


Communication

Copy Number Variants in Two Northernmost Cattle Breeds Are Related to Their Adaptive Phenotypes

Laura Buggiotti ¹ , Nikolay S. Yudin ^{2,3} and Denis M. Larkin ^{1,*}¹ Royal Veterinary College, University of London, London NW1 0TU, UK² The Federal State Budgetary Institution of Science Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (ICG SB RAS), Novosibirsk 630090, Russia³ Kurchatov Genomics Center, the Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Science (ICG SB RAS), Novosibirsk 630090, Russia

* Correspondence: dlarkin@rvc.ac.uk

Abstract: Copy number variations (CNVs) are genomic structural variants with potential functional and evolutionary effects on phenotypes. In this study, we report the identification and characterization of CNVs from the whole-genome resequencing data of two northernmost cattle breeds from Russia: the Yakut and Kholmogory cattle and their phylogenetically most related breeds, Hanwoo and Holstein, respectively. Comparisons of the CNV regions (CNVRs) among the breeds led to the identification of breed-specific CNVRs shared by cold-adapted Kholmogory and Yakut cattle. An investigation of their overlap with genes, regulatory domains, conserved non-coding elements (CNEs), enhancers, and quantitative trait loci (QTLs) was performed to further explore breed-specific biology and adaptations. We found CNVRs enriched for gene ontology terms related to adaptation to environments in both the Kholmogory and Yakut breeds and related to thermoregulation specifically in Yakut cattle. Interestingly, the latter has also been supported when exploring the enrichment of breed-specific CNVRs in the regulatory domains and enhancers, CNEs, and QTLs implying the potential contribution of CNVR to the Yakut and Kholmogory cattle breeds' adaptation to a harsh environment.

Keywords: CNV; cattle; cold adaptation

Citation: Buggiotti, L.; Yudin, N.S.; Larkin, D.M. Copy Number Variants in Two Northernmost Cattle Breeds Are Related to Their Adaptive Phenotypes. *Genes* **2022**, *13*, 1595. <https://doi.org/10.3390/genes13091595>

Academic Editor: Nico M. Van Straalen

Received: 29 July 2022

Accepted: 2 September 2022

Published: 6 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Copy number variations (CNV) refer to a structural variation type where DNA segments of >1 kb are present in individual genomes in varying copy numbers, compared to a reference genome [1]. CNVs are less frequent than single nucleotide polymorphisms (SNPs) and other variations. However, they can potentially have a larger functional and evolutionary impact, such as changing gene structure and dosage, altering gene regulation, and exposing recessive alleles [2,3]. CNVs and their impacts have been extensively studied, especially in humans [4], where CNVs are considered to affect gene expression and, therefore, some phenotypes of interest. For example, CNV loss in the *NPY4R* gene was associated with obesity [5]. Other studies have revealed that genomic diversity could be increased due to the differential selection of CNVs for adaptations to different environments [6–8]. For instance, 30% of young fast evolving duplicated genes in sticklebacks are in CNVs and these genes are enriched in functional categories related to environmental adaptations [8]. Studies in livestock also highlight the role of CNVs in shaping various phenotypes. A partial or complete duplication of the *KIT* gene causes different patterns of white coat coloration in pigs and in some of the cattle breeds [9,10], while a white coat color in sheep has been associated with a duplication of the *ASIP* gene [11].

The present study focuses on CNV detection from the whole-genome resequencing data of two cold-adapted cattle breeds from Russia: the Kholmogory and Yakut cattle. Both breeds live in harsh environments but have very different population histories. The

Kholmogory was formed in the European part of Russia around 300 years ago [12], while the Yakut cattle was formed at the Baikal area of Siberia and likely migrated together with the Yakut people to contemporary Yakutia about 500–800 years ago [13]. To identify CNVs and CNV regions (CNVRs) which could co-evolve with the adaptation of the Russian cattle breeds to harsh climate conditions, we utilized sequences of four breeds (Yakut, Kholmogory, Holstein, and Hanwoo). Previously, the Yakut cattle has been found to be related to Korean Hanwoo, and Kholmogory to Holstein [14]. Therefore, we compared Kholmogory and Yakut cattle to phylogenetically close breeds to search for CNVs that could have an influence on breed-specific biology and adaptations.

2. Materials and Methods

The Yakut and Kholmogory cattle breeds were previously whole-genome resequenced [15] and mapped to the reference Hereford cattle assembly (UMD3.1, BosTau6) using BWA-MEM [16] with default parameters; Hanwoo and Holstein resequencing data were downloaded from the Sequence Repository Archive [17]. A total of 98 high-quality samples of the four cattle breeds (Yakut (29), Kholmogory (32), Hanwoo (19), and Holstein (18)) were used. The cn.MOPS R package (copy number estimation by a Mixture of Poissons [18]) was used for CNV detection. Based on the average sequence coverage of our data (~11X), window length was set to 700 ($\text{windowLength} = \text{readLength} \times 50 / \text{coverage}$) and posterior probabilities were estimated (posteriorProbs). The cn.MOPS tool represents a CNV detection pipeline that models the depths of coverage across multiple samples at each genomic position. Using a Bayesian approach, it decomposes read variants across samples into integer CNVs and noise using mixture components and Poisson distributions, respectively. The multiple samples approach increases statistical power and decreases computational burden and the FDR in CNV detection. CNVs were then used to construct a set of copy number variable regions (CNVRs) for Kholmogory, Yakut, Hanwoo, and Holstein breeds. The CNVRs were constructed by merging CNVs across samples of the same breed that exhibited at least 50% pairwise reciprocal overlap in their genomic coordinates; unique CNVRs per breed were those with less than 10% overlap with CNVRs in other breeds. BEDTools [19] and BEDOPS [20] tools were used to calculate CNVR overlaps. Genomic Regions Enrichment of Annotations Tool (GREAT), [21] was used to assign each gene of the reference Hereford cattle assembly (UMD3.1, BosTau6) to a regulatory domain consisting of a basal domain that extends 5 kb upstream and 1 kb downstream from its transcription start site (total length of the regions per gene is 6000 bp). The karyoploteR [22] package was used to plot the cattle chromosome map and to visualize the CNVRs locations for the four breeds.

3. Results and Discussion

A total of 860,380 autosomal CNVs were detected in the four-breed set, which were then merged into 71,549 CNVRs. Interestingly, the Yakut and Kholmogory breeds shared the largest fraction of CNVRs (138.41 Mb). The second largest shared fraction of Kholmogory CNVRs was with Holstein (61.05 Mbp) and Yakut CNVRs with Hanwoo (27.92 Mbp), confirming known breed relations. A total of 19,502 CNVRs (total length: 106.09 Mbp) were breed-specific for the Yakut, 2238 (18.61 Mbp) for the Kholmogory, 2535 (8.27 Mbp) for the Hanwoo and 1625 (4.95 Mbp) for the Holstein cattle (Figure 1).

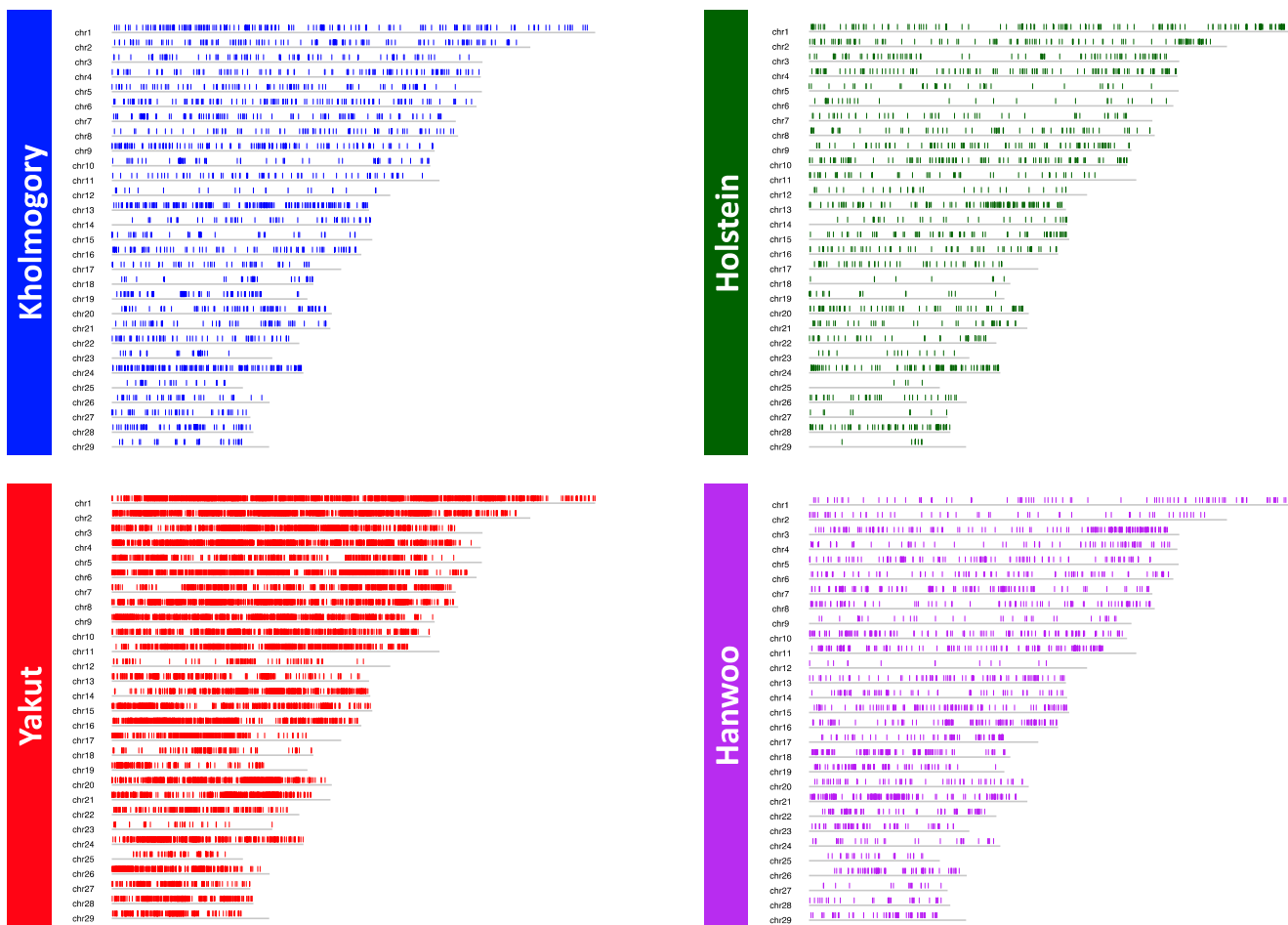


Figure 1. Whole genome distribution of breed-specific CNVRs for the four breeds.

To reveal a possible contribution of CNVRs to breed-specific biology and adaptations, we investigated 5522 genes found in the four breed-specific CNVRs, of which 2962, 1034, 862, and 664 genes were found in the Yakut, Hanwoo, Kholmogory, and Holstein CNVRs, respectively (Table S1). A gene ontology (GO) enrichment analysis highlighted distinct pathways being enriched in these gene sets, with the second largest number (35) found in the Yakut cattle, among which we observed cognition, the regulation of small GTPase-mediated signal transduction, the detection of mechanical stimulus, etc. (Table S2a). The Kholmogory cattle had the third largest number of GO categories (22) of which intracellular signal transduction, GTPase regulator activity, adenyly nucleotide binding, etc. were uniquely present. The Hanwoo had the largest number of GO categories (38), while Holstein had the same number of GO categories as Kholmogory (22). Interestingly, all the breeds showed enrichment for the GO category response to stimulus, although the genes were different. Moreover, a DAVID functional annotation cluster analysis [23] highlighted the enrichment of ubiquitin protein in the Yakut cattle, which is involved in protein degradation and found to be enriched in Antarctic fish [24]. The authors hypothesized that the cost of living for cold-adapted ectotherms commits more effort to maintaining protein homeostasis. Moreover, ATPase activity, microtubule motor activity, and blood coagulation inhibitor were also found to be enriched in the Yakut cattle CNVRs, which could potentially influence thermoregulation (Table S2b). There were 7414 CNVRs shared between the Yakut and Kholmogory breeds, overlapping 2925 genes, which were enriched in various pathways such as keratin filament, kinase, oxytocin signaling pathway, etc. (Table S3a,b).

Yakut cattle-specific CNVRs involved multiple fatty-acid related genes such as the *CYP4A11* (cytochrome P-450 4A11), which is implicated in lipogenesis and growth traits.

This CNVR has previously been reported in various Chinese native taurine cattle breeds (Jaxian, Quinchuan, Nanyang, Jinnan, Luxi, and Chinese Red Steppe [25]). The Kholmogory cattle unique CNVRs covered over one-thousand annotated genes, among which a few were interleukin genes (*IL17RE*, *IL20RA*, *IL10RA*, *IL10*) associated with immune response, as well as *PLA2G4A*, which is involved in inflammatory responses [26]. To further investigate the potential association of CNVRs with gene regulation we investigated the landscape of breed-specific CNVRs overlapping regulatory domains by using the GREAT approach [21]. We found 1109, 663, 659, and 516 genes with breed-specific CNVRs in their GREAT domains in Yakut, Holstein, Hanwoo, and Kholmogory cattle breeds, respectively. The GO enrichment analysis of the latter sets of genes found the response to stimulus to be enriched in the Yakut but not in the Kholmogory cattle (Table S4a), while a DAVID functional annotation cluster analysis revealed the term lipid transport and lipoprotein metabolic process to be unique to Yakut cattle when considering only breed-specific CNVRs in GREAT domains (Table S4b), suggesting a potential fundamental role of CNVRs in regulating thermogenesis [27], a major adaptive mechanism to extreme climates. Moreover, the top DAVID functional cluster (Table S4b) contained genes that are involved in cytoskeletal reorganization (such as *PADI*- genes), which also plays a role in shaping the adaptation to a cold environment [28]. Kholmogory, on the other hand, had the unique GO terms: metal ion transport, cellular response to organic substance and transporter activity, among others, suggesting again their potential role in the adaptation to new ecological niches [29,30]. Finally, we investigated the enrichment of breed-specific CNVRs in three functional classes: (i) conserved non-coding elements (CNEs); (ii) enhancers; (iii) QTLs regions, by using GAT [31]. Interestingly, breed-specific CNVRs were enriched with CNEs [32] in all four breeds, highlighting again their potential involvement in gene regulation, and Yakut cattle-specific CNVRs were significantly negatively enriched of both cetartiodactyla and ruminant CNEs (Figure 2). The QTLs enrichment analysis suggests that Yakut cattle-specific CNVRs are significantly negatively enriched in all major QTLs categories but meat and carcass, while CNVs in the other three breeds show positive or negative associations with a limited number of QTL categories (Figure 2). One explanation for the QTLs enrichment results is that in a natural population subjected to a harsh climate such as Yakut cattle, selection acts on CNVRs contributing to phenotypes that are quite different from those which are normally under selection by humans.

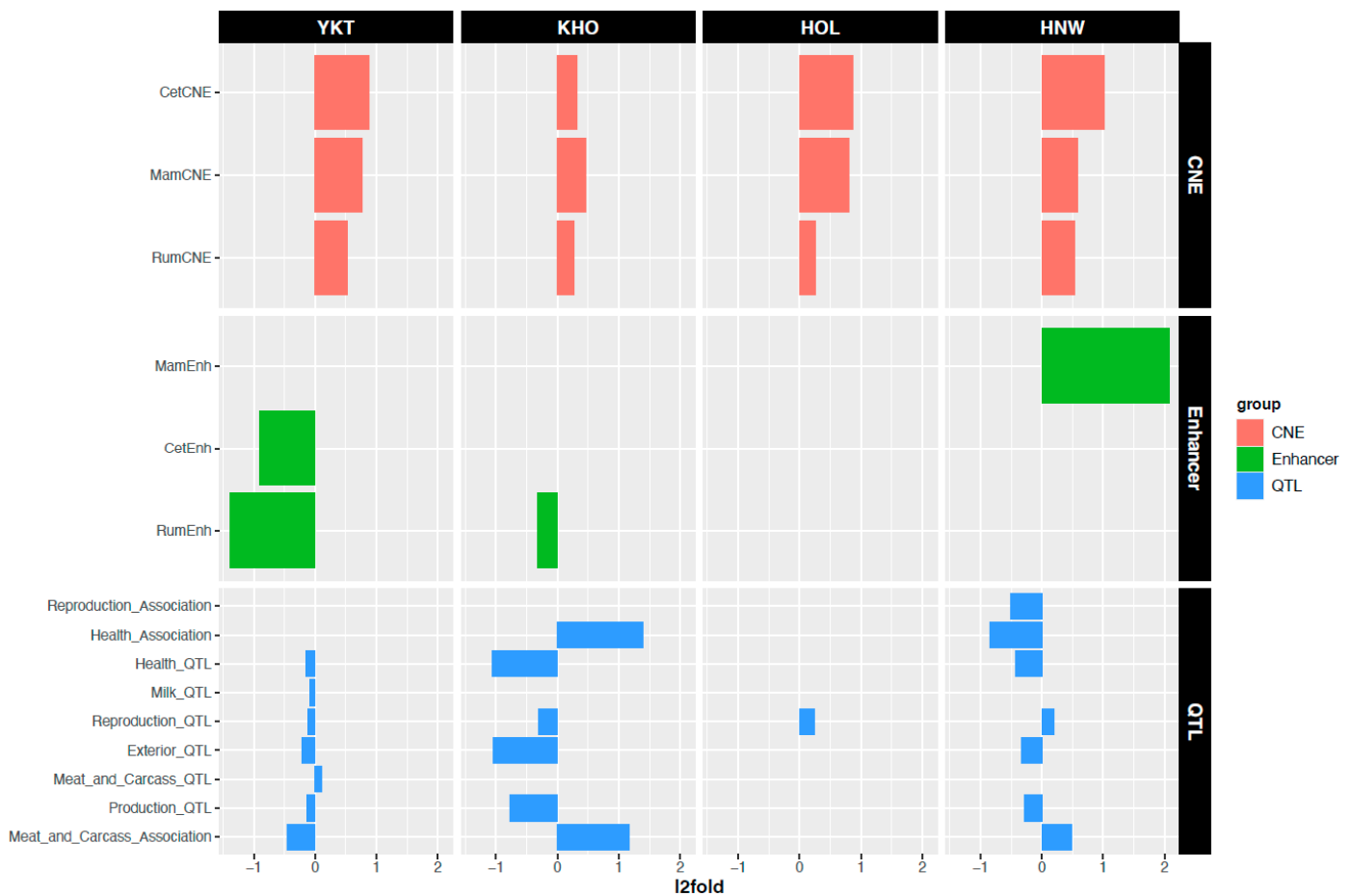


Figure 2. Breed-specific CNVR overlapping CNEs, enhancers, and QTLs. Abbreviations are as follows: YKT—Yakut, KHO—Kholmogory, HOL—Holstein, HNW—Hanwoo, CNE—conserved non-coding elements.

4. Conclusions

Overall, our results point to novel copy number variants and their potential contributions to local adaptations in the northernmost cattle breeds from Russia and shared CNV events involving GO terms such as a response to stimulus related to thermoregulation. The enrichment of ubiquitin proteins in Yakut cattle unique CNVRs might indicate their contribution to maintaining protein homeostasis; moreover, the enrichment of multiple fatty-acid related genes implicated in lipogenesis and growth traits suggests their potential involvement in thermoregulation, as well as the petite size of the Yakut cattle breed.

Finally, both the Yakut and Kholmogory cattle breeds had breed-specific CNVRs enriched in regulatory domains and enhancers, CNEs, and QTLs highlighting their potential contribution to harsh environment adaptations.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13091595/s1>, Table S1: Breed-specific CNVRs overlapping cattle genes (Btau6); Table S2a: GO analysis of genes in breed-specific CNVRs for the four breeds; Table S2b: DAVID functional cluster analysis of breed-specific CNVRs; Table S3a: GO analysis of genes in CNVRs shared between the Yakut and Kholmogory cattle breeds; Table S3b: DAVID functional cluster analysis of YKT-KHO shared CNVRs; Table S4a: GO analysis of breed-specific CNVRs in regulatory domain (GREAT); Table S4b: DAVID functional cluster analysis of breed-specific CNVRs in regulatory domain (GREAT).

Author Contributions: Conceptualization, L.B. and D.M.L.; methodology, L.B.; formal analysis, L.B.; investigation, L.B. and D.M.L.; resources, D.M.L. and N.S.Y.; writing—original draft preparation, L.B.; writing—review and editing, D.M.L. and N.S.Y.; project administration, D.M.L.; funding acquisition, D.M.L. and L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Russian Scientific Foundation (RSF) grant no. 19-76-20026. LB was funded by Marie Skłodowska-Curie grant agreement No. 703376.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Whole genome sequences data that support the findings of this study have been deposited in GenBank (PRJNA642008).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Feuk, L.; Carson, A.R.; Scherer, S.W. Structural variation in the human genome. *Nat. Rev. Genet.* **2006**, *7*, 85–97. [[CrossRef](#)] [[PubMed](#)]
2. Zhang, F.; Gu, W.; Hurles, M.E.; Lupski, J.R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 451–481. [[CrossRef](#)] [[PubMed](#)]
3. Zarrei, M.; MacDonald, J.R.; Merico, D.; Scherer, S.W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **2015**, *16*, 172–183. [[CrossRef](#)]
4. Conrad, D.F.; Pinto, D.; Redon, R.; Feuk, L.; Gokcumen, O.; Zhang, Y.; Aerts, J.; Andrews, T.D.; Barnes, C.; Campbell, P.; et al. Origins and functional impact of copy number variation in the human genome. *Nature* **2010**, *464*, 704–712. [[CrossRef](#)]
5. Aerts, E.; Beckers, S.; Zegers, D.; Van Hoorenbeeck, K.; Massa, G.; Verrijken, A.; Verhulst, S.L.; Van Gaal, L.F.; Van Hul, W. CNV analysis and mutation screening indicate an important role for the NPY4R gene in human obesity. *Obesity* **2016**, *24*, 970–976. [[CrossRef](#)]
6. Bickhart, D.M.; Xu, L.; Hutchison, J.L.; Cole, J.B.; Null, D.J.; Schroeder, S.G.; Song, J.; Garcia, J.F.; Sonstegard, T.S.; Van Tassell, C.P.; et al. Diversity and population-genetic properties of copy number variations and multicopy genes in cattle. *DNA Res.* **2016**, *23*, 253–262. [[CrossRef](#)]
7. Sudmant, P.H.; Mallick, S.; Nelson, B.J.; Hormozdiari, F.; Krumm, N.; Huddleston, J.; Coe, B.P.; Baker, C.; Nordenfelt, S.; Bamshad, M.; et al. Global diversity, population stratification, and selection of human copy-number variation. *Science* **2015**, *349*, aab3761. [[CrossRef](#)]
8. Chain, F.J.J.; Feulner, P.G.D.; Panchal, M.; Eizaguirre, C.; Samonte, I.E.; Kalbe, M.; Lenz, T.L.; Stoll, M.; Bornberg-Bauer, E.; Milinski, M.; et al. Extensive copy-number variation of young genes across stickleback populations. *PLoS Genet.* **2014**, *10*, e1004830. [[CrossRef](#)]
9. Durkin, K.; Coppeters, W.; Drögemüller, C.; Ahariz, N.; Cambisano, N.; Druet, T.; Fasquelle, C.; Haile, A.; Horin, P.; Huang, L.; et al. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature* **2012**, *482*, 81–84. [[CrossRef](#)]
10. Pielberg, G.; Olsson, C.; Syvänen, A.C.; Andersson, L. Unexpectedly high allelic diversity at the KIT locus causing dominant white color in the domestic pig. *Genetics* **2002**, *160*, 305–311. [[CrossRef](#)]
11. Norris, B.J.; Whan, V.A. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res.* **2008**, *18*, 1282–1293. [[CrossRef](#)] [[PubMed](#)]
12. Feliuss, M.; Beerling, M.-L.; Buchanan, D.S.; Theunissen, B.; Koolmees, P.A.; Lenstra, J.A. On the History of Cattle Genetic Resources. *Diversity* **2014**, *6*, 705–750. [[CrossRef](#)]
13. Crate, S.A. *Cows, Kin, and Globalization: An Ethnography of Sustainability*; Rowman Altamira: Lanham, MD, USA, 2006.
14. Yurchenko, A.A.; Yudin, N.; Aitnazarov, R.; Plyusnina, A.; Brukhin, V.; Soloshenko, V.; Lhasaranov, B.; Popov, R.; Paronyan, I.A.; Plemashov, K.V.; et al. Genome-wide genotyping uncovers genetic profiles and history of the Russian cattle breeds. *Heredity* **2018**, *120*, 125–137. [[CrossRef](#)]
15. Buggiotti, L.; Yurchenko, A.A.; Yudin, N.S.; Vander Jagt, C.J.; Vorobieva, N.V.; Kusliy, M.A.; Vasiliev, S.K.; Rodionov, A.N.; Boronetskaya, O.I.; Zinovieva, N.A.; et al. Demographic history, adaptation, and NRAP convergent evolution at amino acid residue 100 in the world northernmost cattle from Siberia. *Mol. Biol. Evol.* **2021**, *38*, 3093–3110. [[CrossRef](#)]
16. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)] [[PubMed](#)]
17. Leinonen, R.; Sugawara, H.; Shumway, M. The Sequence Read Archive. *Nucleic Acids Res.* **2011**, *39*, D19–D21. [[CrossRef](#)]
18. Klambauer, G.; Schwarzbauer, K.; Mayr, A.; Clevert, D.-A.; Mitterecker, A.; Bodenhofer, U.; Hochreiter, S. cn.MOPS: Mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* **2012**, *40*, e69. [[CrossRef](#)]

19. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [[CrossRef](#)]
20. Neph, S.; Kuehn, M.S.; Reynolds, A.P.; Haugen, E.; Thurman, R.E.; Johnson, A.K.; Rynes, E.; Maurano, M.T.; Vierstra, J.; Thomas, S.; et al. BEDOPS: High-performance genomic feature operations. *Bioinformatics* **2012**, *28*, 1919–1920. [[CrossRef](#)]
21. McLean, C.Y.; Bristor, D.; Hiller, M.; Clarke, S.L.; Schaar, B.T.; Lowe, C.B.; Wenger, A.M.; Bejerano, G. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **2010**, *28*, 495–501. [[CrossRef](#)]
22. Gel, B.; Serra, E. karyoploteR: An R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **2017**, *33*, 3088–3309. [[CrossRef](#)]
23. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4*, 44–57. [[CrossRef](#)] [[PubMed](#)]
24. Todgham, A.E.; Crombie, T.A.; Hofmann, G.E. The effect of temperature adaptation on the ubiquitin–proteasome pathway in notothenioid fishes. *J. Exp. Biol.* **2017**, *220*, 369–378. [[CrossRef](#)] [[PubMed](#)]
25. Yang, M.; Lv, J.; Zhang, L.; Li, M.; Zhou, Y.; Lan, X.; Lei, C.; Chen, H. Association study and expression analysis of CYP4A11 gene copy number variation in Chinese cattle. *Sci. Rep.* **2017**, *7*, 46599. [[CrossRef](#)]
26. Dirandeh, E.; Sayyar, M.A.; Ansari-Pirsaraei, Z.; Deldar, H.; Thatcher, W.W. Peripheral leucocyte molecular indicators of inflammation and oxidative stress are altered in dairy cows with embryonic loss. *Sci. Rep.* **2021**, *11*, 12771. [[CrossRef](#)]
27. Laperrousaz, E.; Denis, R.G.; Kassis, N.; Contreras, C.; López, M.; Luquet, S.; Cruciani-Guglielmacci, C.; Magnan, C. Lipoprotein lipase expression in hypothalamus is involved in the central regulation of thermogenesis and the response to cold exposure. *Front. Endocrinol.* **2018**, *9*, 103. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5861133/> (accessed on 26 April 2021). [[CrossRef](#)]
28. Orvar, B.L.; Sangwan, V.; Omann, F.; Dhindsa, R.S. Early steps in cold sensing by plant cells: The role of actin cytoskeleton and membrane fluidity. *Plant J.* **2000**, *23*, 785–794. [[CrossRef](#)]
29. Niimura, Y. On the origin and evolution of vertebrate olfactory receptor genes: Comparative genome analysis among 23 chordate species. *Genome Biol. Evol.* **2009**, *1*, 34–44. [[CrossRef](#)]
30. Gibson, O.R.; Taylor, L.; Watt, P.W.; Maxwell, N.S. Cross-adaptation: Heat and cold adaptation to improve physiological and cellular responses to hypoxia. *Sports Med.* **2017**, *47*, 1751–1768. [[CrossRef](#)]
31. Heger, A.; Webber, C.; Goodson, M.; Ponting, C.P.; Lunter, G. GAT: A simulation framework for testing the association of genomic intervals. *Bioinformatics* **2013**, *29*, 2046–2048. [[CrossRef](#)]
32. Farré, M.; Kim, J.; Proskuryakova, A.A.; Zhang, Y.; Kulemzina, A.I.; Li, Q.; Zhou, Y.; Xiong, Y.; Johnson, J.L.; Perelman, P.L.; et al. Evolution of gene regulation in ruminants differs between evolutionary breakpoint regions and homologous synteny blocks. *Genome Res.* **2019**, *29*, 576–589. [[CrossRef](#)]