Genetic background of productive and adaptive features in the West African indigenous cattle breeds in Benin

S.F.U. Vanvanhossou[1,*], T. Yin[1], I.J. Giambra[1], K. Brügemann[1], L.H. Dossa[2] and S. König[1]

[1] Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany; [2] School of Science and Technics of Animal Production, Faculty of Agricultural Sciences, University of Abomey-Calavi, Abomey-Calavi, 03 BP 2819 Jéricho Cotonou, Benin; [*] Seyi.F.Vanvanhossou-2@agrar.uni-giessen.de.

Abstract

This study aimed to investigate the genetic patterns underlying productive and adaptive features in West African indigenous cattle breeds. Multi-breed genome-wide association studies for two morphometric traits (height at withers and sacrum height) were performed considering 449 animals from four indigenous cattle breeds (Lagune, Somba, Borgou and Pabli) in Benin. In addition, 67 animals were selected from the previous sample for next-generation sequencing targeting six milk protein genes (*CSN1S1, CSN2, CSN1S2, CSN3, LALBA, LGB*). A total of eight single nucleotide polymorphisms were associated with the morphometric traits. The loci and the potential candidate genes were not only involved in conformation, growth or carcass traits, but also in adaptive trait mechanisms including response to heat stress, immune functions or disease resistances. Furthermore, high genetic polymorphism was observed in the milk protein genes, and a total of 17 milk protein variants including four previously undescribed variants ($CSN3^{K}$, $LALBA^{F}$, LGB^{B1}, LGB^{K}) were identified.

Introduction

African animal genetic resources are highly diverse and represent a valuable asset for food security and the livelihood of farmers (Mwai *et al.,* 2015). Unfortunately, they are increasingly threatened by several factors including climate change, diseases and lack of adequate management policies (Mwai *et al.,* 2015). In addition, the lack of proper genetic characterization impedes their promotion, genetic amelioration and sustainable use. Recent genetic studies on some African breeds confirmed their richness and valuable features including resistance to disease, heat stress as well as their economic potential (e.g., Kim *et al.,* 2020). Novel genomic statistical methods offer new possibilities to investigate African breeds and to unravel their genetic potential.

The aim of this study was to use up-to-date genomic tools to characterise West African indigenous cattle breeds in Benin. For this purpose, morphometric traits, known as proper early indicators for animal growth, health, welfare and longevity (Kominakis *et al.,* 2017) were targeted, and their genomic architectures were investigated in genome-wide association studies (GWAS). In addition, four casein genes (*CSN1S1, CSN2, CSN1S2, CSN3*) on *Bos taurus* autosome (BTA) 6, and two whey protein genes (*LALBA* and *LGB*) on BTA5 and BTA11, respectively, were characterized, because they are associated with milk production traits in several cattle breeds (Caroli *et al.,* 2009). Genetic polymorphisms in milk proteins genes and milk protein variants were detected using next-generation sequencing.

Materials & Methods

Animal genotyping and multi-breed GWAS for morphometric traits

In total, 449 animals from four indigenous cattle breeds (Lagune, Somba, Borgou and Pabli) were sampled in Benin and genotyped using the Illumina BovineSNP50 BeadChip. SNP with

a minor allele frequency < 5 %, genotyping call rate < 90 % and which are in Hardy-Weinberg equilibrium ($p \geq 10^{-6}$), as well as animals with genotyping call rate < 95%, were excluded. The final genotype data were used to perform GWAS for the morphometric traits in PLINK (Purcell *et al.,* 2007), including height at withers (HAW) and sacrum height (SH). The genetic structure of the animal population was estimated applying a discriminant analysis of principal components. The genetic-statistical model was defined as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Wg} + \mathbf{e}$$

where **y** was a vector of observations for HAW and SH; **b** was a vector of fixed effects including age, sex, the origin (agro-ecological zone), breed, and linear discriminant functions; **g** was a vector for the SNP fixed effects; **e** was a vector of random residual effects with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$; and **X**, **W** were incidence matrices for **b** and **g**, respectively.

Significantly and suggestively associated SNP were identified using the Bonferroni-corrected significance threshold ($p = 1.55 \times 10^{-6}$) and chromosome-wide Bonferroni-corrected significance thresholds ($pc = 0.05/m_c$, with m_c = effective number of SNP for each chromosome).

Animal sequencing and detection of milk protein variants

Next-generation sequencing was performed using 67 animals selected from the previous samples. For this purpose, genotyping-in-thousands by sequencing (GT-seq) using the Illumina MiSeq V3 platform targeted the exon sequences, flanking intron sequences and parts of the 5′-upstream regions of milk protein genes (*CSN1S1, CSN2, CSN1S2, CSN3, LALBA, LGB*). The generated raw sequence data were processed to identify SNP, insertions and deletions (InDel). Variants detected in the coding regions were considered to annotate milk protein variants following the existing nomenclature (Caroli *et al.,* 2009; Gallinat *et al.,* 2013) and referring to the amino acid (AA) position in the mature proteins.

Results

Genomic regions associated morphometric traits

The multi-breed GWAS for HAW identified one significantly associated SNP (rs109126926 on BTA1, $p = 1.12 \times 10^{-6}$) and five chromosome-wide suggestively associated SNP (on BTA3, 17, 19 and 21, $p = 6.44 \times 10^{-6}$ - 5.04×10^{-5}) (Figure 1). The significantly associated SNP for HAW on BTA1 is located in the *VEPH1* gene. The SNP rs110369628 and rs109889052 on BTA19 are positioned close to each other (7.41 Mb), within the genes *PIK3R6* and *SSH2*, respectively. In addition, the SNP rs4163436 is positioned in the gene *CCDC117* on BTA17, whereas the two remaining associated SNP are not located in the vicinity of any gene. Sacrum height was associated with two suggestive SNP ($p = 2.35 \times 10^{-5}$ - 3.86×10^{-5}) and each is located in close distance to a gene (Figure 1). The SNP rs111001850 is positioned near the *LYPD8* gene on BTA7, and the SNP rs110441360 is mapped in relative proximity to the *PIK3R1* gene on BTA20.

Milk protein variants in the Beninese indigenous cattle breeds

Within the coding regions, a total of 12 DNA polymorphisms caused 17 milk protein variants. The distribution of the milk protein variants in the investigated breeds is presented in Table 1. Two casein variants were identified for *CSN1S1 (CSN1S1^{B}* and *CSN1S1^{c}).* They were due the SNP rs43703010A>G located in exon 17. The variant *CSN1S1^{B}* and *CSN1S1^{C}* were associated with the reference alleles rs43703010A and rs43703010G, respectively. For *CSN2,* the casein variants *CSN2^{A1}, CSN2^{A2}* and *CSN2^{L}*, were identified. *CSN2^{A1}* corresponds to the reference sequence and was differentiated from *CSN2^{A2}* with the SNP rs43703011A>C. *CSN2^{L}* is due to the SNP rs715383373T>C in exon 7. In the *CSN1S2* gene, only one missense variant

(rs441966828C>T) was identified, and is associated with the *CSN1S2*^{B} and *CSN1S2*^{A} protein variants. With regard to *CSN3,* three milk protein variants (*CSN3*^{A,} *CSN3*^{B}, *CSN3*^{K}) were detected. The *CSN3*^{A} and *CSN3*^{B} were associated with the SNP rs43703015T>C and rs43703016C>A in exon 4. The third detected milk protein variant was previously undescribed and was named *CSN3*^{K}, following the previous nomenclature in cattle. *CSN3*^{K} is caused by a novel SNP (BTA6:85656526C>T) in exon 4. Moreover, three known (*LALBA*^{A}, *LALBA*^{B,} *LALBA*^{E}) and one novel (*LALBA*^{F}) milk protein variants were detected in *LALBA*. *LALBA*^{A} differs from the reference variant *LALBA*^{B} with the SNP rs722550244G>C in exon 1. Near this SNP, another SNP rs7144688595C>T induces a new milk protein variant previously undescribed and therefore named *LALBA*^{F}. In addition, the variant *LALBA*^{E} due to the SNP rs465119286A>G was idenified. A missense SNP (rs109625649C>T) and a frameshift insertion induced one known (*LGB*^{B}) und two novel milk protein variants (*LGB*^{B1}, *LGB*^{K}) for *LGB*. *LGB*^{B} is associated with the reference allele rs109625649C, whereas the alternative allele rs109625649T is responsible for a novel milk protein variant called *LGB*^{B1}. *LGB*^{K} is induced by a frameshift insertion causing the emergence of a premature stop codon, and resulting in the reduction of the coding sequence of the *LGB* gene from a total of 178 to 104 AA with a complete exchange of the protein sequence from AA 92 to 104.

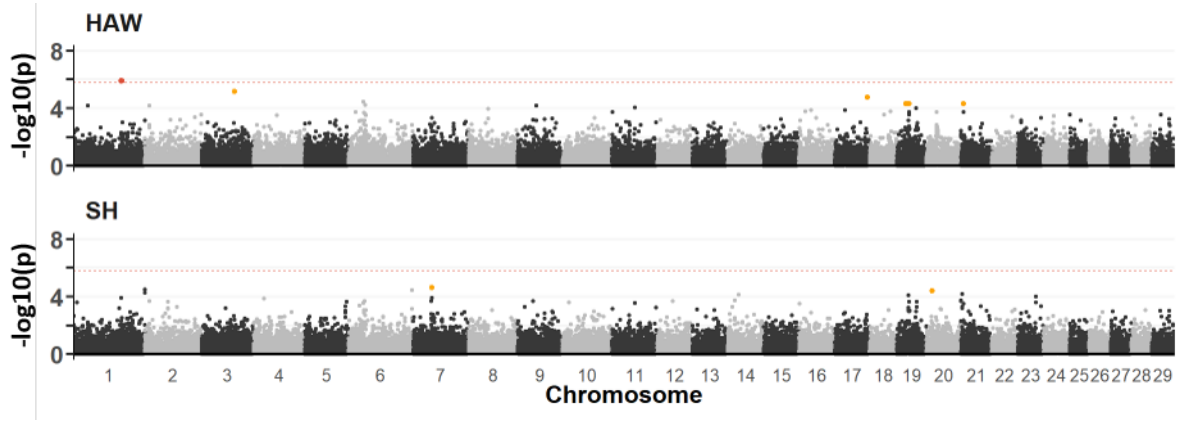

Figure 1. Manhattan Plots displaying GWAS results for height at withers and sacrum height in Beninese indigenous cattle breeds. Genowe-wide significant SNP are in red and chromosome-wide suggestive SNP are in yellow.

Table 1. Distribution of milk protein variants in the Beninese indigenous cattle breeds.

Genes	Variants [1] (Ref. / Alt.)	Protein exchange	Variant Frequency (Ref. / Alt.)			
			Cross-breed (n = 27)	Lagune (n = 20)	Somba (n = 20)	Total (n = 67)
CSN1S1	B / C	p.Glu192Gly	0.62 / 0.38	0.97 / 0.03	0.93 / 0.08	0.81 / 0.19
CSN2	A1 / A2	p.His67Pro	0.39 / 0.61	0.80 / 0.20	0.63 / 0.37	0.57 / 0.43
	A1 / L	p.Val197Ala	0.97 / 0.03	1.00 / 0.00	1.00 / 0.00	0.99 / 0.01
CSN1S2	A / B	p.Ser8Phe	0.98 / 0.02	1.00 / 0.00	1.00 / 0.00	0.99 / 0.01
CSN3	B / A	p.Ile136Thr	0.50 / 0.50	0.75 / 0.25	0.80 / 0.20	0.72 / 0.28
		p.Ala148Asp	0.50 / 0.50	0.50 / 0.50	0.80 / 0.20	0.69 / 0.31
	B / **K**	p. Ala66Val	1.00 / 0.00	0.75 / 0.25	1.00 / 0.00	0.97 / 0.03
LALBA	B / A	p. Arg10Gln	0.88 / 0.12	1.00 / 0.00	1.00 / 0.00	0.95 / 0.05
	B / E	p. Ile41Val	0.88 / 0.12	1.00 / 0.00	0.95 / 0.05	0.93 / 0.07
	B / **F**	p. Arg10Trp	0.96 / 0.04	1.00 / 0.00	0.88 / 0.13	0.94 / 0.06
LGB	B / **B1**	p. Ala118Val	0.89 / 0.11	0.50 / 0.50	0.81 / 0.19	0.80 / 0.20
	B / **K**	p.Thr92Asnfs*13	0.96 / 0.04	1.00 / 0.00	1.00 / 0.00	0.98 / 0.02

[1] The variants in bold have been newly annotated

Discussion

The rather small number of detected SNP for the moderately to highly heritable polygenic traits is in line with our small sample size and its diversity (multi-breed), and is most likely due to a lack of power to detect variants with smaller effects. However, the functional annotation of the identified genes and comparison with previous studies confirm the association of the detected loci with conformation, growth and carcass traits in cattle or in other species. For instance, the *VEPH1* gene affects rump fat thickness in Nellore cattle (Santana *et al.,* 2015). The *PIK3R6* gene was previously related to body size in sheep (Kominakis *et al.,* 2017). In addition, the *CCDC117* gene associated with HAW is implicated in feed intake and heat stress regulation in cattle (Lindholm-Perry *et al.,* 2016). Similarly, *PIK3R1* is associated with resistance to disease infections in cattle (Marino *et al.,* 2017). The observed overlap of genomic regions associated with morphometric and adaptive traits is reported in previous studies, and is in line with the selection strategies for animal adaptation using morphometric traits in African cattle herds.

The results of this study also show that the milk protein gene regions display several known and previously unknown polymorphisms in the Beninese cattle breeds (Caroli *et al.,* 2009; Gallinat *et al.,* 2013). The identified variants confirm the distribution of previously described milk protein alleles in Beninese and African cattle breeds (Moazami Goudarzi *et al.,* 2001). However, the detection of several previously unknown milk protein variants ($CSN3^{K}$, $LALBA^{F}$, LGB^{B1}, LGB^{K}) support the unicity and genetic diversity of the West African cattle breeds. The identified variants are expected to be further investigated for potential association with milk quantity and quality in the Beninese indigenous breeds. This study corroborates the necessity to enhance genetic studies in local breeds from tropical countries, contributing to a deeper understanding of genetic mechanisms for adaptation and the identification of novel genes for productive performances.

References

Caroli A.M., Chessa S., and Erhardt G.J. (2009) J. Dairy Sci. 92(11):5335-5352. https://doi.org/10.3168/jds.2009-2461

Gallinat J.L., Qanbari S., Drögemüller C., Pimentel E.C.G., Thaller G., *et al.* (2013) J. Dairy Sci. 96(1): 699–709. https://doi.org/10.3168/jds.2012-5908

Kim K., Kwon T., Dessie T., Yoo D., Mwai O.A., *et al.* (2020) Nat. Genet. 52(10), 1099–1110. https://doi.org/10.1038/s41588-020-0694-2

Kominakis A., Hager-Theodorides A.L., Zoidis E., Saridaki A., Antonakos G., *et al.* (2017) Genet. Sel. Evol. 49(1): 1–16. https://doi.org/10.1186/s12711-017-0316-3

Lindholm-Perry A.K., Butler A.R., Kern R.J., Hill R., Kuehn L.A., *et al.* (2016) Anim. Genet. 47(4): 408–427. https://doi.org/10.1111/age.12440

Marino R., Capoferri R., Panelli S., Minozzi G., Strozzi F., et *al.* (2017) Immun. 91: 259–271. http://dx.doi.org/10.1016/j.molimm.2017.08.017

Moazami Goudarzi K., Belemsaga D.M., and Ceriotti G. *et al.*, (2001) Rev. Elev. Med. Vet. Pays Trop., 54(2):129-138. https://doi.org/10.19182/remvt.9791

Mwai O., Hanotte O., Kwon Y.-J., and Cho S. (2015) Asian-australas J. Anim. Sci. 28(7), 911–921. http://dx.doi.org/10.5713/ajas.15.0002R

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R, *et al.* (2007) Am. J. Hum. Genet. 81(3): 559–575. https://doi.org/10.1086/519795

Santana M., Ventura R.V., Utsunomiya Y.T., Neves H.H.R., Alexandre P.A. *et al.* (2015) J. Anim. Breed. Genet. 132, 420–427. https://doi.org/10.1111/jbg.12167