

Detection of genomic signatures of selection in the North African Brune de l’Atlas cattle



S. Ben Jemaa¹, O. Rahal², S. B. S. Gaouar³, S. Mastrangelo⁴, E. Ciani⁵ & M. Boussaha⁶

¹ INRA-Tunisia, Tunis, Tunisia ² Centre de Recherche de Biotechnologie, Constantine, Algeria ³ Université de Tlemcen, Tlemcen, Algeria ⁴ Università degli Studi di Palermo, Palermo, Italy ⁵ Università degli Studi di Bari, Bari, Italy ⁶ GABI, INRA, AgroParisTech, Université Paris Saclay, France

Abstract

The identification of signatures of selection in indigenous livestock genome is able to elucidate the adaptive mechanisms allowing them to cope with environmental pressures. In the present study, we used genotyping data from 106 Brune de l’Atlas cattle individuals and tests based on the extent of haplotype homozygosity (EHH) to look for footprints of natural selection. We identified 17 genomic regions displaying strong signal of selection ($p\text{-value} < 10^{-6}$) overlapping with previously described QTL regions affecting many traits such as meat and milk composition, tuberculosis susceptibility, tick and mastitis resistance. Some of these highly significant candidate regions overlap with four previously described copy number variation regions (CNVRs) suggesting a possible involvement of these CNVRs in local adaptation of the Brune de l’Atlas population. Twelve candidate genes located within 25 kb of the most significant SNPs of the highly significant regions, were identified. Their physiological functions were mainly related to protein methylation, regulation of transcription, immune system homeostasis, ATP transport and regulation of GTPase activity. Our results represent the first study providing evidence for selection footprints in the North African cattle and could help to emphasize the importance to preserve local breeds to cope with challenges related to climate change.

Introduction

Local cattle in the Maghreb region (i.e. the present-day Morocco, Algeria, Tunisia, and Libya) are commonly called «Brune de l’Atlas» (BRAT) and regroup populations with different morphological attributes. For instance, in Tunisia, three ecotypes are distinguished according to coat color. These are the Blonde du Cap Bon (white), the Brune de l’Atlas Grise (Grey) and the Brune de l’Atlas Fauve (tawny). BRAT population had been subjected, for centuries, to strong environmental pressures imposing strong adaptive constraints. Accordingly, their genome has been shaped by natural selection. Understanding how natural selection has shaped genome variability represents a powerful approach to identify the mechanisms of evolutionary change and the genes that have played a major role in host survival. Here, we used information on nearly 42 000 autosomal SNPs to perform a whole genome scan for selection footprints based on the examination of extended haplotype homozygosities (EHH) within the BRAT and across BRAT and European (EUT), African (AFT) and Zebu (ZEB) populations.

Material and Methods

Samples and genotyping data

BRAT individuals (n = 106), from 12 locations in northeastern Algeria, Northern and central part of Tunisia (Figure 1) were genotyped using the BovineSNP50 BeadChip Ver. 2 (Illumina, San Diego, CA, USA). After data quality control, we combined our data set with publicly available genotyping data for 195 individuals from eight reference cattle populations : 73 African individuals from Baoule (BAO), Ndama (ND1) and Somba (SOM) populations, 69 European individuals from Aubrac (AUB), Gascon (GAS) and Limousin (LMS) breeds and 53 African zebu from two breeds: Zebu Bororo (ZBO) and Zebu Fulaini (ZFU) (Gautier et al., 2009; Matukumalli et al., 2009; Gautier et al., 2010). The final data set consisted of 42 622 common SNPs typed on 301 animals.



Figure 1. Map of the sampling areas of the Brune de l'Atlas cattle

Detection of signatures of selection

The intra-population integrated Haplotype (iHS) and inter-population Rsb scores were computed according to Voight et al. (2006) and Tang et al. (2007). in iHS computation, we defined the ancestral allele as the one present on the bovine reference genome sequence. A Z-test was applied to identify significant SNPs under selection. A $p\text{-value}$ of 10^{-3} was used as a threshold. A genomic region was considered as harboring a selection footprint if, within a 1 Mb window, it carried at least three SNPs with $p\text{-value} < 10^{-3}$. The Bovine QTL database was used to identify any overlap of the significant regions with previously published bovine QTL.

Identification and annotation of candidate genes

The candidate regions were annotated using the UCSC genome browser. A gene was considered as a candidate if the peak position was located up to 25 kb from its boundaries. A functional enrichment analysis was performed using DAVID software.

Results and Discussion

Identification of signatures of selection

Plots of the four different scores (iHS BRAT, Rsb BRAT/EUT, Rsb BRAT/AFT and Rsb BRAT/ZEB) over the bovine genome are given in Figure 2.

- Five chromosomes including 17 genomic regions were shown to display highly significant scores ($p\text{-value} < 10^{-6}$) for Rsb statistics: BTA02 and BTA14 for Rsb BRAT/EUT, BTA06 and BTA16 for Rsb BRAT/AFT and BTA05 for Rsb BRAT/ZEB (Figure 1).

- Most of these highly significant regions (16 out of 17) overlapped with previously reported regions harboring QTL that affect milk and meat composition, mastitis resistance, feed intake, average daily gain, tick resistance, gestation length, fertility, tuberculosis susceptibility and some morphological traits.

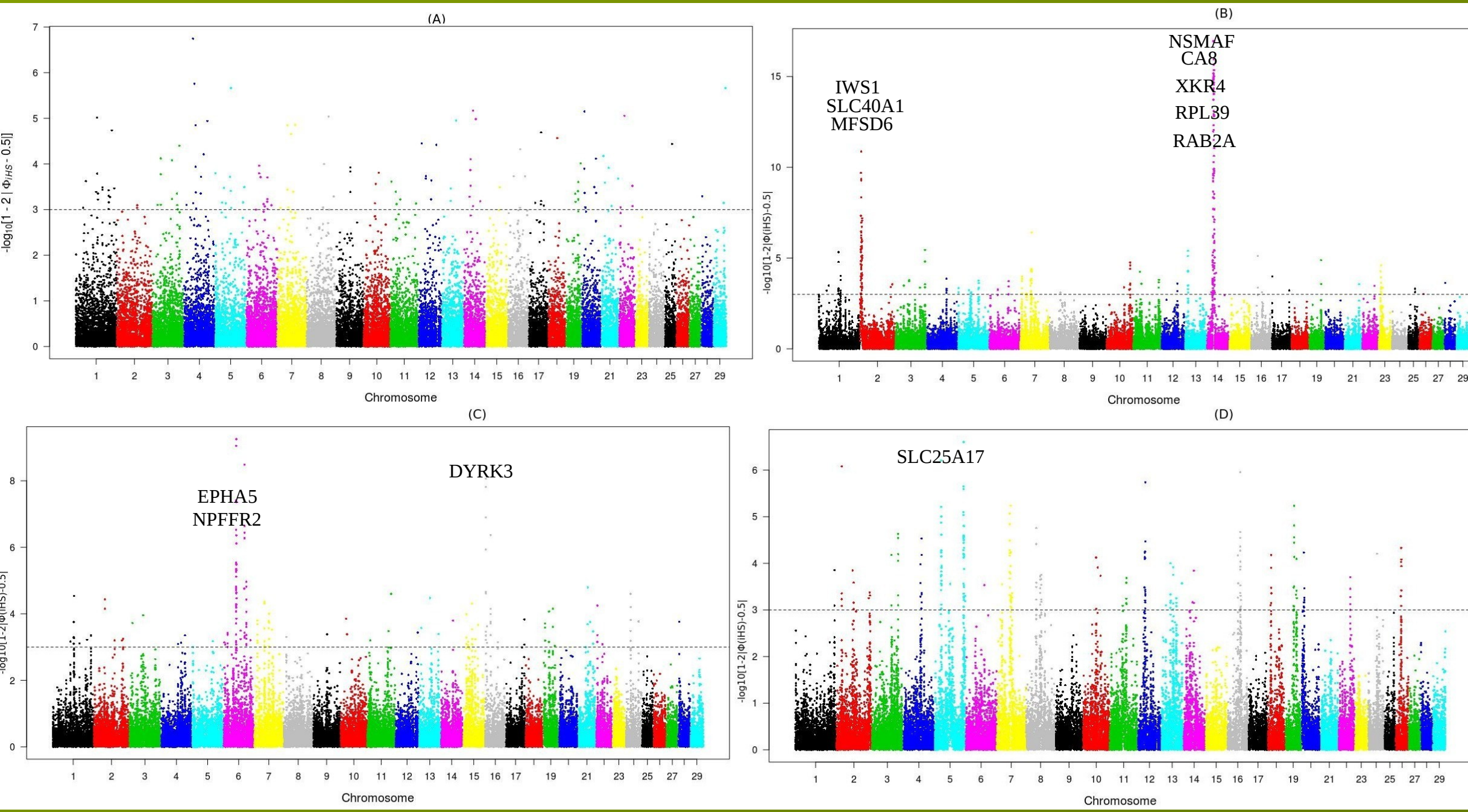


Figure 2. Manhattan plots over the genome of the Brune de l’Atlas cattle : piHS (A) and the EUTvsBRAT (B), AFTvsBRAT (C), ZEBvsBRAT (D) pRsb values. The 12 genes within 25 kb of the most significant SNPs ($p\text{-value} < 10^{-6}$) are shown.

Identification of candidate genes

- The 17 highly significant regions included 94 genes grouped into 6 enriched biological process terms including anterior/posterior pattern specification (highest enrichment score (3.21) and collagen fibril organization (Figure 3).

- 12 candidate genes involved in several biological processes such as protein methylation, regulation of transcription, immune system homeostasis, ATP transport and regulation of GTPase activity were identified within these regions (Figure 2).

- 4 CNVRs overlapped with some of the 17 highly significant regions. CNVRs near genes can result in different phenotypes through changes in the regulatory architecture that lead to alterations of gene expression levels(Spielmann & Mundlos 2013).

- Faillure to identify genes that are strongly associated with known QTL in cattle (e.g. DGAT1, GH1, GHR, ABCG2, MSTN) is mostly due to the low genome coverage of the BovineSNP50 BeadChip within these genes.

Annotation Cluster 1		Enrichment Score: 3.21		Count	P_Value	Benjamini
☐ GOTERM_BP_DIRECT	anterior/posterior pattern specification	RT		7	3.0E-6	1.5E-3
☐ INTERPRO	Homeodomain_metazoa	RT		7	6.2E-6	1.6E-3
☐ INTERPRO	Homeobox_conserved_site	RT		8	3.8E-5	4.9E-3
☐ INTERPRO	Homeobox_protein_antennapedia_type_conserved_site	RT		4	1.0E-4	8.7E-3
☐ UP_KEYWORDS	Homeobox	RT		8	1.1E-4	1.2E-2
☐ INTERPRO	Homeodomain	RT		8	1.5E-4	9.5E-3
☐ SMART	HOX	RT		8	2.5E-4	1.6E-2
☐ INTERPRO	Homeodomain-like	RT		8	8.2E-4	4.1E-2
☐ GOTERM_MF_DIRECT	sequence-specific DNA binding	RT		9	1.5E-3	2.0E-1
☐ UP_KEYWORDS	DNA-binding	RT		12	5.1E-3	1.8E-1
☐ GOTERM_MF_DIRECT	transcription factor activity_sequence-specific DNA binding	RT		7	6.7E-2	9.7E-1
☐ UP_KEYWORDS	Nucleus	RT		17	9.1E-2	9.4E-1
☐ UP_KEYWORDS	Developmental protein	RT		4	1.7E-1	9.3E-1
Annotation Cluster 2		Enrichment Score: 1.38		Count	P_Value	Benjamini
☐ GOTERM_BP_DIRECT	collagen fibril organization	RT		3	1.2E-2	8.8E-1
☐ KEGG_PATHWAY	Angiogenesis	RT		4	2.8E-2	9.6E-1
☐ UP_KEYWORDS	Secreted	RT		8	2.1E-1	9.1E-1

Figure 3. Functional annotation clustering of the 94 genes located in the highly significant regions

Conclusion

Our findings unravel, for the first time, the presence of candidate regions for signataures of selection in the genome of the North African local cattle. We identified several candidate genes involved in several important biological processes. A further network analysis looking for interactions between these candidate genes and other molecules is needed to gain a better understanding of the role of these genes in genetic and phenotypic adaptations of local cattle. We also highlight a possible involvement of some CNVRs in local adaptation of cattle through the modulation of the expression levels of individual genes. The use of higher density SNP array with uniform genome coverage may offer additional accuracy in the identification of footprints within the BRAT population.

References

- Gautier, M., L. Flori, A. Riebler, F. Jaffrézic, D. Laloé, I. Gut, K. Moazami-Goudarzi & J.-L. Foulley, 2009. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics 10: 550.
- Gautier, M., D. Laloë & K. Moazami-Goudarzi, 2010. Insights into the Genetic History of French Cattle from Dense SNP Data on 47 Worldwide Breeds. PLoS ONE 5: e13038.
- Matukumalli, L.K., C.T. Lawley, R.D. Schnabel, J.F. Taylor, M.F. Allan, M.P. Heaton, J. O’Connell, S.S. Moore, T.P.L. Smith, T.S. Sonstegard & C.P.V. Tassell, 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLOS ONE 4: e5350.
- Spielmann, M. & S. Mundlos, 2013. Structural variations, the regulatory landscape of the genome and their alteration in human disease. BioEssays News Rev. Mol. Cell. Dev. Biol. 35: 533–543.
- Tang, K., K.R. Thornton & M. Stoneking, 2007. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 5: e171.
- Voight, B.F., S. Kudaravalli, X. Wen & J.K. Pritchard, 2006. A map of recent positive selection in the human genome. PLoS Biol. 4: e72.