Detection of selection signatures in Limousin cattle using whole-genome resequencing

M. Mariadassou, Y. Ramayo-caldas, M. Charles, M. Féménia, G. Renand, D. Rocha

doi: 10.1111/age.12982 Animal Genetics, 51, 815–819

Detection of selection signatures in Limousin cattle using whole-genome resequencing

M. Mariadassou*, Y. Ramayo-Caldas[†,‡], M. Charles[†,§], M. Féménia[†], G. Renand[†] and D. Rocha[†]

*INRAE, MaIAGE, Université Paris-Saclay, Jouy-en-Josas F-78350, France. [†]INRAE, AgroParisTech, GABI, Université Paris-Saclay, Jouy-en-Josas F-78350, France. [‡]Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture, Torre Marimon, Caldes de Montbui 08140, Spain. [§]INRAE, SIGENAE, Université Paris-Saclay, Jouy-en-Josas F-78350, France.

Summary

Limousin, a renowned beef breed originating from central France, has been selectively bred over the last 100 years to improve economically important traits. We used whole-genome sequencing data from 10 unrelated Limousin bull calves to detect polymorphisms and identify regions under selection. A total of 13 943 766 variants were identified. Moreover, 311 852 bi-allelic SNPs and 92 229 indels located on autosomes were fixed for the alternative allele in all sequenced animals, including the previously reported missense deleterious F94L mutation in *MSTN*. We performed a whole-genome screen to discover genomic regions with excess homozygosity, using the pooled heterozygosity score and identified 171 different candidate selective sweeps. In total, 68 candidate genes were found in only 57 of these regions, indicating that a large fraction of the genome under selection might lie in non-coding regions and suggesting that a majority of adaptive mutations might be regulatory in nature. Many QTL were found within candidate selective sweep regions, including QTL associated with shear force or carcass weight. Among the putative selective sweeps, we located genes (*MSTN*, *NCKAP5*, *RUNX2*) that potentially contribute to important phenotypes in Limousin. Several candidate regions and genes under selection were also found in previous genome-wide selection scans performed in Limousin. In addition, we were able to pinpoint candidate causative regulatory polymorphisms in *GRIK3* and *RUNX2* that might have been under selection. Our results will contribute to improved understanding of the mechanisms and targets of artificial selection and will facilitate the interpretation of GWASs performed in Limousin.

Keywords Limousin breed, polymorphisms, signature of selection, whole-genome resequencing

Human-driven selection has resulted, since the domestication of cattle approximately 10 000 years ago (Loftus *et al.* 1994), in the creation of more than 1000 breeds exhibiting differences in appearance, adaptation and productivity. This artificial selection has induced specific changes in the patterns of DNA variation. Detection of these selection signatures could be used to reveal mutations responsible for improved traits. Several genome-wide scans have been performed to detect selection signatures in cattle (Gutiérrez-Gil *et al.* 2015). Most studies have used SNP genotypes to discover regions under selection and often have only allowed the identification of candidate genes that might be involved in the genetic variability of the traits under selection. More recently, genome-wide scans for selection signatures using WGSs have been performed in cattle, allowing a nearly complete interrogation of genome-wide polymorphisms and sometimes even pinpointing underlying candidate causal mutations (Xu *et al.* 2015). To date, several genome-wide selection studies have been carried out in Limousin (Druet *et al.* 2013); however, these studies have used only SNP genotypes and no candidate causal mutations have been identified.

We present here the first WGS-based selective sweep scan performed in Limousin. In this study, we used whole-genome sequencing data recently generated from 10 unrelated Limousin bull calves and then used the discovered

Address for correspondence

D. Rocha, INRAE, AgroParisTech, GABI, Université Paris-Saclay, Jouy-en-Josas F-78350, France.
E-mail: dominique.rocha@inrae.fr

Accepted for publication 16 June 2020

polymorphisms to identify regions under selection (Guillocheau *et al.* 2019). A total of approximately 3.3 billion reads were mapped to the bovine reference sequence assembly (UMD 3.1) with an average of approximately 13-fold coverage per animal (Table S1). A total of 13 943 766 variants, including 11 943 766 bi-allelic SNPs, were identified. As all sequenced animals were previously genotyped with Illumina BovineSNP50 BeadChips (Allais *et al.* 2014), it was possible to analyse SNP calling accuracy by comparing WGS variant genotypes and microarray-derived genotypes. The percentage of concordance was calculated for >19 600 common SNPs (according to their genomic positions) and considering read depth (>2 reads). The genotype concordance ranged from 98.26 to 99.15% (Table S1).

Interestingly 311 852 bi-allelic SNPs and 92 229 indels located on autosomes were fixed for the alternative allele in all sequenced animals. We focused on autosomal polymorphisms as all animals were males and to avoid false polymorphisms located on the pseudo-autosomal part of the sexual chromosomes. Polymorphisms were annotated using Variant Effect Predictor (McLaren *et al.* 2010) and deleterious variants were predicted with sift (Ng & Henikoff 2001). We identified 27 missense deleterious bi-allelic autosomal SNP fixed polymorphisms among all samples (Table S2), including the known F94L mutation (rs110065568: BTA2 g.6279278C>A) in *myostatin* (*MSTN*), a known muscle growth factor inhibitor. Mutations in this gene are responsible for the double-muscling phenotype in cattle (Grobet *et al.* 1997; Kambadur *et al.* 1997; McPherron & Lee 1997). Several studies have shown that the F94L variation is associated with increased muscle mass, carcass yield and meat tenderness in Limousin and Limousin-cross cattle (Sellick *et al.* 2007; Esmailizadeh *et al.* 2008; Alexander *et al.* 2009). The high frequency of the mutant allele in Limousin most likely reflects the effects of selection for increased muscle mass.

To detect putative selective signatures, we performed a whole-genome screen to identify genomic regions with excess homozygosity, according to a previously described method (Rubin *et al.* 2010). A total of 11 475 036 autosomal polymorphic bi-allelic SNPs were used to calculate the Z transformation of the pooled heterozygosity (ZHp). ZHp scores were calculated in 40 kb windows sliding in 20 kb steps and we removed windows with fewer than 10 SNPs (Fig. 1). The average value and standard deviation observed for the Hp score distribution were 0.11 and 0.02. Extremely low ZHp scores indicate putative selective sweeps because of excess homozygosity. The major challenge of such analyses is the difficulty of setting a threshold to distinguish selection from drift. We used the method described by Rubin *et al.* and chose the *MSTN* locus as a known selective sweep to select the ZHp score threshold ($ZHp \leq -3$, corresponding to the bottom 0.35% of all windows). A total of 440 windows passed this threshold. We then merged overlapping or adjacent windows, into 171 different genomic regions and recomputed a new ZHp score for each such region (Table S3). Overall 68 candidate genes were identified in 57 of these regions.

To evaluate whether these genes were significantly enriched for specific gene ontology terms and KEGG pathways, gene enrichment analyses were conducted using *ClueGO* (Bindea *et al.* 2009). Thirty-five GO terms were significantly enriched (*P*-value corrected with Benjamini-Hochberg <5%, Table S4). Interestingly, the GO terms with the highest percentage of associated genes corresponded to biological processes related to embryonic body morphogenesis and chondrocyte development.

The positions of the putative selective sweep regions were compared with the location of known bovine QTL deposited in the database AnimalQTLdb (Hu *et al.* 2007). A total od 440 different QTL were found within 101 candidate selective sweep regions (Table S5). These results indicate that many of the candidate selective sweep regions highlight loci related to traits of interest, including several meat-related phenotypes.

Among the putative selective sweeps, we located several loci that potentially contribute to important phenotypes in Limousin. For example, a putative selective sweep was observed in a region with an extremely low ZHp score of −4.16 (region 2.4 in BTA2), which included *MSTN*. This region and adjacent regions (regions 2.2–2.11) are located within QTL found in Limousin associated with carcass fat and carcass conformation (Purfield *et al.* 2019). Furthermore, region 2.16 (ZHp = −3.17) contains *NCKAP5*, a gene associated in Nellore with flight speed, a temperament trait (Valente *et al.* 2016). This gene has been associated with human attention deficit hyperactivity disorders (Lasky-Su *et al.* 2008), schizophrenia and bipolar disorders (Wang *et al.* 2010). Interestingly, since the early 1990s, selection for improving temperament has been integrated among the main objectives for Limousin cattle (Boivin *et al.* 1992). This information pinpoints *NCKAP5* as a candidate gene under selection in Limousin. Moreover, region 23.4 (ZHp = −3.05) contains *RUNX2*, a gene encoding a transcription factor essential for osteoblast differentiation and chondrocyte maturation (Komori 2018). The knockout of this gene, also known as *CBFA1*, in mice results in skeletal abnormalities and *Cbfa1*-null mutant embryos are smaller and their weight is reduced, compared with normal littermates (Otto *et al.* 1997). Interestingly this gene has also been associated with carcass- and growth-related traits in broilers (Grupioni *et al.* 2017) and goats (Jiang *et al.* 2019). It might therefore be possible that *RUNX2* has been the target of selection.

Comparison of the results of our scan with results of previous genome-wide bovine selection signature studies performed in Limousin revealed some overlap, including near *MSTN* (Table S3). We also found regions under selection not previously described in Limousin.

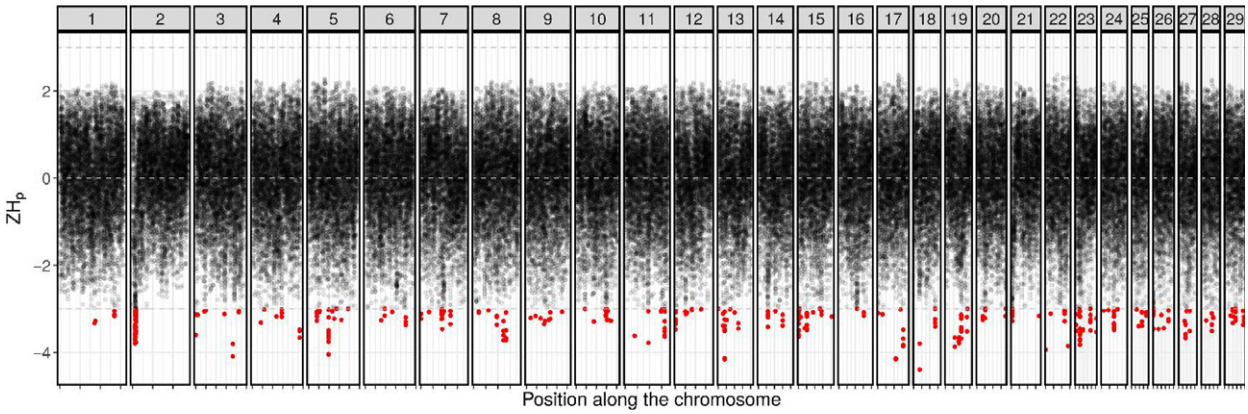

Figure 1 Manhattan plot showing the distribution of pooled heterozygosity (ZHp) scores. Points with ZHp ≤−3 are highlighted in red.

Interestingly, region 3.8 (ZHp = −3.00) contains *GRIK3*, a gene encoding glutamate ionotropic receptor kainate type subunit 3 and found in a selection scan performed in Lidia breed, a bovine breed known for its aggressive behaviour (Eusebi *et al.* 2018). *GRIK3* is highly expressed in the central nervous system and has been associated with personality traits in human (Minelli *et al.* 2009). Like *NCKAP5*, *GRIK3* might have been under selection in Limousin. We also found one region previously identified in genome-wide whole-genome sequence-based selection scans performed in other cattle breeds. Indeed, region 16.6 (containing *RERE*) was found by Boitard *et al.* (2016) in four cattle breeds. Observing this overlap among different breeds, including Limousin, suggests that this region is truly under selection.

Genome-wide scans for selection signatures using whole-genome sequences can potentially pinpoint underlying candidate causal mutations. Interestingly, 114 of our 171 candidate selective sweep regions do not contain any gene. This is the case for the region with the lowest ZHp score (region 2.9, ZHp = −4.26). Moreover, six out of the 10 regions with the lowest ZHp score are intergenic. This shows that a large fraction of the genome under selection might lie in non-coding regions, suggesting that a majority of adaptive mutations might be regulatory in nature. However, it is also possible that some of these 114 regions contain genes encoding long non-coding RNAs, yet to be discovered. A better annotation of the cattle genome is therefore needed to identify the polymorphisms within these regions directly targeted by artificial selection. On another hand, a detailed analysis of polymorphisms located within candidate genes suggests putative causative polymorphisms. Most SNPs are located upstream of genes and we therefore investigated if some of those SNPs might have regulatory effects. The identification of SNPs potentially altering transcription factor binding sites was predicted *in silico* with a custom script using transcription factor binding site models from the JASPAR (Sandeli *et al.* 2004), HOCOMOCO (Khamis *et al.* 2018) and TRANSFAC (Knüppel *et al.* 1994) databases. We found 11 and five putative regulatory SNPs upstream of *GRIK3* and *RUNX2* respectively (Table S6). It might be possible that some of these putative regulatory SNPs have been selected in Limousin. Further work on these suggested causative mutations is required to support their involvement in phenotypes of interest.

In conclusion, we identified 171 candidate selective sweeps in Limousin. We found that most of our strongest signals are in non-coding regions of the genome, suggesting that several mutations could potentially be regulatory. In addition, we pinpointed some candidate regulatory polymorphisms that might be under selection. We have probably missed other regions under selection because of limitations in sample size and the power of the study owing to the use of only one population and one specific test in scanning for signals of selection. In addition, some of the identified regions could be false-positive discoveries. However, our results will contribute to improved understanding of the mechanisms and targets of artificial selection and will facilitate the interpretation of GWASs performed in Limousin.

Acknowledgements

This work was funded by INRAE, Agence Nationale de la Recherche (contracts ANR-05-GANI-005, ANR-05-GANI-017-01 and ANR-09-GENM-011) and APIS-GENE (contract 01-2005-QualviGenA-02). YRC was funded by Marie Skłodowska-Curie grant (P-Sphere) agreement no. 6655919 (EU). We are grateful to the Genotoul bioinformatics facility for providing computing and storage resources and we would like to thank Coline Billerey and Mekki Boussaha for their help. We would like to thank the two anonymous reviewers for their help in improving our manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability

The WGSs and the VCF file containing variant information for the 10 sequenced Limousin animals are not publicly available because the sequenced animals belong to commercial breeding companies, but are available from the corresponding author on reasonable request.

References

Alexander L.J., Kuehn L.A., Smith T.P., Matukumalli L.K., Mote B. & Koltes J.E. (2009) A Limousin specific myostatin allele affects *longissimus* muscle area and fatty acid profiles in a Wagyu-Limousin F2 population. *Journal of Animal Sciences* **87**, 1576–81.

Allais S., Levéziel H., Hocquette J.F., Rousset S., Denoyelle C., Journaux L. & Renand G. (2014) Fine mapping of quantitative trait loci underlying sensory meat quality traits in three French beef cattle breeds. *Journal of Animal Sciences* **92**, 4329–41.

Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., Kirilovsky A., Fridman W.H., Pagès F., Trajanoski Z. & Galon J. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–3.

Boitard S., Boussaha M., Capitan A., Rocha D. & Servin B. (2016) Uncovering adaptation from sequence data: lessons from genome resequencing of four cattle breeds. *Genetics* **203**, 433–50.

Boivin X., Le Neindre P., Chupin J.M., Garel J.P. & Trillat G. (1992) Influence of breed and early management on ease of handling and open-field behaviour of cattle. *Applied Animal Behavioural Science* **20**, 259–77.

Druet T., Pérez-Pardal L., Charlier C. & Gautier M. (2013) Identification of large selective sweeps associated with major genes in cattle. *Animal Genetics* **44**, 758–62.

Esmailizadeh A.K., Bottema C.D., Sellick G.S., Verbyla A.P., Morris C.A. & Cullen N.G. (2008) Effects of the myostatin F94L substitution on beef traits. *Journal of Animal Sciences* **86**, 1038–46.

Eusebi P.G., Cortés O., Carleos C., Dunner S. & Cañon J. (2018) Detection of selection signatures for agonistic behaviour in cattle. *Journal of Animal Breeding and Genetics* **135**, 170–7.

Grobet L., Martin L.J., Poncelet D., Pirottin D., Brouwers B. & Riquet J. (1997) A deletion in the bovine *myostatin* gene causes the double-muscled phenotype in cattle. *Nature Genetics* **17**, 71–4.

Grupioni N.V., Stafuzza N.B., Carvajal A.B., Ibelli A.M., Peixoto J.O., Ledur M.C. & Munari D.P. (2017) Association of *RUNX2* and *TNFSF11* genes with production traits in a paternal broiler line. *Genetics and Molecular Research* **16**.

Guillocheau G.M., El Hou A., Meersseman C. *et al.* (2019) Survey of allele specific expression in bovine muscle. *Scientific Reports* **9**, 4297.

Gurgul A., Szmatoła T., Ropka-Molik K., Jasielczuk I., Pawlina K., Semik E. & Bugno-Poniewierska M. (2016) Identification of genome-wide selection signatures in the Limousin beef cattle breed. *Journal of Animal Breeding and Genetics* **133**, 264–76.

Gutiérrez-Gil B., Arranz J.J. & Wiener P. (2015) An interpretive review of selective sweep studies in *Bos taurus* cattle populations: identification of unique and shared selection signals across breeds. *Frontiers in Genetics* **6**, 167.

Hu Z.-L., Fritz E.R. & Reecy J.M. (2007) AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Research* **35**, D604–9.

Jiang E., Kang Z., Wang X., Liu Y., Liu X., Wang Z., Li X. & Lan X. (2019) Detection of insertions/deletions (InDels) within the goat *Runx2* gene and their association with litter size and growth traits. *Animal Biotechnology* **8**, 1–9.

Kambadur R., Sharma M., Smith T.P. & Bass J.J. (1997) Mutations in *myostatin* (*GDF8*) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Research* **7**, 910–6.

Kemper K.E., Saxton S.J., Bolormaa S., Hayes B.J. & Goddard M.E. (2014) Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics* **15**, 246.

Khamis A.M., Motwalli O., Oliva R., Jankovic B.R., Medvedeva Y.A., Ashoor H., Essack M., Gao X. & Bajic V.B. (2018) A novel method for improved accuracy of transcription factor binding site prediction. *Nucleic Acids Research* **46**, e72.

Knüppel R., Dietze P., Lehnberg W., Frech K. & Wingender E. (1994) TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *Journal of Computational Biology* **1**, 191–8.

Komori T. (2018) Runx2, an inducer of osteoblast and chondrocyte differentiation. *Histochemistry and Cell Biology* **149**, 313–23.

Lasky-Su J., Neale B.M., Franke B. *et al.* (2008) Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *American Journal of Medical Genetics Part B Neuropsychiatric Genetics* **147B**, 1345–54.

Loftus R.T., MacHugh D.E., Bradley D.G., Sharp P.M. & Cunningham P. (1994) Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences USA* **91**, 2757–61.

McLaren W., Pritchard B., Rios D., Chen Y., Flicek P. & Cunningham F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–70.

McPherron A.C. & Lee S.J. (1997) Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Sciences USA* **94**, 12457–61.

Minelli A., Scassellati C., Bonvicin C., Perez J. & Gennarelli M. (2009) An association of *GRIK3* Ser310Ala functional polymorphism with personality traits. *Neuropsychobiology* **59**, 28–33.

Ng P.C. & Henikoff S. (2001) Predicting deleterious amino acid substitutions. *Genome Research* **11**, 863–74.

Otto F., Thornell A.P., Crompton T. *et al.* (1997) *Cbfa1*, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89**, 765–71.

Purfield D.C., Evans R.D. & Berry D.P. (2019) Reaffirmation of known major genes and the identification of novel candidate genes associated with carcass-related metrics based on whole genome sequence within a large multi-breed cattle population. *BMC Genomics* **20**, 720.

Ramey H.R., Decker J.E., McKay S.D., Rolf M.M., Schnabel R.D. & Taylor J.F. (2013) Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* **14**, 382.

Rubin C.J., Zody M.C., Eriksson J. *et al.* (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–91.

Sandeli A., Alkema W., Engström P., Wasserman W.W. & Lenhard B. (2004) JASPAR: an open access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research* **32**, D91–4.

Sellick G.S., Pitchford W.S., Morris C.A., Cullen N.G., Crawford A.M. & Raadsma H.W. (2007) Effect of myostatin F94L on carcass yield in cattle. *Animal Genetics* **38**, 440–6.

Shen J., Hanif Q., Cao Y., Yu Y., Lei C., Zhang G. & Zhao Y. (2020) Whole genome scan and selection signatures for climate adaptation in Yanbian cattle. *Frontiers in Genetics* **11**, 94.

Szmatoła T., Gurgul A., Jasielczuk I., Ząbek T., Ropka-Molik K., Litwińczuk Z. & Bugno-Poniewierska M. (2019) A comprehensivce analysis of runs of homozygosity of eleven cattle breeds representing different production types. *Animals* **9**, 1024.

The Bovine HapMap Consortium, Gibbs R.A., Taylor J.F. *et al.* (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–32.

Valente T.S., Baldi F., Sant'Anna A.C., Albuquerque L.G. & Paranhos da Costa M.J. (2016) Genome-wide association study between single nucleotide polymorphisms and flight speed in Nellore cattle. *PLoS ONE* **11**, e0156956.

Wang K.S., Liu X.F. & Aragam N. (2010) A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophrenia Research* **124**, 192–9.

Xu L., Bickhart D.M., Cole J.B., Schroeder S.G., Song J., Tassell C.P., Sonstegard T.S. & Liu G.E. (2015) Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Molecular Biology and Evolution* **32**, 711–25.

Zwane A.A., Schnabel R.D., Hoff J., Choudhury A., Makgahlela M.L., Maiwashe A., Van Marle-Koster E. & Taylor J.F. (2019) Genome-wide SNP discovery in indigenous cattle breeds of South Africa. *Frontiers in Genetics* **10**, 273.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Summary of sequencing results.

Table S2. List of the fixed non-synonymous deleterious SNPs.

Table S3. List of the candidate selective sweep regions with their gene content.

Table S4. List of enriched biological processes.

Table S5. List of the QTL overlapping with candidate selective sweep regions.

Table S6. List of putative regulatory SNPs found in *GRIK3* and *RUNX2*.