYPD8*
7	46.56 − 46.99	0.0006	EUT	–
7	49.91 − 50.25	0.0390	IND	–
11	36.81 − 37.13	0.0578	IND	*ACYP2*
12	28.64 − 29.05	0.0578	IND	–
12	76.82 − 76.93	0.0390	EUT	*CLDN10*
16	4.52 − 4.89	0.0995	IND	–

CONCLUSIONS

This study provides evidence that the genome of the admixed cattle in Kenya may have been shaped by adaptive selection in response to the challenging environment in which they exist. If our findings can be substantiated, the information might be used in breeding programmes to enhance productivity and adaptation traits in smallholder dairy systems of Kenya.

REFERENCES

Al-Husseini W., Gondro C., Quinn K., Herd R.M., Gibson J.P. and Chen Y. (2014) *Anim. Genet.* **45**: 12.

Baran Y., Pasaniuc B., Sankararaman S., Torgerson D.G., Gignoux C., Eng C., Rodriguez-Cintron W,. Chapela R., Ford J.G., Avila P.C., Rodriguez-Santana J., Burchard E.G. and Halperin E. (2012) *Bioinformatics* **28**: 1359.

Gautier M. and Naves M. (2011) *Mol. Ecol.* **20**: 3128.

Gautier M., Klassmann A. and Vitalis R. (2017) *Mol. Ecol. Resour.* **17**: 78.

Mallikarjunappa S., Sargolzaei M., Brito L.F., Meade K.G., Karrow N.A. and Pant S.D. (2018) *J. Dairy Sci*. **101**: 7280.

Scanu A.M., Bull T.J., Cannas S., Sanderson J.D., Sechi L.A., Dettori G., Zanetti S. and Hermon-Taylor J. (2007) *J. Clin. Microbiol.* **12**: 3883.

Song M., He Y., Zhou H., Zhang Y., Li X. and Yu Y. (2016) *Sci. Rep.* **6**: 29390.

Villani A.C., Lemire M., Fortin G., Louis E., Silverberg M.S., Collette C., Baba N., Libioulle C., Belaiche J., Bitton A., Gaudet D., Cohen A., Langelier D., Fortin P.R., Wither J.E., Sarfati M., Rutgeerts P., Rioux J.D., Vermeire S., Hudson T.J. and Franchimont D. (2009) *Nat. Genet.* **41**: 71.

Voight B.F., Kudaravalli S., Wen X. and Pritchard J.K. (2006) *PLoS Biol.* **4**: e5350.