



OPEN

Exploring the genetic factors behind the discrepancy in resistance to bovine tuberculosis between African zebu cattle and European taurine cattle

SangJung Lee^{1,2}, Charton Clémentine^{1,2} & Heebal Kim¹✉

Caused by the pathogenic agent *Mycobacterium bovis*, bovine tuberculosis (bTB) is a major concern in cattle breeding due to both its zoonotic potential and economic impact. Greater resistance to this disease has been reported in certain African zebu breeds compared to European taurine breeds. However the genetic basis for the lower susceptibility to bTB infection observed in zebu cattle remains poorly explored. This study was conducted on whole genome sequencing data of three bTB infection-resistant African zebu breeds and two bTB infection-susceptible taurine breeds to decipher the genetic background. A set of four selection signature statistics based on linkage disequilibrium, site frequency spectrum, and population differentiation were used on SNPs whereas between population variance based VST and t-test were used on CNVs. As a complement, genes from previous literature reported as candidate genes for bTB resistance were also inspected to identify genetic variations. Interestingly, the resulting nine candidate genes had deleterious missense variants (*SHC3*, *IFNGR1*, *TLR2*, *TLR6*, *IL1A*, *LRRK2*, *EP300* and *IRAK4*) or a CNV difference (*CD48*) segregating between the groups. The genes found in the study play a role in immune pathways activated during *Mycobacterium* infection, contributing to the proliferation of immune cells and the granuloma formation, ultimately modulating the outcome of the infectious event. In particular, a deleterious variant in the *LRRK2* gene, whose deficiency has been linked to improved prognosis upon tuberculosis infection, was found in the bTB infection-resistant zebu breeds. Therefore, these genes constitute credible candidates in explaining the discrepancy in *Mycobacterium bovis* infection susceptibility among different breed.

With around two million deaths reported annually, tuberculosis (TB) remains to be the world's deadliest infectious disease. In developing countries 10 to 15% of human TB cases are estimated to be caused by the bovine pathogenic agent *Mycobacterium bovis*¹. To date, bTB remains a major public health threat and a source of substantial economic loss in cattle breeding.

Mycobacterium bovis is part of the wider *Mycobacterium tuberculosis* complex (MTBC). Though it was originally believed that cattle transmitted the causative agent to humans, today's evidence show that it was more likely the other way around^{2,3}. Transmission was thought to have occurred during domestication, around 10,000 years ago, but a recent study from 2020 provides a more modern (3rd to twelfth century AD) estimation for *Mycobacterium bovis* origin in East Africa⁴.

Though the prevalence of bTB has been shown to be determined by cattle management, evidence also point to individual genetic variation in cattle resistance to bTB infection⁵. Bovine tuberculosis susceptibility appears to be a trait of moderate heritability, ranging from 0.21 to 0.37^{6,7}. Resistance to infection by disease-causing bacteria of the MTBC appears to be under polygenic control in numerous species⁸. Fewer TB infection resistance-related loci have been identified compared to other complex diseases when implementing GWAS⁹. Accordingly, bTB infection resistance studies, among which are^{10,11}, also revealed few candidate genes. Some of these genes could be related to non-specific immunity, particularly the ones involved in the defense of the respiratory tract (for

¹Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea. ²These authors contributed equally: SangJung Lee and Charton Clémentine. ✉email: heebal@snu.ac.kr

example genes encoding proteins regulating bronchial mucus secretion or non-specific macrophages in lungs) and might participate in the elimination of a low dose bTB contamination⁵. However, the genes tend not to replicate between studies.

On the other hand, cattle sub-species also influence resistance to bTB infection^{12,13}. Zebu (*Bos indicus*) breeds were historically shown to exhibit a higher resistance towards bTB than Taurine (*Bos taurus*) breeds^{14–17}. Later reports show a higher prevalence of bTB in the European Taurine (EUT) Holstein breed than in African Zebu (AFZ) (mainly Arsi breed) or AFZ × EUT crosses (22.2, 11.6 and 11.9% respectively) by conducting a comparative intradermal tuberculin test with injection of purified protein derivatives of *Mycobacterium bovis* on a total of 5,424 cattle¹³. In the same study, lesion severity scores of lungs and lymph nodes (according to the scoring procedure developed by Vordermeier et al.¹⁸) showed a significant difference among breeds. Subsequently, Vordermeier et al.¹⁹ found the AFZ Boran breed calves to be less susceptible to bTB infection than the EUT Holstein calves were. Moreover, after developing the disease, the severity of the lesions was significantly lower in indicine individuals than in Holstein. The indicine individuals were also found to derive more protection from the BCG Bacillus Calmette-Guerin (BCG) vaccination¹⁶. Also, previous studies show that bTB has a geographical origin overlapping with the long time habitats of the more resistant breeds, which coincides with our basis of the study^{4,20} (Fig. S1).

Because bTB appeared after the divergence between *Bos indicus* and *Bos taurus*, the higher resistance to bTB in zebu breeds may be due to one or more of the three following scenarios: (1) a recent (less than 2000 generations) and ongoing selection towards bTB infection resistance, (2) an ancient selection in immune genes due to exposure to various pathogens, occurring in one of the two sub-species and conferring by chance a pre-adaptation against bTB infection in zebu or (3) a random genetic drift alone that led to alleles more adapted towards resistance to bTB infection by chance. While *Bos indicus* breeds have been shown to exhibit a higher resistance to bTB infection than *Bos taurus* breeds, only one study aiming at identifying the genes related to this resistance was conducted in *Bos indicus* × *Bos taurus* crosses¹¹. However, due to the use of low-density markers, only one loci associated to the Toll-Like Receptor (TLR) complex was discovered. The causal mutations at the origin of the bTB infection resistance discrepancy between *Bos indicus* and *Bos taurus* breeds thus remains vastly unknown.

This study aimed to decipher the genetic basis for the differences in bTB susceptibility to infection at high resolution using WGS data from three African zebu breeds reported as more resistant to bTB infection from the aforementioned studies (Arsi, Ethiopian Boran and Kenyan Boran) and two bTB infection susceptible breeds (Holstein, and Jersey). In order to gain maximum insight, this study proceeded in two complementary steps. The first step identifies genomic regions related to immunity having undergone selection. The second step explores genomic regions already reported in the literature as modulating resistance to bTB infection.

Results

Alignment statistics

The CNVs calling accuracy has been shown to highly depend on the coverage and sequencing depth. Previous studies reported that a $5 \times$ depth coverage was sufficient for CNV detection^{21–23}, with accuracies as high as 97% being reached for a depth $> 8 \times$ ²³. The minimum and average mean depths reached in our study being $7.9 \times$ and $9.4 \times$, respectively, with associated alignment rate and coverage of 99.3 and 95.0% (Table. S2), we estimated that CNVs calling accuracy was sufficient.

Population structure

Population structure of the five breeds (30 individuals in the Resistant (Res) group and 20 individuals in the non-resistant (NRes) group) was examined to decipher the genetic distance between and within each group.

The number of ancestries was then estimated using the Admixture software²⁴. As evidenced from the cross-validation error estimation plot (Fig. 1a), three ancestral populations seemed the most sensible modeling choice considering the current dataset. The first hypothesis (three ancestral populations) isolates Res groups and NRes groups from each other while the second hypothesis (two ancestral populations) isolates NRes (Holstein and Jersey) from Res breeds. The split identified between the NRes and Res breeds using the pairwise fixation index ($F_{ST} > 0.19$, Table 1) was slightly higher than the F_{ST} between two NRes breeds (0.16). On the other hand, a low (< 0.03) F_{ST} was observed between the three Res breeds, indicating a closer genetic composition within the Res group than within the NRes group.

PCA eigenvalues evidenced a substantial jump between the first and the second dimensions. The first dimension of the PCA explained 19.4% of the variability in the dataset, while the second dimension explained 3.9% of the variability. As illustrated on (Fig. 1b), dimension 1 clearly separated the three African zebu breeds (Arsi, Ethiopian Boran, and Kenyan Boran) from the two European taurine breeds (Holstein and Jersey), whilst the second dimension segregated between the Holstein and Jersey breeds.

The CNV-based distance was determined by drawing a dendrogram based on the CNV difference between individuals (Fig. 1c). Similarly, to what was observed in the SNP-based population structure results, the Res breeds clustered together while some of the Jerseys in the NRes breeds seemed to be grouped separately. Identical to the SNP based methods, 3 clusters were recommended as the optimal population number from the results of the Gap statistics (Fig. 1d).

Candidate genes issued from the selection signature analyses

The six complementary methods (4 SNP based methods and 2 CNV based methods) implemented in the present study correspond to different selection signatures and are associated with various time scales. By annotating the upper 1% window of the empirical distribution obtained for each method, a total of 783, 882, 371, 619, 37, and

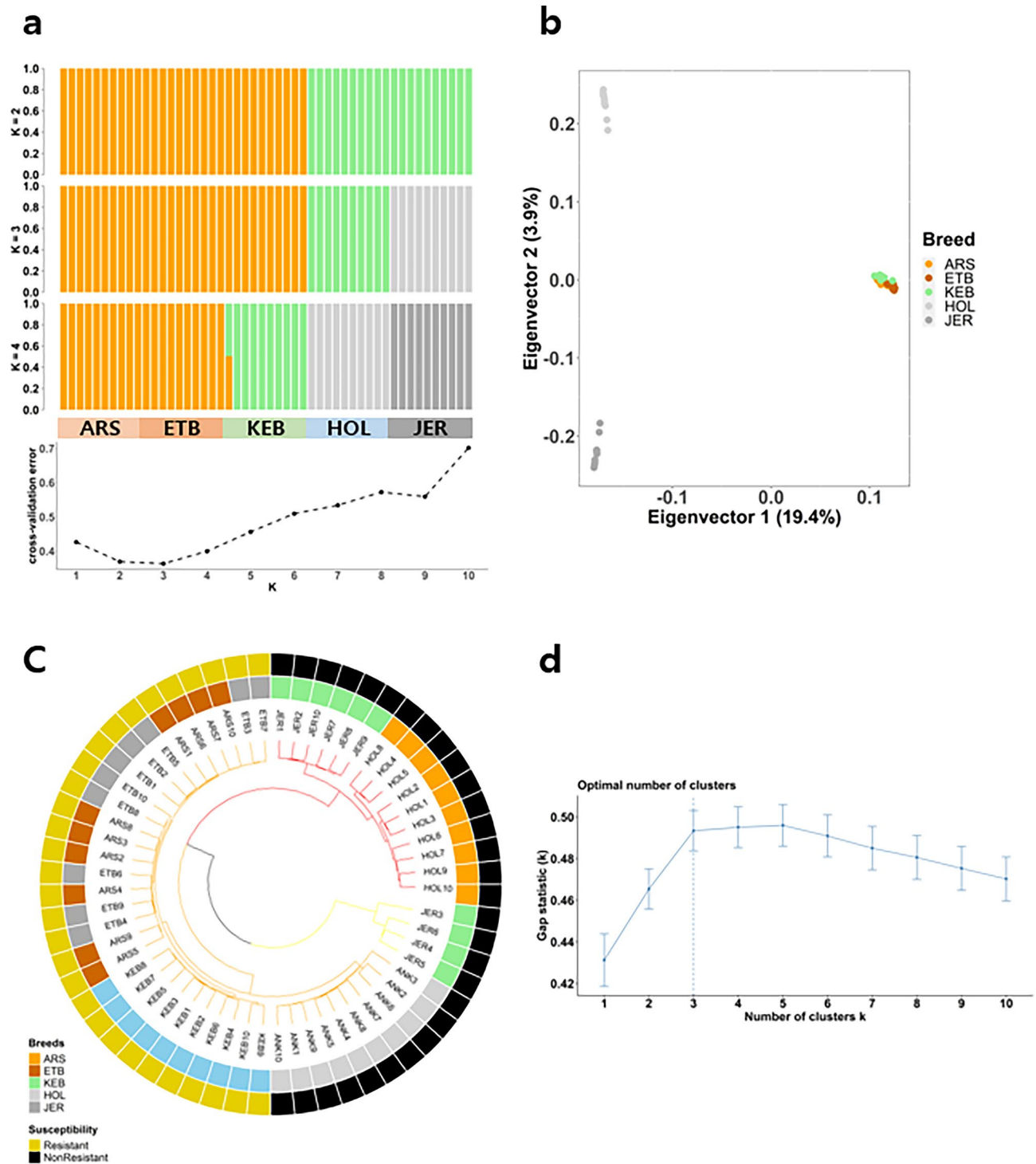


Figure 1. (a) Admixture analysis result for K 2–4 and cross-validation error plot. (b) PCA result for the 50 samples used in the study. (c) Dendrogram according to the copy number differences among individuals. (d) Gap statistic plot indicating the most probable K for clustering.

303 genes were identified as under recent or ongoing selection using the FLK, FST, XP-EHH, XP-CLR, VST, and t-test methods, respectively.

As for the SNP based methods, 635 genes were found to replicate for at least two methods. Among these genes, a total of 165 genes were found using three or more methods and 18 genes (*ABLM3*, *ADGRV1*, *AFAPIL1*, *BBS9*, *BBX*, *CLCC1*, *CMSS1*, *COL8A1*, *CSNK1A1*, *CTS*, *FILIP1L*, *GRPEL2*, *MED27*, *MSRB3*, *NELFCD*, *REEP1*, *TFEC*, and *ZNF831*) were found common to 4 out of the 4 SNP based methods. A total of 35 genes were found to replicate in the two CNV-based methods (Table S4).

The DAVID gene ontology analysis was implemented on every gene from the six methods above and resulted in the finding of 268 significant GO terms (GO:BP 114, GO:CC 71, and GO:MF 83) and 60 KEGG pathways.

	Arsi	Ethiopian Boran	Kenyan Boran	Holstein
Ethiopian Boran	0.00303			
Kenyan Boran	0.02085	0.01896		
Holstein	0.19126	0.20201	0.20822	
Jersey	0.20327	0.21396	0.22126	0.16426

Table 1. Pairwise Weir and Cockerman weighted F_{ST} computed between each of the five breeds: Arsi, Ethiopian Boran, Kenyan Boran, Holstein, and Jersey.

Resistance to bTB solicits both innate and adaptive immunity responses. The innate immunity of lung alveolar epithelial cells in response to *Mycobacterium* exposure plays a crucial role in eliminating the pathogen. T-cells, in turn, act as a regulatory force, mitigating the overabundance of pro-inflammatory factors that could otherwise lead to immunopathology. As a result, we chose to thoroughly explore all GO terms and KEGG pathways associated with relevant immune responses, through a comprehensive literature review. The most pertinent terms were “Bacterial invasion of epithelial cells”, “T cell receptor signaling pathway”, “Leukocyte trans endothelial migration”, “Negative regulation of T cell proliferation”, “Negative regulation of leukocyte cell–cell adhesion”, “Apoptosis”, “Cell death”, and “Endocytosis”. Many other terms, involved in protein adhesion, interaction and transportation, that could be potentially related to resistance to *Mycobacterium*, were also found. In total 282 genes were investigated, among which are 242 SNP-based genes and 40 CNV-based genes (Table S3).

The 242 SNP-based genes were inspected for the presence of missense variants segregating between the Res and NRes populations. The resulting 32 genes are displayed in (Figs. 2 and 3a). These missense variants were further filtered based on whether they were estimated as having a deleterious effect on the protein formation (Table 2). The *CDC42BPA*, *CLCC1*, *FNBP1*, *NWD1*, *SECEL*, *SHC3*, *TEX14*, and *UNC13B* missense variants were estimated as deleterious by the SIFT score and the *CDC42BPA*, *CDKAL*, *CLCC1*, *FRMPD1*, *IRAK3*, *LRRK2*, *NWD1*, *SEC24A*, *SHC3*, *SNX31*, *UNC13B* and *USP24* variants were found deleterious by the PROVEAN score.

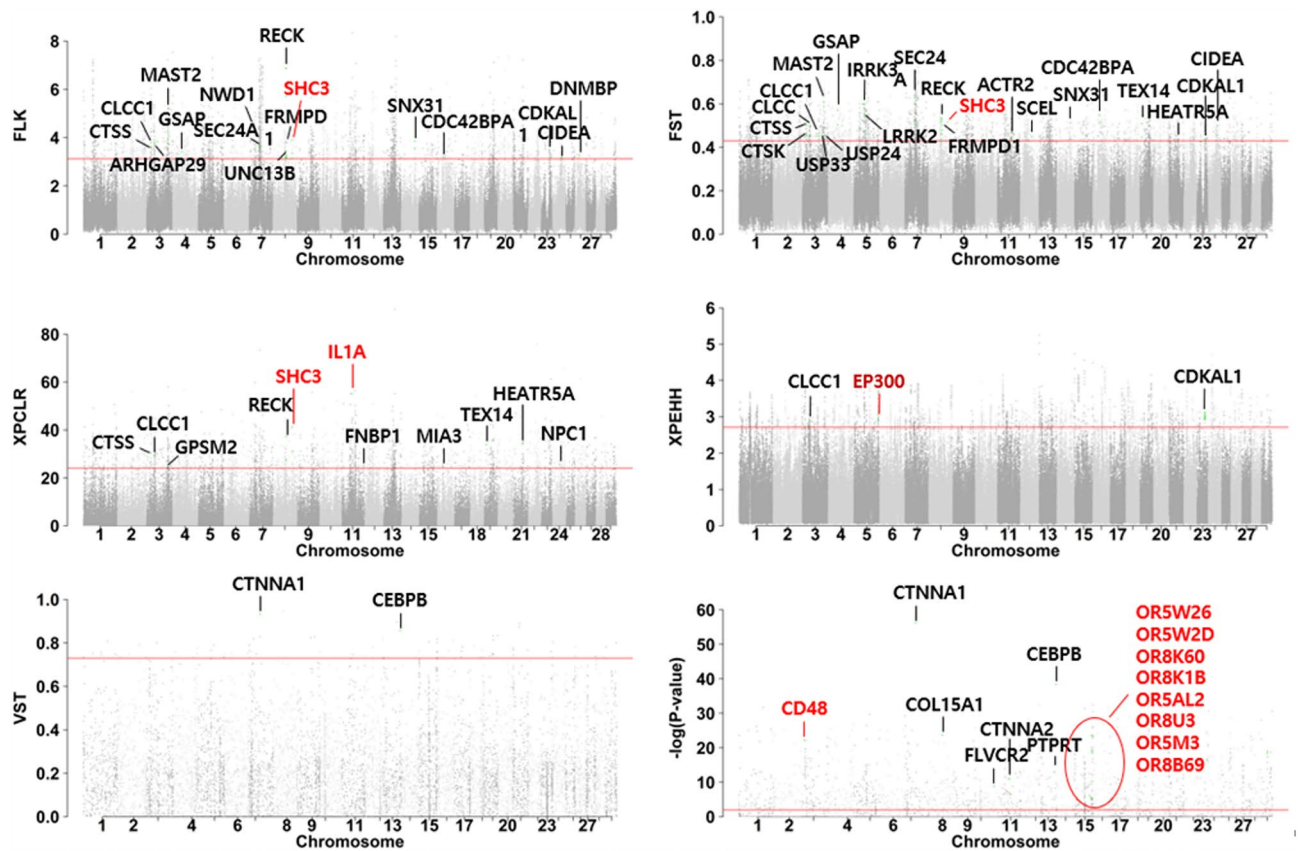


Figure 2. The score values for each window selected from the selection signature analyses (FLK, FST, XP-CLR, XP-EHH, VST, t-test) used in the study. The horizontal line in red shows the 0.01 significance level of the scores without binning. Labeled dots are the highest scoring windows in each gene. Genes labeled in red are the strongest candidate genes chosen from the study.

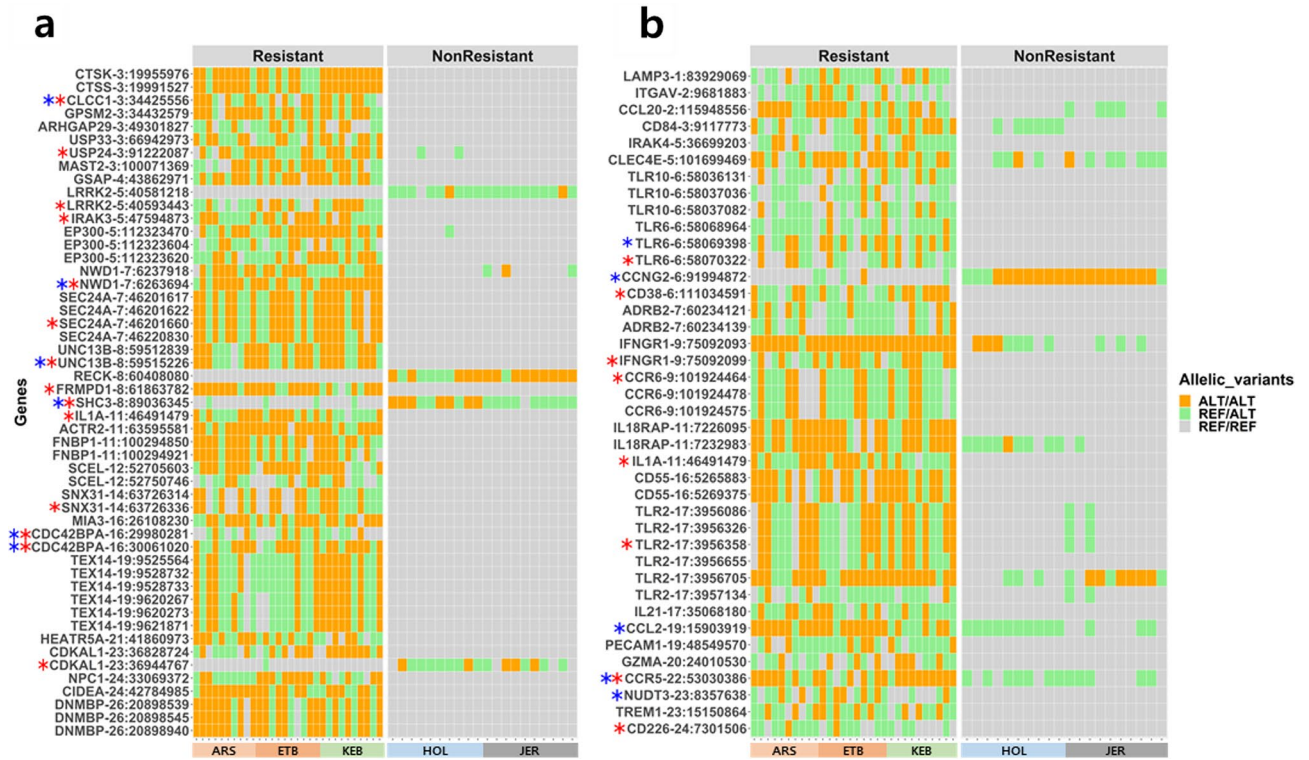


Figure 3. (a) Missense variants filtered after the selection signature analysis and gene ontology analysis. (b) Missense variants filtered after the literature review and gene ontology analysis. Blue and red asterisks indicate the significance in SIFT score (< 0.05) and PROVEAN score (< -1.3) respectively.

Gene, position (ref/alt)	AFZ alt	EUT alt	SIFT	Provean
<i>SHC3</i> , chr8:89,036,345 (C/T)	0.07	0.675	0	-4.197
<i>IFNGR1</i> , chr9:75,092,099 (C/A)	0.67	0	0.13	-2.2
<i>TLR2</i> , chr17:3,956,358 (A/T)	0.683	0.05	0.23	-1.301
<i>TLR6</i> , chr6:58,069,398 (T/C)	0.65	0	0.04	-0.399
<i>TLR6</i> , chr6:58,070,322 (T/C)	0.47	0	0.06	-5.737
<i>IL1A</i> , chr11:46,491,479 (C/G)	0.717	0	0.2	-2.159
<i>LRRK2</i> , chr5:40,593,443 (C/T)	0.6	0	0.07	-3.803
<i>EP300</i> , chr5:112,323,620 (G/A)	0.7	0	0.26	-0.248
<i>IRAK4</i> , chr5:36,699,203 (G/A)	0.483	0	0.25	-0.559

Table 2. Missense variant allele frequencies of the most serious candidate genes. SIFT score < 0.05 and Provean score < -1.3 were considered as significant.

The 40 CNV-based genes were further inspected for copy number differences within the gene. Remarkably, the *CD48* gene exhibited a copy number difference of up to 10 in certain intronic regions (Fig. 4a and b).

Candidate genes issued from the literature

A total of 239 genes were considered when implementing the second strategy, i.e. exploring the genomic difference in the genes already reported in the literature as having potential relation to resistance to bTB infection. Among the 239 genes investigated, 25 genes exhibited missense variants that contrasted to some extent between the groups. The *CCL2*, *CCNG2*, *CCR5*, *NUDT3*, and *TLR6* genes were identified as deleterious by the SIFT score and *CCR5*, *CCR6*, *CD226*, *CD38*, *IFNGR1*, *IL1A*, *TLR2*, and *TLR6* genes were chosen by PROVEAN score (Fig. 3b).

In the final step, we restricted the most serious candidate genes to the ones presenting a clear causative relationship to bTB resistance according to the literature: *SHC3*, *IFNGR1*, *TLR2*, *TLR6*, *IL1A*, *LRRK2*, *CD48*, *EP300* and *IRAK4*. The *SHC3* missense variant stood out as the sole gene variant found in the NRes breeds, whereas missense variants were exclusively identified in the Res breeds for *IFNGR1*, *TLR2*, *TLR6*, *IL1A*, *LRRK2*, *CD48*, *EP300*, and *IRAK4* genes (Table 2).

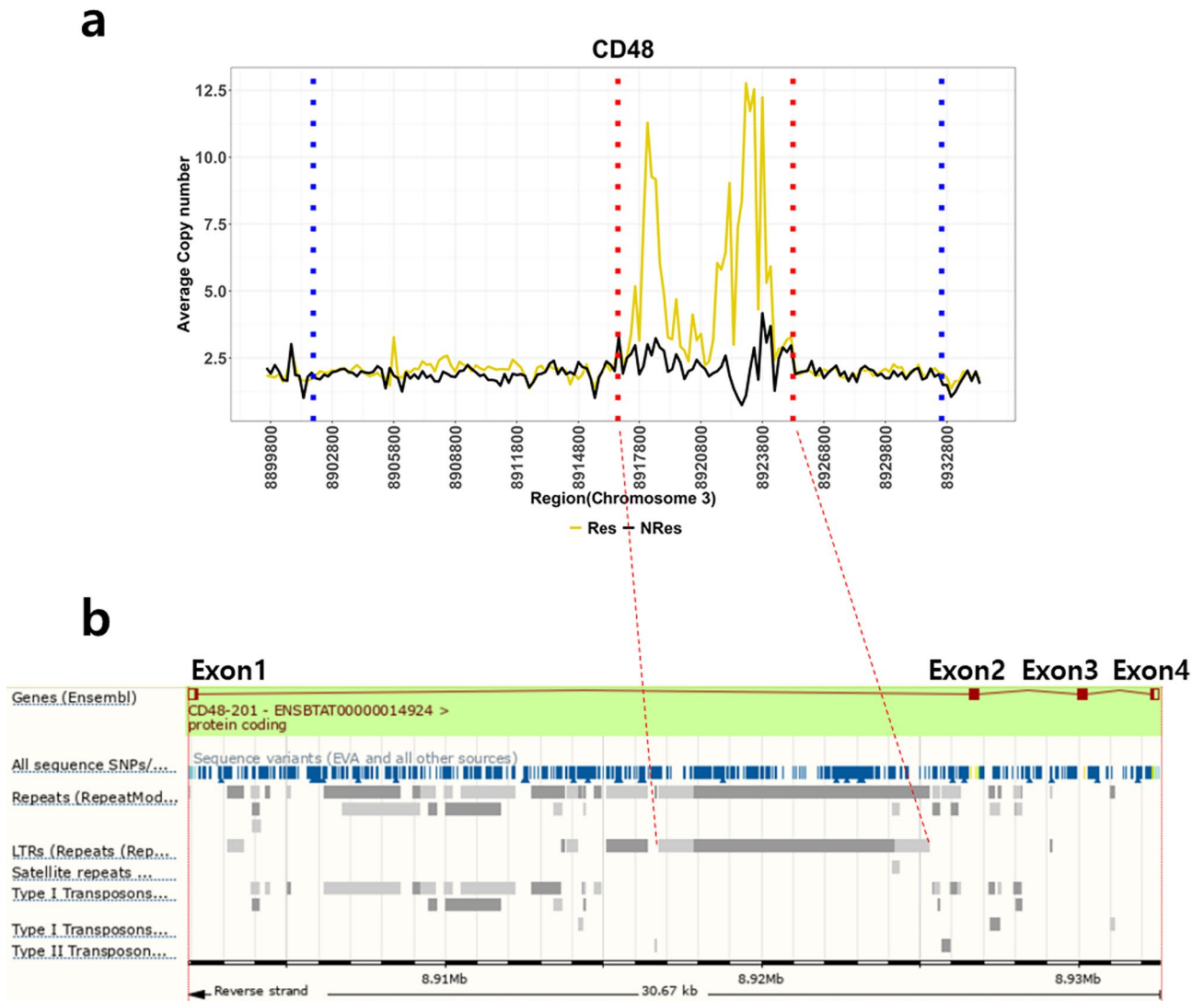


Figure 4. (a) Average copy number differences of every 200 bp regions were visualized as a line graph between Res and NRes in *CD48* (chromosome 3:8,901,877–8,932,543) with flanking regions of 2 kb on both sides. The x and y axes indicate the genomic region and average copy number, respectively. (b) Region of repeat was shown by Ensembl Genome browser. Repeats were called with Repeat Masker and Repeat Modeler. Region in-between red dotted lines indicate the overlapping repeat region which is mostly LTR (long tandem repeats). Blue line is the starting position of flanking regions.

Discussion

Mycobacterium has successfully developed the ability to sustain latent and persistent infections by limiting the development of adaptive immune responses^{25,26}. Accordingly, innate immunity is the key to determining the disease outcome upon *Mycobacterium* exposure. Several components of the innate immune system contribute to the pathogenesis of *Mycobacterium* infection, especially macrophages but also dendritic cells, neutrophils, mast cells, and NK (natural killer) cells. The recognition and phagocytosis of the bacteria are essential for the initiation of pro-inflammatory responses. Recognition is mediated through different receptors (toll-like receptors, nuclear oligomerization domain receptors, C-type lectin receptors and mannose receptor families^{27,28}) while the inflammatory response appears to be mainly mediated through TLRs and possibly dectin-1^{29–31}. Inflammation is a key process that can locally resolve TB infection via the recruitment of macrophages, neutrophils, NK cells, and eventually T-cells³². In turn, these cells clear invading pathogens or can control bacilli growth while containing them within granulomas³³. It is thus probable that tuberculosis sensitivity is modulated by both the host *Mycobacterium* recognition and the balance between pro- and anti-inflammatory factors. Indeed, weak inflammatory responses are reported to lead to mycobacterial infection propagation while strong inflammatory responses are likely to induce immunopathology³⁴. Part of this management is thought to be genetically determined as it was reported in diverse species^{34–37}.

In this study, we found evidence of differentially selected genes between the Res and the NRes breeds in several of the genes related to the pathways of known tuberculosis responses and NK cell-mediated cytotoxicity

(Fig. 5). Among these genes, the most serious candidates were *SHC3*, *IFNGR1*, *TLR2*, *TLR6*, *IL1A*, *LRRK2* and *CD48*. These genes harbored deleterious missense variants which allelic frequencies contrasted between Res and NRes individuals. In addition, we also considered the *EP300* and *IRAK4* genes, which scores were close to being significantly deleterious, as putative candidate genes. We describe the relationship between these genes and the immune pathways modulating bTB resistance.

A strong candidate gene for the discrepancy in bTB infection resistance found in this study is the *LRRK2* gene (Leucine-rich repeat kinase 2). Indeed, upon infection, TB mostly targets the host's macrophages, within which they establish a replicative niche. In reaction, the immune system targets TB in the macrophages' phagolysosomes, leading to the pathogen's elimination. In its active state, the *LRRK2* gene has been reported to have a negative impact on the TB outcome by delaying the maturation of phagosomes in macrophages, acting potentially by the recruitment of Class III phosphatidylinositol-3 kinase complex and Rubicon^{38,39}. Accordingly, *LRRK2* deficiency led to a clear decrease in the TB impact on mice. Compellingly, the missense variant observed in this gene was absent in the NRes group but found at an allelic frequency of 0.6 in the Res group. The *LRRK2* missense variant in the Res group could thus contribute to the bTB resistance phenotype and functional validation of this variant in further study might be decisive in deciphering bTB resistance.

The other candidate genes also have well-documented effects on the immune system. The SHC adaptor protein 3 (*SHC3*) gene is part of the SHC (Src Homology and Collagen) family, which transmits the extracellular signal recognized by the CD16 receptor of the NK cells. In turns, this signal activates different intracellular signaling pathways, among which the MAPK pathway, to release cytokines related to NK cell-mediated cytotoxicity^{40–42}. The missense variant, exclusively observed in NRes breeds with an allelic frequency of 0.6, warrants further investigation regarding its impact on managing *Mycobacterium bovis* infection, particularly in modulating

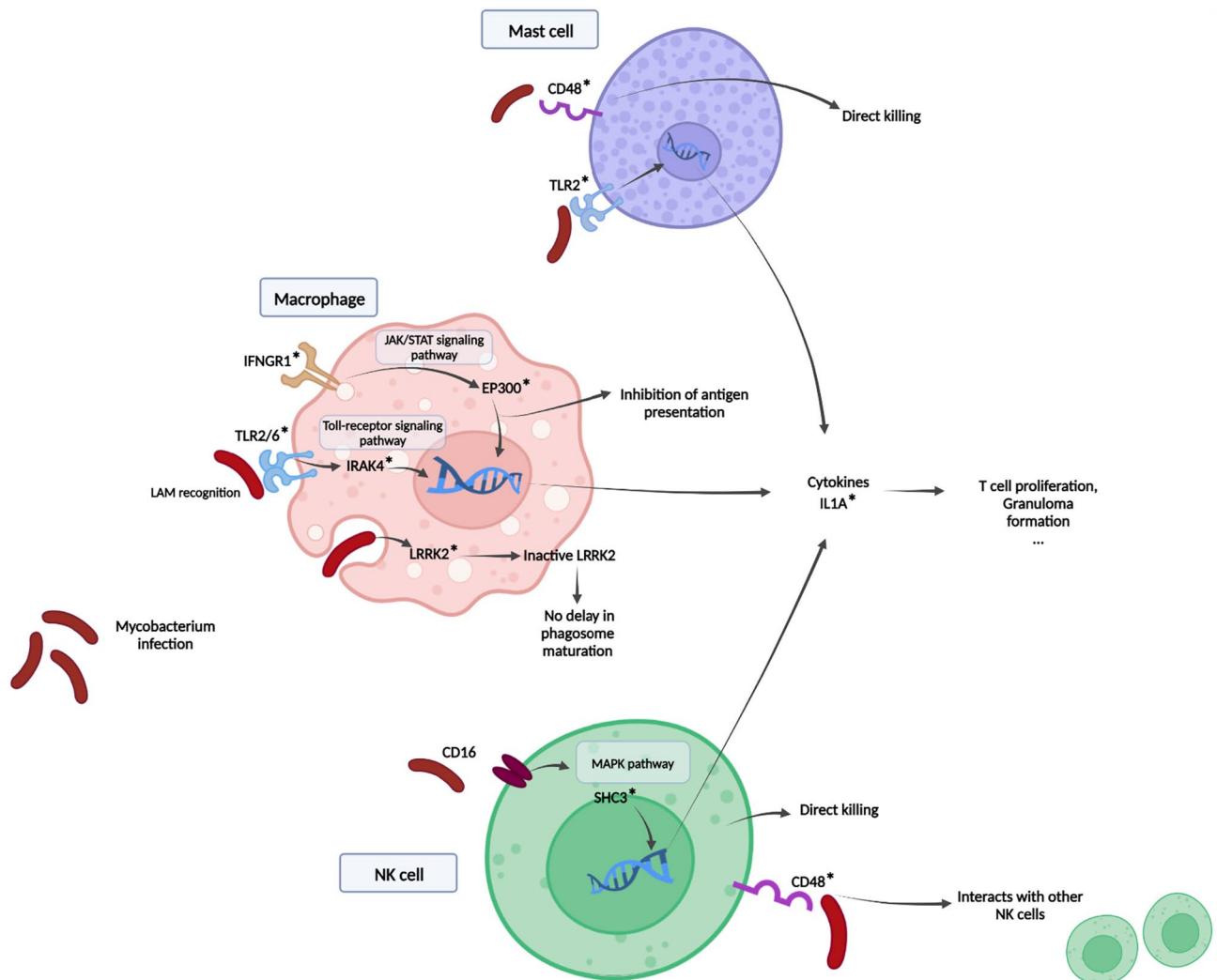


Figure 5. Involvement of the candidate genes in innate immune signaling after *Mycobacteria* encounters the immune cells. Genes presenting at least one missense variant in one of the two breed groups are identified with an asterisk, red in case this variant is probably deleterious (based on SIFT and Provean) score. The figure was drawn with Photoshop CC 2022 (<https://www.adobe.com/products/photoshop.html>).

inflammation through the *SHC3* gene. Functional validation is crucial to determine its effect, yet the significant finding of this deleterious variant solely in NRes breeds positions it as a compelling candidate for further study.

The remaining deleterious missense variants observed in the *EP300*, *IFNGR1*, *TLR2*, *TLR6*, *IL1A* and *IRAK4* genes all presented a higher allelic frequency in the Res breeds. The P300 proteins, coded by the *EP300* gene, interact with the STAT proteins in the JAK-STAT signaling pathways to control the antigen presentation by the Major Histocompatibility Complexes (MHC) 1 and 2^{43–46}. *IFNGR1* which is also a receptor for the JAK-STAT signaling pathway and is one of most frequently reported genes to show an association with TB and bTB susceptibility in human, buffalo, and cattle^{47–49}. Toll-like members, which are expressed on the membrane of sentinel cells (such as macrophages, dendritic cells and mast cells) can recognize a wide repertoire of microbes including the 19-kDa lipoprotein and lipoarabinomannan (LAM) of TB^{31,50–52}. Especially, *TLR1*, *TLR2*, *TLR4*, and *TLR6* in monocytes recruit adaptor proteins such as TIRAP, interleukin-1 receptor associated kinase (IRAK) or the innate immune signal transduction adaptor MyD88 and further activates other components, resulting in the over-expression of pro inflammatory cytokines such as IL-1 and IL-18⁵³. *TLR2* also contributes to the activation of mast cells after the TB infection^{54,55}. Interestingly, polymorphism in *TLR2*, *TLR4*, and *TLR6* genes has repeatedly been reported to have an association with susceptibility to *Mycobacterium* in cattle and humans^{56–61}. Inflammatory cytokine Interleukin 1-alpha, which is secreted after the recognition of *Mycobacterium* by TLRs in monocytes, plays a role in T cell proliferation, chemokine secretion and granuloma formation^{62–64}. Polymorphisms of *IL1A* in mice, human, and African buffalo were also modulated susceptibility to tuberculosis^{65–68}. The *IRAK4* gene codes for an essential enzyme involved in the TLR pathways. It activates NF- κ B and sequentially regulates cytokines and T cell activation^{69,70}. In case the *IRAK4* gene is disrupted, the signaling cascade can be interrupted and consequently alter the response to TB^{71,72}.

The higher prevalence of the deleterious missense variants in these genes observed in the Res breeds might seem counter-intuitive given their role in mediating immunity upon infection. However, as stated above, resistance to tuberculosis depends on a complex balance between pro and anti-inflammatory factors and a better understanding of how the observed variants interact with each other and might contribute to bTB resistance could only be verified through functional validation.

Last, though the *CD48* gene did not display any obvious differences in missense variants, it was both detected through the XP-EHH selection signature and had one of the highest average copy number differences between populations. CD48, combined with the immunoreceptors CD2 and CD244, plays various roles in the activation of cytotoxicity and proliferation of immune cells such as natural killer (NK) cells, B lymphocytes, T lymphocytes, dendritic cells, neutrophils, eosinophils, and mast cells (MC). In particular, the CD48 receptor has a crucial role in the function of NK cells^{73–75}, cell type that is reported to directly kill the *Mycobacteria* and activate other immune cells (monocytes, T cells) after TB infection^{76–78}. In addition, CD48 has been reported to activate mast cells, leading to the release of proinflammatory mediators controlling the uptake of TB⁷⁹.

Conclusions

In this study, we propose *LRRK2*, *SHC3*, *IFNGR1*, *TLR2*, *TLR6*, *IL1A*, *CD48* and other genes related to the innate immune system as candidate genes for the difference in bTB susceptibility reported between African zebu and European taurine cattle breeds. The *LRRK2* gene appears as a strong candidate as its deficiency has been shown to improve bTB resistance. The results, issued from the screening of WGS data at the base pair resolution, contribute a new layer of information regarding the genetic background of resistance towards bovine tuberculosis. Additionally, these results present potential candidates that warrant further investigation through functional validation. However, this study is limited due to lack of actual infection studies comparing more breeds and individuals. As, the genes reported in this study are expected to work in a connected way in modulating immunity through regulation of the pro- and anti-inflammatory factors, the candidate variants should be integrated in any further validation. After validation, these variants might serve as valuable tools for managing bTB infections in herds through future selective breeding programs.

Methods

Data description

The sequences were obtained from the publicly available database of NCBI under the following accession numbers: PRJNA574857 (Arsi, ETB, African buffalo), PRJNA312138 (Kenyan), PRJNA210521 (Holstein), PRJNA318089 (Jersey). Five cattle breeds reported in previous literature for different degrees of resistance to bovine tuberculosis infection and two African buffalo (included as an outgroup in compliance with the recommendations of the population differentiation method FLK) were used in this study. More precisely, the breeds selected as relatively resistant to bTB infection were three African indigenous Zebu breeds (*Bos indicus*) Arsi, Ethiopian Boran, and Kenyan Boran which will be referred to as the Res group throughout the manuscript. The breeds selected as relatively susceptible to bTB were two European *Bos taurus* dairy breeds Holstein, and Jersey. These susceptible breeds will be termed as NRes. A total of ten individuals per breed except for African buffalo (AFB) were used in the comparison. The 52 samples used in this study were part of a previously published larger dataset⁸⁰.

SNP variant calling

The raw sequence data was processed as described below. First, per-base sequence quality was checked using the fastQC software v.0.11.9⁸¹. Low quality bases and artifact sequences were removed using Trimomatic v.0.39⁸². BWA-MEM v.0.7.17⁸³ was then used to map high-quality sequence reads to the reference bovine genome (ARS-UCD1.2) with default parameters. Index files for reference and bam files were generated using Samtools v.1.9⁸⁴ and potential PCR duplicates were marked by 'MarkDuplicates' of Picard v.2.20.2⁸⁵. To perform base quality

score recalibration (BQSR), 'BaseRecalibrator' and 'PrintReads' arguments of genome analysis toolkit (GATK) v.4.40⁸⁶ were used. For masking known sites, variants file (ARS1.2PlusY_BQSR_v3.vcf.gz) from the 1000 Bull genomes Project was used, except for the two AFB samples. The base quality improvement was then checked with the 'AnalyzeCovariates' argument of GATK. Variant calling for AFB was performed as described in the previous paper⁸⁰. Alignment quality and overall alignment statistics were calculated with Sambamba⁸⁷.

To call SNPs from bam files, we used 'HaplotypeCaller' of GATK to create GVCF files. The GVCF files were then combined with 'CombinedGVCFs' of GATK. From combined GVCF files, SNPs were called with 'GenotypeGVCFs' and selected with 'SelectVariants' of GATK, respectively. The 'VariantFiltration' argument was then implemented to avoid possible false-positive calls with following options: cluster size 3; cluster window size 10; SNPs with mean depth (for all individuals) < 1/3 and > 3x (x, overall mean sequencing depth across all SNP sites); quality by depth < 2; phred scaled variant quality score < 30; strand odds ratio > 3; Fisher strand > 60; mapping quality < 40; mapping quality rank sum test < -12.5; read position rank sum test < -8. We also removed SNPs with missing genotype rates > 0.01 and non biallelic SNPs. Then with BEAGLE 4.0⁸⁸, genotype refinement, imputation, and phasing were successively applied to the remaining SNPs. Finally, after removing minor allele frequency < 0.01, high-quality SNPs were obtained and annotated with SnpEff v.4.3⁸⁹ for the downstream analysis.

Analysis of population structure

A set of analyses (3 SNP-based methods and one CNV-based method) were performed to comprehend the population structure of the dataset. First, a Principal Component Analysis (PCA) was performed using the Genome-wide Complex Traits Analysis (GCTA) v.1.93.2⁹⁰. To further investigate the genetic distance between the 5 breeds included in this study, the fixation index F_{ST} was computed pairwise between each of the 5 breeds using VCFtools v.4.0⁹¹, considering 10 kb windows without overlapping. Next, Admixture²⁴ was used to determine the most probable number of ancestral populations and assign an ancestral population to each of the 50 individuals. Finally, we drew a CNV based dendrogram among individuals by coding the events in the CNVR regions as -1, 0, and 1 for copy number deletion, no variation, and amplification, respectively. The optimal cluster (K) parameter according to the CNV was found using Gap statistics with nstart 25, K.max 10, B 50 options.

Detection of selection signatures in SNP

The appearance of bTB having most probably occurred less than 2,000 generations ago, selective sweeps triggered by this pathogen would necessarily be recent or still ongoing. The cross-population extended haplotype homozygosity statistics (XP-EHH), which is based on long haplotype detection, detects recent and ongoing selection having occurred within one of the populations⁹². This method is optimally used for selections having occurred within the last 30,000 years in humans⁹³, or 7200 years in equivalent for cattle considering standard approximates of 25- and 6- year generation intervals for the human and bovine species, respectively. XP-EHH was first computed using hapbin 1.3.0⁹⁴ with 10 kb nonoverlapping windows. The average score (absolute value) of all SNPs within a window was used as a summary statistic for each window. To account for SNPs number variation in each window, genomic windows were binned in increments of 200 SNPs with dropping windows containing less than 10 SNPs and combining windows over 600 SNPs into one bin. The empirical P value was defined as described in Pickrell et al.⁹⁵: within each bin, for each window i , P value is the fraction of a window with a value of the statistic greater than that in i . The results above P value of 0.01 in each bin were taken as candidates.

As a complement, three other methods that might detect selection in time frames occurring before and after the presumed apparition of the bTB causative agent were employed to test for putative ancient selection related to various pathogens exposure and conferring by chance a selective advantage to one of the groups in case of bTB exposure (pre-adaptation).

The cross-population composite likelihood ratio (XP-CLR) method⁹⁶ is one of the statistics developed to detect allele frequency differentiation signature which has been estimated to persist over the equivalent of 18,000 to 12,000 years in cattle. The XP-CLR method was implemented using scripts available at <https://github.com/hardnngnj/xpclr>⁹⁷, using the following options: -size 10,000; -maxsnps 1000; -minsnps 10. The results were averaged, binned and the P -value was defined in the same way informed for XP-EHH.

To make the most insights out of the available data, the commonly used statistic F_{ST} , based on population differentiation, was applied with VCFtools v4.0 using the -weir- F_{ST} -pop option and setting the - F_{ST} -window-size to 10,000. The results were averaged, binned and the P -value was defined according to the afore-mentioned protocol.

Finally, the extended Lewontin and Krakauer (FLK) method⁹⁸ which accounts for the hierarchical structure by using kinship matrix was employed. Two African buffalo were used as outgroups with the default options. The results were averaged, binned and the P -value was defined according to the afore-mentioned protocol.

All the regions selected from methods above were annotated based on the reference genome ARS-UCD1.2 from NCBI RefSeq database⁹⁹. Genes that overlap partly or entirely with the significant regions were defined as candidate genes.

CNV calling and CNVR identification

CNVs were called with the aligned read depth-based tool CNVnator v0.4.1¹⁰⁰, using a bin size of 200 bp and filtered with length > 1 kb. The p -value was based on t-test statistics < 0.001 and a fraction of reads with zero mapping quality q_0 < 0.5 as frequently described in the literature^{101,102}. After removing unmatched scaffolds, the 'CNV_overlap.py' script from (<https://github.com/bjtrost/TCAG-WGS-CNV-workflow>)¹⁰³ was applied to identify copy number variation region (CNVR) as the 50% overlap between remaining CNVs. To minimize false-positives, only CNVRs that were found in more than 3 individuals were kept for downstream analyses¹⁰⁴.

Detection of selection signatures in CNV

Two statistical methods were applied for detecting the population difference of CNVR between Res and NRes. First, V_{st} , a population differentiation estimator similar to F_{ST} , was calculated between the two groups as $V_{st} = (V_t - V_s) / V_t$ with V_t being the total variance of copy number among every individuals, V_s being the average variance within each population weighted by the number of individuals in the population¹⁰⁵. As small within-variance might generate false-positive CNVs when employing the V_{st} method, a two sample t-test between the populations was conducted to compare the copy numbers between the two groups (P -values adjusted with a Bonferroni correction) as a complement. Regions that were within the top 1% of the highest V_{st} and which had an adjusted t-test P -value below 0.01 were considered as population differentiated CNVRs. Genes overlapping with the CNVR were annotated.

Selection and characterization of candidate genes

To get more insights into the bovine tuberculosis susceptibility discrepancy between Res and NRes, a rough screening of the annotated genes from the putatively selected regions was performed using two different approaches. First, candidate genes issued from each method which are possibly related to immunity for tuberculosis susceptibility were filtered by implementing gene ontology using the Database for Annotation, Visualization and Integrated Discovery (DAVID)¹⁰⁶. Overrepresented Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways¹⁰⁷ were searched, using a P -value of 0.05 as the threshold for statistical significance. Genes that could be related to tuberculosis susceptibility were kept in further steps.

A literature review was performed as a second approach. The studies based on differentially expressed genes were considered for further steps (Table S1). As a result, a total of 239 genes from the literature that were reported for possible modulation of bTB sensitivity were individually considered as candidate genes and proceeded to the last selection step along with the genes found from the selection signature and GO analysis.

The candidate genes resulting from the SNP-based selection signature filtered with DAVID analysis and the literature-based candidate genes were further refined by keeping the genes whose visual inspection revealed segregation (more than 0.5 difference in either allele frequency) in missense variants between the 2 groups (performed using ARS-UCD1.2). The Sorting Intolerant from Tolerant (SIFT) score¹⁰⁸ provided by the Ensembl Variant Effect Predictor (VEP) and the Protein Variation Effect Analyzer (PROVEAN) score¹⁰⁹ were used to determine the outcome of the missense variants on the proteins. The SIFT score is calculated with a normalized probability of observing a different amino acid at a considered position and ranges from 0 to 1 with values comprised between 0 and 0.05 predicted to affect protein function. PROVEAN uses an alignment-based score approach which measures the similarity of a functional sequence and the query sequence. A score threshold of -1.3 was used as it was sufficient to detect the functional changes in the variants of non-human dataset with a balanced accuracy of 75.51%¹¹⁰. The strongest candidate genes were visualized with Haploview¹¹¹ to assess the degree of linkage among the selected regions and the adequacy of the window size used in the selection detecting methods (Figs. S2, S3, S4, S5). Haplotype blocks were detected based on the method by Gabriel et al.¹¹².

The CNV-related genes filtered with the DAVID analysis were also examined together with the literature genes and candidate genes that showed more than one average copy number difference (Table S4) were further inspected by visualizing the average copy number in 200 bp windows sliding along the gene.

Data availability

The dataset supporting the conclusions of this article is available in the Sequence Read Archive (SRA) repository with the Bioproject accession number PRJNA574857 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA574857/>), PRJNA312138 (<https://www.ncbi.nlm.nih.gov/bioproject/prjna312138>), PRJNA210521 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA210521>), and PRJNA318089 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA318089>).

Received: 27 October 2023; Accepted: 21 January 2024

Published online: 29 January 2024

References

1. Michel, A. L., Müller, B. & van Helden, P. D. Mycobacterium bovis at the animal–human interface: A problem, or not?. *Vet. Microbiol.* **140**, 371–381 (2010).
2. Hershkovitz, I. et al. Tuberculosis origin: The Neolithic scenario. *Tuberculosis* **95**, S122–S126 (2015).
3. Smith, N. H., Hewinson, R. G., Kremer, K., Brosch, R. & Gordon, S. V. Myths and misconceptions: The origin and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **7**, 537–544 (2009).
4. Loiseau, C. et al. An African origin for *Mycobacterium bovis*. *Evol. Med. Public Health* **2020**, 49 (2020).
5. Phillips, C. J. C., Foster, C. R. W., Morris, P. A. & Teverson, R. Genetic and management factors that influence the susceptibility of cattle to *Mycobacterium bovis* infection. *Anim. Health Res. Rev.* **3**, 3–13 (2002).
6. Raphaka, K. et al. Genomic regions underlying susceptibility to bovine tuberculosis in Holstein-Friesian cattle. *BMC Genet.* **18**, 27 (2017).
7. Bermingham, M. L. et al. Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis. *Heredity (Edinb)* **112**, 543 (2014).
8. Allen, A. R. et al. Bovine tuberculosis: The genetic basis of host susceptibility. *Proc. R. Soc. B Biol. Sci.* **277**, 2737 (2010).
9. Curtis, J. et al. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nat. Genet.* **47**, 523 (2015).
10. Kadarmideen, H. N., Ali, A. A., Thomson, P. C., Müller, B. & Zinsstag, J. Polymorphisms of the SLC11A1 gene and resistance to bovine tuberculosis in African Zebu cattle. *Anim. Genet.* **42**, 656–658 (2011).
11. Kassahun, Y. et al. Admixture mapping of tuberculosis and pigmentation-related traits in an African-European hybrid cattle population. *Front. Genet.* **6**, 210 (2015).

12. Ameni, G. *et al.* Cattle husbandry in Ethiopia is a predominant factor affecting the pathology of bovine tuberculosis and gamma interferon responses to mycobacterial antigens. *Clin. Vaccine Immunol.* **13**, 1030 (2006).
13. Ameni, G. *et al.* High prevalence and increased severity of pathology of bovine tuberculosis in Holsteins compared to zebu breeds under field cattle husbandry in central Ethiopia. *Clin. Vaccine Immunol.* **14**, 1356 (2007).
14. Carmichael, J. Bovine tuberculosis in the tropics, with special reference to Uganda. Part I. *J. Compar. Pathol Ther.* **52**, 322–335 (1939).
15. Soparkar, M. B. The relative susceptibility of indian milch cattle of various breeds to tuberculosis. *Indian J. Med. Res.* **13**, 755–780 (1926).
16. Ellwood, D. C. & Waddington, F. G. A second experiment to challenge the resistance to tuberculosis in B.C.G. vaccinated cattle in Malawi. *Br. Vet. J.* **128**, 619–626 (1972).
17. Carmichael, J. Bovine tuberculosis in the tropics with special reference to Uganda. Part II. *Vet. J.* **97**, 329–339 (1941).
18. Vordermeier, H. M. *et al.* Correlation of ESAT-6-specific gamma interferon production with pathology in cattle following *Mycobacterium bovis* BCG vaccination against experimental bovine tuberculosis. *Infect. Immunol.* **70**, 3026–3032 (2002).
19. Vordermeier, M. *et al.* The influence of cattle breed on susceptibility to bovine tuberculosis in Ethiopia. *Comp. Immunol. Microbiol. Infect. Dis.* **35**, 227–232 (2012).
20. Rito, T. *et al.* Evolution and molecular characteristics of *Mycobacterium tuberculosis* and *Mycobacterium bovis*. In *Tuberculosis. Integrated Science* (ed. Rezaei, N.) (Springer, Cham, 2023).
21. Altshuler, D. M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
22. Bickhart, D. M. *et al.* Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* **22**, 778–790 (2012).
23. Sudmant, P. H. *et al.* Diversity of human copy number variation and multicopy genes. *Science* **330**, 641 (2010).
24. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
25. Goldberg, M. F., Saini, N. K. & Porcelli, S. A. Evasion of innate and adaptive immunity by *Mycobacterium tuberculosis*. *Microbiol. Spectr.* **2**, 742 (2014).
26. Korb, V. C., Chuturgoon, A. A. & Moodley, D. *Mycobacterium tuberculosis*: Manipulator of protective immunity. *Int. J. Mol. Sci.* **17**, 131 (2016).
27. Bassessar, V. *et al.* Immunological host responses against *Mycobacterium tuberculosis* and *M. bovis* infection: A review. *Int. J. Curr. Microbiol. App. Sci* **9**, 2150–2171 (2020).
28. Coll, R. C. & O'Neill, L. A. J. New insights into the regulation of signalling by toll-like receptors and nod-like receptors. *J. Innate Immunol.* **2**, 406–421 (2010).
29. Schlesinger, L. S. Macrophage phagocytosis of virulent but not attenuated strains of *Mycobacterium tuberculosis* is mediated by mannose receptors in addition to complement receptors. *J. Immunol.* **150**, 2920 (1993).
30. Maeda, N. *et al.* The cell surface receptor DC-SIGN discriminates between *Mycobacterium* species through selective recognition of the mannose caps on lipoarabinomannan. *J. Biol. Chem.* **278**, 5513–5516 (2003).
31. Brightbill, H. D. *et al.* Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. *Science* **1979**(285), 732–736 (1999).
32. Allen, M. *et al.* Mechanisms of control of *Mycobacterium tuberculosis* by NK cells: Role of glutathione. *Front. Immunol.* **6**, 508 (2015).
33. Zuiga, J. *et al.* Cellular and humoral mechanisms involved in the control of tuberculosis. *Clin. Dev. Immunol.* **2012**, 193923 (2012).
34. Tobin, D. M. *et al.* The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* **140**, 717 (2010).
35. Wu, H. *et al.* TALE nickase-mediated SP110 knockin endows cattle with increased resistance to tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1530–E1539 (2015).
36. Mahdaviyani, S. A. *et al.* Mendelian susceptibility to mycobacterial disease (MSMD): Clinical and genetic features of 32 Iranian patients. *J. Clin. Immunol.* **40**, 872–882 (2020).
37. Sugawara, I. *et al.* Mycobacterial infection in natural killer T cell knockout mice. *Tuberculosis* **82**, 97–104 (2002).
38. Weindel, C. G. *et al.* LRRK2 maintains mitochondrial homeostasis and regulates innate immune responses to *Mycobacterium tuberculosis*. *Elife* **9**, e51071 (2020).
39. Härtlova, A. *et al.* LRRK2 is a negative regulator of *Mycobacterium tuberculosis* phagosome maturation in macrophages. *EMBO J.* **37**, e98694 (2018).
40. Chiesa, S., Tomasello, E., Vivier, E. & Vély, F. Coordination of activating and inhibitory signals in natural killer cells. *Mol. Immunol.* **42**, 477–484 (2005).
41. Ahmed, S. B. M. & Prigent, S. A. Insights into the Shc family of adaptor proteins. *J. Mol. Signal* **12**, 1–17 (2017).
42. Galandrini, R. *et al.* The adaptor protein shc is involved in the negative regulation of NK cell-mediated cytotoxicity. *Eur. J. Immunol.* <https://doi.org/10.1002/1521-4141> (2001).
43. Zhu, G., Pei, L., Li, Y. & Gou, X. EP300 mutation is associated with tumor mutation burden and promotes antitumor immunity in bladder cancer patients. *Aging (Albany NY)* **12**, 2132 (2020).
44. Zhou, Y. *et al.* Activation of NF- κ B and p300/CBP potentiates cancer chemoimmunotherapy through induction of MHC-I antigen presentation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2025840118 (2021).
45. Schindler, C., Levy, D. E. & Decker, T. JAK-STAT signaling: From interferons to cytokines. *J. Biol. Chem.* **282**, 20059–20063 (2007).
46. Paulson, M. *et al.* Stat protein transactivation domains recruit p300/CBP through widely divergent sequences. *J. Biol. Chem.* **274**, 25343–25349 (1999).
47. Driscoll, E. E., Hoffman, J. I., Green, L. E., Medley, G. F. & Amos, W. A preliminary study of genetic factors that influence susceptibility to bovine tuberculosis in the british cattle herd. *PLoS One* **6**, e18806 (2011).
48. Velez, D. R. *et al.* NOS2A, TLR4, and IFNGR1 interactions influence pulmonary tuberculosis susceptibility in African-Americans. *Hum. Genet.* **126**, 643 (2009).
49. Cheng, L., Zhang, F., Wang, Y., Chen, J. & Yuan, X. Association between IFNGR1 gene polymorphisms and tuberculosis susceptibility: A meta-analysis. *Front. Public Health* **10**, 976221 (2022).
50. Tang, X. *et al.* Mannose-capped lipoarabinomannan (ManLAM) binding TLR2 activates mast cells to release exosomes and induces M2 polarization of macrophages. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* **37**, 481–486 (2021).
51. Supajatura, V. *et al.* Differential responses of mast cell Toll-like receptors 2 and 4 in allergy and innate immunity. *J. Clin. Invest.* **109**, 1351 (2002).
52. Aliprantis, A. O. *et al.* Cell activation and apoptosis by bacterial lipoproteins through Toll-like receptor-2. *Science* **1979**(285), 736–739 (1999).
53. Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: Update on toll-like receptors. *Nat. Immunol.* **11**, 373–384 (2010).
54. Garcia-Rodriguez, K. M. *et al.* Bacillus Calmette–Guérin-induced human mast cell activation relies on IL-33 priming. *Int. J. Mol. Sci.* **23**, 7549 (2022).

55. Carlos, D. *et al.* TLR2-dependent mast cell activation contributes to the control of *Mycobacterium tuberculosis* infection. *Microbes Infect.* **11**, 770–778 (2009).
56. Song, Y., Sun, L., Guo, A. & Yang, L. Toll-like receptor 6 gene polymorphisms increase the risk of bovine tuberculosis in Chinese Holstein cattle. *Acta Histochem.* **116**, 1159–1162 (2014).
57. Sharma, B. S. *et al.* Association of TLR4 polymorphisms with *Mycobacterium avium* subspecies paratuberculosis infection status in Canadian Holsteins. *Anim. Genet.* **46**, 560–565 (2015).
58. Koets, A. *et al.* Susceptibility to paratuberculosis infection in cattle is associated with single nucleotide polymorphisms in Toll-like receptor 2 which modulate immune responses against *Mycobacterium avium* subspecies paratuberculosis. *Prev. Vet. Med.* **93**, 305–315 (2010).
59. Velez, D. R. *et al.* Variants in toll-like receptors 2 and 9 influence susceptibility to pulmonary tuberculosis in Caucasians, African-Americans, and West Africans. *Hum. Genet.* **127**, 65 (2010).
60. Sánchez, D. *et al.* Role of TLR2- and TLR4-mediated signaling in *Mycobacterium tuberculosis*-induced macrophage death. *Cell Immunol.* **260**, 128–136 (2010).
61. Ben-Ali, M., Barbouche, M. R., Bousnina, S., Chabbou, A. & Dellagi, K. Toll-like receptor 2 Arg677Trp polymorphism is associated with susceptibility to tuberculosis in Tunisian patients. *Clin. Diagn. Lab. Immunol.* **11**, 625 (2004).
62. Kurt-Jones, E. A., Fiers, W. & Pober, J. S. Membrane interleukin 1 induction on human endothelial cells and dermal fibroblasts. *J. Immunol.* **139**, 2317–2324 (1987).
63. Hernandez-Pando, R. *et al.* Analysis of the local kinetics and localization of interleukin-1 alpha, tumour necrosis factor-alpha and transforming growth factor-beta, during the course of experimental pulmonary tuberculosis. *Immunology* **90**, 607 (1997).
64. Juffermans, N. P. *et al.* Interleukin-1 signaling is essential for host defense during murine pulmonary tuberculosis. *J. Infect. Dis.* **182**, 902–908 (2000).
65. Mayer-Barber, K. D. *et al.* Innate and adaptive interferons suppress IL-1 α and IL-1 β production by distinct pulmonary myeloid subsets during *Mycobacterium tuberculosis* infection. *Immunity* **35**, 1023 (2011).
66. Guler, R. *et al.* Blocking IL-1 α but not IL-1 β increases susceptibility to chronic *Mycobacterium tuberculosis* infection in mice. *Vaccine* **29**, 1339–1346 (2011).
67. Bellamy, R. *et al.* Assessment of the interleukin 1 gene cluster and other candidate gene polymorphisms in host susceptibility to tuberculosis. *Tubercle Lung Dis.* **79**, 83–89 (1998).
68. le Roex, N., Koets, A. P., van Helden, P. D. & Hoal, E. G. Gene polymorphisms in African buffalo associated with susceptibility to bovine tuberculosis infection. *PLoS One* **8**, 64494 (2013).
69. Suzuki, N. *et al.* A critical role for the innate immune signaling molecule IRAK-4 in T cell activation. *Science* **1979**(311), 1927–1932 (2006).
70. Suzuki, N. & Saito, T. IRAK-4—a shared NF- κ B activator in innate and acquired immunity. *Trends Immunol.* **27**, 566–572 (2006).
71. Wang, Z., Wesche, H., Stevens, T., Walker, N. & Yeh, W.-C. IRAK-4 inhibitors for inflammation. *Curr. Top. Med. Chem.* **9**, 724 (2009).
72. Boro, M., Singh, V. & Balaji, K. N. *Mycobacterium tuberculosis*-triggered Hippo pathway orchestrates CXCL1/2 expression to modulate host immune responses. *Sci. Rep.* **6**, 1–14 (2016).
73. Kim, E. O. *et al.* Homotypic cell to cell cross-talk among human natural killer cells reveals differential and overlapping roles of 2B4 and CD2. *J. Biol. Chem.* **285**, 41755 (2010).
74. Nakajima, H., Cella, M., Langen, H., Friedlein, A. & Colonna, M. Activating interactions in human NK cell recognition: The role of 2B4-CD48. *Eur. J. Immunol.* [https://doi.org/10.1002/\(SICI\)1521-4141\(199905\)29:05](https://doi.org/10.1002/(SICI)1521-4141(199905)29:05) (1999).
75. Lee, K. M. *et al.* Requirement of homotypic NK-cell interactions through 2B4(CD244)/CD48 in the generation of NK effector functions. *Blood* **107**, 3181 (2006).
76. Crome, S. Q., Lang, P. A., Lang, K. S. & Ohashi, P. S. Natural killer cells regulate diverse T cell responses. *Trends Immunol.* **34**, 342–349 (2013).
77. Siddiqui, N. & Hope, J. Differential recruitment and activation of natural killer cell sub-populations by *Mycobacterium bovis*-infected dendritic cells. *Eur. J. Immunol.* **43**, 159–169 (2013).
78. Allen, M. *et al.* Mechanisms of control of *Mycobacterium tuberculosis* by NK cells: Role of glutathione. *Front. Immunol.* **6**, 1 (2015).
79. Muñoz, S., Hernández-Pando, R., Abraham, S. N. & Enciso, J. A. Mast cell activation by *Mycobacterium tuberculosis*: Mediator release and role of CD48. *J. Immunol.* **170**, 5590–5596 (2003).
80. Kim, K. *et al.* The mosaic genome of indigenous African cattle as a unique genetic resource for African pastoralism. *Nat. Genet.* **52**, 1099–1110 (2020).
81. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
82. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
83. Md, V., Misra, S., Li, H. & Aluru, S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. in *Proceedings - 2019 IEEE 33rd International Parallel and Distributed Processing Symposium, IPDPS 2019* 314–324 (2019) doi:<https://doi.org/10.1109/IPDPS.2019.00041>.
84. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
85. Picard Tools - By Broad Institute. <https://broadinstitute.github.io/picard/>.
86. McKenna, A. *et al.* The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
87. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
88. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084 (2007).
89. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80 (2012).
90. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76 (2011).
91. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156 (2011).
92. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913 (2007).
93. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **1979**(312), 1614–1620 (2006).
94. Maclean, C. A., Chue Hong, N. P. & Prendergast, J. G. D. hapbin: An efficient program for performing haplotype-based scans for positive selection in large genomic datasets. *Mol. Biol. Evol.* **32**, 3027 (2015).
95. Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826 (2009).
96. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393 (2010).
97. GitHub - hardingnj/xpclr: Code to compute the XP-CLR statistic to infer natural selection. <https://github.com/hardingnj/xpclr>.

98. Bonhomme, M. *et al.* Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics* **186**, 241–262 (2010).
99. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754 (2018).
100. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974 (2011).
101. Genova, F. *et al.* First genome-wide CNV mapping in FELIS CATUS using next generation sequencing data. *BMC Genom.* **19**, 1–13 (2018).
102. Dorshorst, B. *et al.* Dominant red coat color in holstein cattle is associated with a missense mutation in the coatmer protein complex, subunit alpha (COPA) gene. *PLoS One* **10**, e0128969 (2015).
103. GitHub - bjtrost/TCAG-WGS-CNV-workflow: Scripts involved in our workflow for detecting CNVs from WGS data using read depth-based methods. <https://github.com/bjtrost/TCAG-WGS-CNV-workflow>.
104. Pierce, M. D., Dzama, K. & Muchadevi, F. C. Genetic diversity of seven cattle breeds inferred using copy number variations. *Front. Genet.* **9**, 340644 (2018).
105. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
106. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2008).
107. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27 (2000).
108. Sim, N. L. *et al.* SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452 (2012).
109. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688 (2012).
110. PROVEAN | J. Craig Venter Institute. <https://www.jcvi.org/research/provean#faq>.
111. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
112. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **1979**(296), 2225–2229 (2002).

Acknowledgements

The authors thank the institutions and personnel that participated in sampling the African cattle: ILRI Kapiti Ranch (Kenya) and the Institute of Biodiversity (Ethiopia). We also thank the Directors of Veterinary Services and cattle keepers from Ethiopia and Kenya for their assistance and permission to sample the animals. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C209411113).

Author contributions

H.B.K led the study. S.J.L and C.C. analyzed the data, drafted the manuscript and revised the manuscript. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-52606-2>.

Correspondence and requests for materials should be addressed to H.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024