Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Marker-assisted selection for heat tolerance in Criollo cattle misses slick PRLR carriers when only one allele is genotyped

Luiz Dione Barbosa De-Melo

luiz.melo@ifrj.edu.br

Federal Institute of Education, Science and Technology of Rio de Janeiro (IFRJ)

Article

Keywords: heat tolerance, slick coat, prolactin receptor (PRLR), Criollo cattle, marker-assisted selection, local ancestry

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11629430/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Marker-assisted selection for heat tolerance in Criollo cattle misses slick *PRLR* carriers when only one allele is genotyped

Luiz Dione Barbosa De-Melo[1,*]
[1] Molecular Genetics Laboratory, Federal Institute of Education, Science and Technology of Rio de Janeiro (IFRJ), Rio de Janeiro, Brazil
* Corresponding author: luiz.melo@ifrj.edu.br

Abstract

Some tropical cattle breeds have a short, sleek "slick" coat that helps them shed heat. This arises from mutations that shorten the prolactin receptor (*PRLR*), and six such alleles, SLICK1 to SLICK6, have been published. Breed surveys found several in the same Criollo breeds. I screened 1,880 phased genomes for all six, checking most carrier calls against reads. Four were found in Criollo populations: SLICK2 on 22 chromosome copies, SLICK1 on five, SLICK6 on three, and SLICK3 on one. SLICK4 was absent, and SLICK5 also occurs in European and African cattle not known to be slick, so its use as a marker awaits phenotypic confirmation. Two animals carry a different slick allele on each chromosome copy, one phased directly from reads. A single-allele test counts one copy in these animals and misses carriers of untested alleles. A SLICK2-only test detects 22 of the 31 carrier chromosomes. All four alleles sit on European taurine DNA, Iberian or continental, but so do most Criollo chromosomes without them. A classifier suggesting otherwise rested on two linked sites. Ancestry therefore does not flag carriers. Even without thermal phenotypes, which I did not measure, heat tolerance breeding programmes for Criollo cattle should genotype all four.

Keywords: heat tolerance; slick coat; prolactin receptor (*PRLR*); Criollo cattle; marker-assisted selection; local ancestry

Introduction

Heat stress is one of the most consequential constraints on cattle production in tropical regions, reducing reproductive performance, impairing feed conversion, and cutting yields of both milk and meat [1]. A handful of breeds keep a short, sleek coat year-round in the humid lowlands where most *Bos taurus* herds struggle to thermoregulate. Affected animals carry a truncated form of the prolactin receptor (*PRLR*), which signals through the Janus kinase and signal transducer and activator of transcription (JAK-STAT) pathway, and the truncation appears to enhance body heat dissipation [2]. The slick locus is among the few major-effect loci for heat tolerance

identified in cattle, and the coat is inherited as a dominant trait [3]. Under heat stress, slick-coated Criollo Limonero heifers keep a lower rectal temperature and respiration rate than normal-haired herd mates [4], and Holstein cows carrying an introgressed slick allele regulate body temperature better and sweat more than non-carriers [5,6]. In those cows, the summer decline in milk yield is smaller [6]. Holstein heifers carrying SLICK1 also show a lower rectal temperature under humid heat [7], and their sweat glands hold twice as much FOXA1, a factor required for sweating in mice, although the area of skin the glands occupy does not change [8]. Slick alleles have been introgressed into dairy cattle [9,10], the case has been made for extending them to other temperate dairy breeds [11], and comparable *PRLR* truncations have more recently been introduced into Angus and Jersey cattle by gene editing [12].

At least six independent truncating mutations have been described in the final coding exon of the gene [2,13,14], an apparent case of convergent adaptation at the molecular level. Two of these mutations, SLICK1 (c.1382del, p.Ala461ValfsTer2) and SLICK2 (c.1489C>T, p.Arg497Ter), both segregate in the same Caribbean and South American Criollo populations; both abolish the distal tyrosines required for STAT5 docking while sparing the proximal Box1 motif that engages JAK2 [13]. A third allele, SLICK3 (c.1394C>A, p.Ser465Ter), has since been identified in one of those populations, Criollo Limonero [13,15]. Breed-level surveys have already found several of the six in the same breed. Sanger sequencing of slick-coated animals recorded SLICK1, SLICK2, SLICK5, and SLICK6 in both Hartón del Valle and Costeño con Cuernos [14]. Genotyping of all six positions in 14 Criollo and Criollo-derived breeds found the same pattern across more breeds, and showed that chips restricted to SLICK1 and SLICK2 misclassified carriers [15]. Phased whole genomes can answer three further questions: whether two alleles in one animal lie on the same chromosome or on opposite ones, how many carriers a single-position assay detects, and on what haplotype background each allele sits.

Convergence indicates the number of times the receptor was disrupted but not the origin of each disruption. Modern Criollo cattle descend from Iberian founders that began arriving in the Americas in 1493, when the first bovines landed on Hispaniola, and spread progressively across the Antilles and mainland throughout the sixteenth century [16,17]. These Iberian founders were not genetically homogeneous, and autosomal and mitochondrial markers documented historical introgression from African cattle into Iberian herds well before the colonial crossing [18,19]. A separate and later African contribution to the Caribbean is traditionally attributed to the Senepol of St. Croix, formed in the early twentieth century from a presumed cross between Red Poll and N’Dama [20,21]. Single-nucleotide polymorphism (SNP) data from 147 Senepol attribute 0.6% African taurine ancestry to this breed [20]. Each of these historical channels represents a plausible conduit through which a slick allele could have entered the Criollo gene pool. Distinguishing between them depends on the reference panels, and I tested each one by simulation before reading the carriers.

[22] reported SLICK2 in both pure taurine and zebu-admixed cattle from Burkina Faso, the first formal record of a slick allele in cattle native to Africa. They proposed two explanations: SLICK2 either originated independently within West African herds or was

transported there from South America during the colonial transatlantic cattle trade [22]. Both remain open.

The geographic origin of the slick alleles has been approached from several directions. [21] compared the genomic clustering of slick-coated Criollo breeds against non-slick ancestral populations in progressively narrower windows centred on the locus. The breed-specific ancestry signal of the Senepol dropped from 93% genome-wide to just 45% at the receptor locus itself [21], a pattern that points to two independently arisen mutations rather than a shared common ancestor. At that time, reference sequences for continental Criollo and Iberian cattle were not yet available, so clustering around the locus could not separate the flanking ancestry of a haplotype from the demographic history of the breed that carries it. [2] phased the causal SLICK1 mutation and ruled out *Bos indicus* introgression at the adaptive flank, which narrowed the question to European or ancient African taurine sources. [17] detected a composite selection signal at *PRLR* across Blanco Orejinegro and Hartón del Valle, and [10] tracked the introgressed segment around the receptor through absorbing crosses, using the two breeds of the directed cross as references. In the Montana Tropical composite, tropically adapted taurine ancestry also concentrates at the locus: the largest region in its top 1% spans 17.4 Mb of chromosome 20 and contains the slick locus [23]. Allele dating offers no alternative route. Linkage disequilibrium conflates haplotype age with the recent inbreeding history of the host animal, and both the mutation rate and the recombination clock are far too slow to disentangle those signals within the colonial timescale.

A more tractable approach is to ask not when a mutation arose, but which reference population the haplotype carrying it is most similar to. This comparison uses the alleles at flanking sites, which are not themselves under selection and therefore record the lineage of the haplotype that carries the mutation. For breeding, this answer decides how carriers can be found. If each allele sat on its own background, linked markers could flag its carriers, whereas a background shared with non-carriers leaves only the alleles themselves to genotype.

Two methodological requirements shape credible answers. Each carrier haplotype must be resolved directly at the mutation site rather than inferred by statistical imputation, which means preserving the causal slick variants through phasing and then returning carrier calls to the aligned reads wherever the alignments are held. The window used to characterise the background must be wide enough that no single variant can carry the answer, with reference panels large enough to recover their own simulated haplotypes before any query is interpreted. Neither requirement is visible in the output of a likelihood classifier, which reports a posterior whether or not the evidence behind it is distributed across sites.

Where the slick coat is selected through markers, the value of a marker panel depends on how fully it covers the alleles a population carries. Here, I screened all six published slick positions in 1,880 phased whole genomes, for which no thermal phenotypes are recorded, and returned to the aligned reads the eighteen of twenty-seven carrier calls for which I hold the alignments. I then asked what haplotype background the segregating alleles sit on, using three approaches: a coding-site classifier over 39.1 kb, local ancestry over 10 Mb, and direct haplotype distance. The three approaches

disagree, and the disagreement is itself informative, in that it shows how much of an ancestry assignment at a short window can rest on a single linked variant, and that six Iberian reference animals are too few to separate Iberian from continental European backgrounds.

Results

Four of the six published alleles segregate in these populations

All six published slick alleles lie within 208 bp of the final coding exon of *PRLR* on BTA20 (*Bos taurus* autosome 20) (Fig. 1a, b). Screening every one of them in the phased panel recovered four in Criollo animals. SLICK1 (c.1382del, p.Ala461ValfsTer2) appeared in five carrier haplotypes across four animals, SLICK2 (c.1489C>T, p.Arg497Ter) in 22 across 15, SLICK6 (c.1435C>T, p.Gln479Ter) in three across three, and SLICK3 (c.1394C>A, p.Ser465Ter) in one. All four occur in Limonero, in which 18 of the 24 animals screened carried at least one of them. SLICK1 was also present in both Costeño con Cuernos animals, and SLICK6 was present in one of the six Hartón del Valle (Fig. 2c). SLICK4 (c.1281C>G, p.Tyr427Ter) was monomorphic across all 1,880 animals at this site, consistent with its original description in a single Mexican Criollo Lechero Tropical animal [14], a population absent from this panel. Each allele truncates the cytoplasmic tail at a different point, removing between 85 and 155 residues and leaving the Box1 motif intact. All four remove the distal tyrosine that docks STAT5 in the rat receptor [24] (Fig. 1c).

The sixth position, chr20:39,099,228, carries SLICK5 (c.1396A>T, p.Lys466Ter) in 123 phased haplotypes across 109 animals. SLICK5 was originally described in slick-coated Criollo breeds [14], and [15] identified it as the most frequent slick allele in Blanco Orejinegro. Within the Criollo breeds, the calls here agree with that description, including nine of ten Blanco Orejinegro animals. Published frequencies vary between samples: [22] did not detect the change in 40 Blanco Orejinegro, and the frequency reported for Hartón del Valle is 0.24 in one survey [14] and 0.89 in another [22], differences that sampling of different herds can produce. Outside the Criollo breeds, the same change segregates in pure European taurines and occurs in one N'Dama and one Ogaden zebu. It never lies on a haplotype that carries SLICK1, SLICK2, SLICK3, or SLICK6. No coat phenotype is recorded for the animals in this panel, so whether the change segregates with the slick coat outside the breeds in which it was described cannot be tested here (Supplementary Note on SLICK5). An ancient origin, recurrent mutation at the same position, or penetrance that varies with genetic background could each plausibly produce the distribution seen here, and none can be excluded.

Carrier calls checked against reads, and two animals carrying two alleles

Imputation during phasing can create carrier genotypes that have no read support, and the alleles here are rare enough that a single spurious call can change a count. Of the 27 carrier calls, 18 came from the newly incorporated runs, for which I hold the alignments, and those were returned to the reads. Four were not supported by the

reads. Two SLICK2, one SLICK3, and one SLICK6 calls lacked the supporting alternative reads required (Fig. 2a; Table 2; Supplementary Methods). The counts above are those that remain.

The remaining nine calls came from Run 8, which I hold as called genotypes rather than alignments, and they could not be checked this way. All four SLICK1 carriers fall in that group, so the allele that carries the only African signal in these data is also the one whose calls rest on the published genotypes.

Two animals carry two different slick alleles. [15] suggested that some of the Senepol they genotyped are compound heterozygotes. In the Limonero ERR12035149 the two alleles, SLICK6 and SLICK2, sit on opposite chromosomes. The two positions lie 54 bp apart, within the span of a single 150 bp read, and none of the six reads covering both carries the two alternative alleles together: four carry SLICK6 alone and two SLICK2 alone (Fig. 2b). ERR12035149 is the animal LMN10 of ENA project ERP150979 (PRJEB65887), whose genomes [15] also genotyped; the reads resolve its phase. In SAMN08166129 the SLICK1 and SLICK2 alleles are likewise placed in trans, there by statistical phasing alone. A compound carrier of this kind presumably expresses the phenotype through two independent truncations and would be scored as a simple heterozygote by any assay targeting one position.

The coding classifier rests on a single pair of linked sites

Restricting the likelihood comparison to coding sites, as the informativeness of exonic markers would motivate, assigns the four alleles to different continental panels: SLICK1 to N'Dama in four of five haplotypes, SLICK2 to continental European in nineteen of twenty-two, and SLICK3 to the Iberian panel, as were two of the three SLICK6 haplotypes (Table 2).

This contrast does not appear to come from the coding core as a whole. Of the 56 coding sites, a single pair at chr20:39,077,999 and 39,078,000 carries between 59 and 85% of the total likelihood-ratio evidence for every one of the four alleles (Fig. 3b). The two positions are adjacent bases of the same codon, inherited together with a correlation of 1, and the annotation places one in the synonymous class and the other in the non-synonymous class.

The agreement between the two classes, which would otherwise read as independent corroboration, is essentially this one variant counted twice.

The alternative allele at this pair reaches 0.79 to 1.00 in the African panels and 0.12 to 0.15 in the European panels (Fig. 3a). Four of the five SLICK1 haplotypes carry it; every SLICK2, SLICK3, and SLICK6 haplotype carries the reference base. Removing the pair and repeating the assignment changes the result. SLICK2 moves to N'Dama in nineteen of twenty-two haplotypes, and SLICK1 to Iberian or N'Dama depending on the class (Fig. 3c). Resampling the sites 2,000 times keeps the observed continental group in 71% to 90% of replicates for SLICK1 and in 59% to 74% for SLICK2 (Fig. 3d). The SLICK2 carriers from Run 8 and from ERP150979, the only allele drawn from both sources, return the same panel, so the assignment does not follow the sequencing source (Supplementary Note on classifier robustness). The reference base at this pair is also present in 21% of

N’Dama haplotypes, so carrying it probably does not by itself separate a European background from an African one.

Local ancestry over 10 Mb, and what the reference panels resolve

A window wide enough to carry ancestry information independent of any single variant was genotyped directly from the aligned reads for the 38 newly incorporated animals and seven Asturiana genomes; the 40 with at least 20% of sites called were phased together with the 1,842 Run 8 genomes with the CLARITY genetic map [25] and submitted to RFMix. The genotypes were in agreement with the panel used for the haplotype analysis at 99.0 to 100 percent of shared sites.

The questions that the method can answer appear to depend on the reference configuration (Fig. 4a, b; Table 3). With four panels, simulated Iberian haplotypes were recovered as Iberian only 18% of the time and as continental European 66%. Adding a fifth, North and East Asian taurine panel did not help, and left the Iberian panel at 13%. Six Iberian reference animals were unlikely to support a distinction between Iberian and continental European backgrounds at this locus. Merging the two into a single European taurine panel appears to remove the problem. In the three-way configuration, every panel was recovered in 83% to 99% of simulations. Reference animals held out and run as queries recover their own panel in 20 of 20 continental European and 19 of 20 indicine haplotypes. Ankole, Boran, and Ogaden, African breeds of known taurine by indicine admixture, returned as mixtures rather than pure panels (Fig. 4c, d).

Read in that validated three-way configuration, which merges the Iberian and continental European panels, all four alleles sit on European taurine background at the locus, in all thirty-one carrier haplotypes, whether 15, 50 or 150 generations since admixture are assumed. So do the Criollo haplotypes that carry no slick allele, 60 of 67 at the same position (Fig. 5a shows twelve of them).

On this evidence, ancestry at the locus does not separate carriers from the Criollo haplotypes around them. The four-way and five-way configurations disagree with each other and with the three-way result for SLICK1, SLICK2, and SLICK6 (Fig. 5c), and that disagreement appears to track the panels that fail validation. I read the three-way result as the estimate these data support, on two grounds: its panels are the only ones that recover their own simulated haplotypes, and its window seems wide enough that no single variant can carry the answer.

Haplotype distance without a model

Any ancestry call will tend to name the least distant reference panel, whether or not a close relative of the query haplotype is present in it. Counting differences directly avoids this. Over the 766 SNPs within 50 kb of the locus, the SLICK2, SLICK3, and SLICK6 haplotypes lie 1 to 46 differences from their closest European taurine or Criollo neighbours, 23 of the 26 within 27, and 38 to 92 from the closest N’Dama haplotype (Fig. 5b). For two SLICK2 haplotypes, the single closest haplotype is indicine (Table 2). The SLICK1 haplotypes are the only ones with N’Dama among their nearest neighbours, at 21 and 37 differences for two of the five. The other three are closest to Criollo or American composite breeds. Widening the window to 250 kb brings Mongolian and

other North Asian taurine cattle in as nearest neighbours of several SLICK2 haplotypes, populations absent from the four-way reference, which may account for the indicine assignment that configuration returns.

Discussion

Four of the six published slick alleles segregated in the Criollo populations sampled here, and two animals carried two of them on opposite chromosomes. These data identify the alleles that a marker panel for heat tolerance must cover in the Criollo populations sampled here; measuring the effect of each allele on thermoregulation in herds under selection is the next step. An assay targeting a single position would score a compound carrier as a simple heterozygote and miss the carriers of the three alleles it does not interrogate. A test for SLICK2 alone detects 22 of the 31 carrier haplotypes, and because both compound carriers have SLICK2 in trans with another allele, it would count one slick copy in each. The coat is inherited as a dominant trait [3,10], so such an animal is expected to show it and to pass a slick allele to every offspring, where a simple heterozygote passes a slick allele to half of them. [15] found that chips restricted to SLICK1 and SLICK2 misclassified carriers, and here such a chip would miss the four SLICK3 and SLICK6 haplotypes. Adding SLICK6, present here on three haplotypes, to assays such as the SLICK1 to SLICK5 panel of [22] would cover the four alleles segregating in these herds. SLICK4, described in a single Mexican Criollo Lechero Tropical animal [14] and absent here, would matter for panels used in that population.

SLICK5 also segregates in European taurine breeds not known for the slick coat, so calling it there would count animals not known to show the phenotype as carriers. It is best added to a panel once its association with the coat has been confirmed within the breeds where it was described.

Gene editing now adds truncations of its own. CRISPR/Cas9 edits of *PRLR* in Angus and Jersey cattle improved the regulation of body temperature and growth, with no adverse effect detected on carcass traits or male reproduction [12]. A single gene-edited Angus sire, mosaic at the locus, passed five different slick truncations to his calves. He also passed a large deletion that leaves the receptor non-functional and gives the opposite phenotype, a heavier coat and a higher body temperature than slick heifers [26]. Gene editing is also being developed in bovine zygotes [27] and in dairy cattle cells and cloned embryos [28]. Therefore, edited and natural alleles may circulate in the same breeding populations, and a panel built for one truncation will miss the others, natural or edited.

A classifier restricted to coding sites assigns different continents to different alleles, and the assignment is driven by one pair of adjacent bases 21 kb from the locus. Two properties of that pair limit what it can show. The pair is in complete linkage, so the agreement between synonymous and non-synonymous classes is one observation, not two. Its reference base is common in N'Dama as well as in Europe, so carrying it is weak evidence either way. Over a window wide enough to escape that dependence, and in the only reference configuration whose panels are all recovered in simulation, all four alleles, including SLICK1, are assigned the European taurine background of the Criollo

genomes around them. Earlier evidence from Senepol points in the same direction. With BovineSNP50 genotypes, before *PRLR* was identified as the causal gene, [20] found a selection signature at the slick locus on BTA20. They noted a suggestive excess of local European ancestry at this footprint and proposed that the favourable variant sits on European haplotype backgrounds. Whether any of them arose on an African taurine background remains open on these data. The SLICK1 haplotypes retain what may be the only direct signal. Two of five have N'Dama among their nearest neighbours at 21 and 37 differences, and the African allele at the exon 2 pair. This may indicate West African ancestry at this locus. It is equally consistent with a European haplotype that happens to carry a widespread African allele. Separating the two requires more carrier haplotypes than four animals provide. All four SLICK1 carriers also come from Run 8, held as called genotypes, so the calls and their phase need confirmation in aligned data before this signal is read as ancestry.

The Iberian question cannot be answered using public data. Iberian cattle are the expected taurine source for Criollo populations, and the Iberian reference assembled here, six animals from five breeds, was recovered as itself in 18% of the simulations. Reporting an Iberian assignment from a panel that cannot recover its own simulated haplotypes would be reporting the panel, not the data. Poorly matched reference panels bias local ancestry inference in ways that the posterior alone does not reveal [29]. Closing this gap requires Iberian and Ibero-American whole genomes at a scale that the public archives do not currently hold. Whichever taurine source the alleles came from, the carriers share the European taurine background with 60 of the 67 non-carrier Criollo haplotypes, so linked markers cannot flag them, and the alleles themselves must be genotyped.

The methodological point probably generalises beyond this locus. The functional class is arguably not a unit of independent evidence when a codon carries two variants in complete linkage, and a likelihood classifier over a short window can rest almost its entire answer on a handful of sites without any summary statistic revealing it. Counting how many sites carry the assignment and repeating it with those sites removed is a cheap check for any short-window classifier.

Conclusions

Phased genomes add haplotype resolution to breed-level surveys of the slick alleles [14,15]: four truncating *PRLR* alleles lie on separate haplotypes within the same Criollo populations, all four of which are in Limonero, and two compound carriers are resolved in phase, one of them directly from the reads. A single-position assay counts one slick copy in such animals and misses the carriers of the alleles it does not test. Marker-assisted selection for the slick phenotype should therefore genotype the four segregating alleles, SLICK1, SLICK2, SLICK3, and SLICK6, rather than a single position, and add SLICK5 once it is shown to segregate with the coat. No coat or thermoregulatory phenotype was recorded for these animals, so the effect of each allele on heat tolerance in Criollo herds remains to be measured, including whether SLICK3 and SLICK6 act on thermoregulation as SLICK1 and SLICK2 do.

At the resolution supported by the available reference panels, the four alleles appear to sit on the European taurine background, Iberian or continental, that characterises the Criollo genomes carrying them. Distinguishing an Iberian from a continental European source and testing the West African background suggested for SLICK1 requires reference panels larger than those currently available in public archives.

Methods

Data and genomic quality control

Two independent datasets support the two halves of this study, and they are not interchangeable (Table 1). **Haplotype analysis at *PRLR*** used whole-genome sequences from Run 8 of the 1000 Bull Genomes Project [30], 1,842 animals aligned to ARS-UCD1.2. Within the gene window, no minor allele frequency filter was applied, for the reason given below; for the wider 10-Mb window, sites with a minor allele frequency below 0.01 or more than 10% missing genotypes were removed. Reference panels were built only from animals whose breed was declared in the BioSample record, which yielded 180 labelled animals across seven panels: European taurine (81 animals from Simmental, Charolais, Limousin, Salers, and Shorthorn, two of which were Limousin crosses), Criollo (50), N'Dama (17), Boran (11), Ogaden (9), Ankole (8), and Iberian (4: Sayaguesa, Maronesa, Limia, and Pajuna). Breed assignments inferred from accession ranges were discarded because they disagreed with both the declared breed and genetic pole.

The **genome-wide ancestry of the Senepol**, a supervised admixture analysis of the BovineSNP50 genotypes of [19] against lineage-matched controls, is reported in the Supplementary Note on Senepol ancestry. This addresses a different question from the haplotype analysis and rests on a different resource, and the estimate it returns sits below the detection limit of the estimator, so it is presented as a bound rather than as a point estimate.

Three supplementary sources extend the dataset. Fifty-six whole-genome sequencing runs were incorporated, forty-nine of them runs of Criollo cattle from ENA study ERP150979 (project PRJEB65887) [17], covering thirty-one animals after technical duplicates were collapsed, together with seven supporting N'Dama runs (ENA PRJEB36894, PRJEB39353 and PRJEB39924; SRA PRJNA604048). All were processed through the same quality-control pipeline. Senepol genotypes at the focal *PRLR* sites were called from the liver RNA-seq of GEO series GSE192471. Transcriptome reads establish which allele an animal carries, but they cover only expressed sequence and so cannot phase the surrounding haplotype; the Senepol therefore contributes genotypes to this study and no carrier haplotype. Slick Holstein RNA-seq deposited in the SRA was not used, in either the genotype calls or the reference panels, because those animals carry introgressed slick alleles of documented recent transfer and would place a carrier haplotype in a background unrelated to the colonial-era question examined here. Tract-length analysis used the same 1,842 Run 8 genomes, genotyped at 108,388 biallelic SNPs across a 10 Mb window of BTA20 centred on *PRLR* after the same quality filters. No array panel enters this analysis.

Genomic phasing and haplotype assignment by functional class

Phasing across the *PRLR* was performed in BEAGLE 5.4 with the imputation module disabled and no external genetic map; therefore, the reconstruction rested on sequencing data alone [31]. Disabling imputation did not leave genotypes missing: BEAGLE returned a complete matrix, filling missing genotypes from the haplotypes of the other animals in the sample rather than from an external reference panel. Before phasing, the called sites of the annotated set carried 0.78% missing genotypes in the Run 8 animals and 4.25% in the newly incorporated ones. Haplotypes at those positions were therefore statistically reconstructed rather than observed. Switch errors in that reconstruction would mix the flanking backgrounds of an animal's two haplotypes and were not measured here. The phasing run did not output per-site phase probabilities, and returning a carrier call to the reads checks its genotype, not its phase; phase was read directly only in ERR12035149, where two focal sites fall within single reads. The genome-wide minor-allele-frequency filter removes rare sites that are more likely to be genotyping errors, but it would also remove the SLICK1 and SLICK2 sites, at frequencies of 0.0014 and 0.0019 in Run 8. It was therefore not applied within the gene window, which raised the number of usable sites from 316 to 1,454 across 39.1 kb; the SLICK3 and SLICK6 sites enter with the newly incorporated runs (Fig. 1b). Biallelic sites were retained when fewer than 15% of genotypes were missing and annotated using bcftools csq against Ensembl structural definitions [32]. The phased matrix also carried 37,683 positions called only in the incorporated runs, 37,298 of which were invariant there, at which every Run 8 genotype was filled by BEAGLE. The coding and intronic classes were scored on the 1,454 annotated sites alone. Five annotated sites exceeded 10% missing before phasing: three between 13.5% and 14.2%, and two (chr20:39,085,837 and chr20:39,090,123) at which the Run 8 record is an indel and the merged matrix holds only the SNP called in the incorporated runs, so that every Run 8 genotype there was filled by BEAGLE. Each carrier haplotype was scored by the log-likelihood of observed alleles under population-specific frequency estimates bounded to the interval between 10^{-4} and $1-10^{-4}$, so that private alleles could not dominate the sum; the posterior probability was derived by normalising those likelihoods across competing panels. Animals carrying the target mutations were excluded from the frequency estimates of the reference panels used to evaluate them, preventing a circular comparison.

Informativeness tests, bootstrap and background homozygosity control

The discriminatory power of each exonic marker was quantified using Rosenberg's I_n informativeness statistic [33], with stability confirmed by site restriction and direct weighting on I_n. Geographic assignment stability was further evaluated by resampling with replacement over 2,000 independent draws, taken within each functional class in isolation rather than across the gene. For the background homozygosity control, the physical extent of the linkage-disequilibrium tract at *PRLR* was benchmarked against tract lengths at 150 random positions within the same 10 Mb window on BTA20, each at least 1 Mb from the *PRLR* coordinates; the empirical P value is the fraction of random loci whose shared blocks equalled or exceeded the focal tract. Global ancestry

proportions were estimated using ADMIXTURE 1.3.0 under supervised mode at K = 3 [34,35], with an independent cross-check provided by the sNMF algorithm of the LEA package [36,37].

The posterior was normalised separately within each class of functional consequence, so that the value reported for a haplotype in a given class is the share of the total likelihood of that class held by the winning panel. Panels entered the comparison only with at least eight haplotypes, which left European taurine (162 haplotypes), N'Dama (34), Ankole (16), Boran (22), Ogaden (18), Criollo (100) and Iberian (8).

Validation of carrier calls against the aligned reads

Phasing fills missing genotypes; therefore, every carrier call for which I hold the alignments was returned to the reads before it was counted, on the assumption that an imputed carrier cannot be trusted. The nine calls from Run 8, held as called genotypes, could not be checked this way. Read depth and allele counts at each focal position were taken with samtools mpileup at a minimum mapping and base quality of 20, with base alignment quality recalibration disabled. A call was retained when at least three reads carried the alternative allele and those reads were at least a quarter of the depth at the site. A heterozygote is expected near half the reads; therefore, a quarter of the depth allows for unequal sampling of the two alleles at moderate depth, and three reads exclude calls resting on one or two erroneous bases. Zygosity was not reassigned from the reads; therefore, every retained call keeps its phased genotype, including one SLICK2 homozygote resting on three reads, where the only read carrying the reference allele had a base quality of 11, below the threshold of 20. Where two focal positions fell within the span of one 150 bp read, phase was determined directly by walking the CIGAR (Compact Idiosyncratic Gapped Alignment Report) string of every read covering both, which applies to the SLICK6 and SLICK2 pair 54 bp apart.

Local ancestry over a 10 Mb window

The 39.1 kb gene window appears to carry too few independent sites to establish haplotype background; therefore, ancestry was also inferred over chr20:34,100,000-44,100,000 of the ARS-UCD1.2 assembly [38]. The 45 animals absent from Run 8 (31 Criollo, seven N'Dama, and seven Asturiana) were genotyped directly from their alignments at the 108,388 biallelic Run 8 sites. The genotype was chosen by likelihood under a base error rate of 0.01, and emitted only where at least four reads covered the site and the best genotype led the next by a log10-likelihood of at least one. The remainder was left missing, and the 40 animals with at least 20% of sites called were combined with Run 8 for phasing. The combined matrix of 1,882 animals was phased in BEAGLE 5.4 against the CLARITY genetic map for BTA20 [25], and local ancestry was inferred with RFMix 2.03 [39] at 50 generations since admixture. At a generation interval near five years, the colonial founding lies roughly 100 generations back, and admixture within the Criollo breeds continued after it; therefore, 50 is an intermediate value. The three-way and four-way configurations were repeated at 15 and 150 generations. In the three-way configuration, every carrier haplotype stayed on the European taurine panel at *PRLR* at all three values, the non-carrier Criollo background did not change, and the held-out European and indicine haplotypes recovered their own panel in 20 of 20 and in

19, 19, and 17 of 20 at 15, 50, and 150 generations. In the four-way configuration, the modal panel of SLICK1 and SLICK2 did not change; that of SLICK3, on one haplotype, moved at 15 generations, and that of SLICK6, on three, at 150.

Three reference configurations were run. The four-way configuration separated continental European taurine, Iberian taurine, N'Dama, and indicine. The five-way added North and East Asian taurine, and the three-way merged the two European panels. Panels of more than 40 animals (European taurine, 81; indicine, 88) were subsampled to 40, and ten of the animals left over from each were run as queries, so that each had to recover its own panel. The Asian taurine panel of the five-way configuration entered whole, at 41 animals. Ankole, Boran, and Ogaden were run as queries throughout as a positive control for admixture. The 49 Criollo animals run as queries carried 98 haplotypes, 31 of them with a slick allele; the other 67 formed the non-carrier comparison set. The European taurine, Criollo, N'Dama, Ankole, Boran, and Ogaden labels, and four of the six Iberian reference animals, came from the same source as the haplotype analysis. The indicine and Asian taurine labels came from the curated provenance table for Run 8, and the other two Iberian animals were Asturiana genomes (ENA PRJEB38981) with at least 40% of sites called.

Dependence of the coding classifier on individual sites

For each allele, the contribution of every coding site to the log likelihood ratio between the best-fitting European and African panels was computed on the dominant coding configuration, and the sites were ranked by the absolute value of that contribution. The assignment was then repeated with the two exon 2 sites removed. Model-free distance was measured as the number of differences between each carrier haplotype and every other phased haplotype in windows of ±25, ±50, ±100, and ±250 kb around the locus, excluding the focal positions themselves and the other carriers of the same allele.

Ethics statement

No animals were handled in this study. All genotypes and sequence reads were obtained from public repositories, listed under Data availability.

Use of large language models

Claude (Anthropic), including the Claude Code environment, was used to assist with writing and debugging the analysis code and editing the manuscript text. All methodological decisions, interpretations, and conclusions are the author's own, and every reported value, estimate, and citation was checked against its primary source.

Data availability

No new sequencing or array genotype data were generated in this study. All datasets analysed are public and were produced by third parties. Run 8 of the 1000 Bull Genomes Project is available in the European Nucleotide Archive (ENA) under accession PRJEB42783. The Criollo runs added to the Run 8 panel are available in ENA under ERP150979 (project PRJEB65887), and the N'Dama runs in ENA under

PRJEB36894, PRJEB39353, and PRJEB39924 and in the NCBI Sequence Read Archive (SRA) under PRJNA604048. The Asturiana genomes used as Iberian references are in ENA under PRJEB38981. The BovineSNP50 genotypes used in the Supplementary Note on Senepol ancestry were generated and deposited by Decker et al. [19] in the Dryad Digital Repository, where they are openly available (https://doi.org/10.5061/dryad.th092). The Senepol genotypes at the focal sites derive from NCBI Gene Expression Omnibus (GEO) series GSE192471. The CLARITY physical-genetic map is available in Zenodo (https://doi.org/10.5281/zenodo.17909700). All analysis code, configuration files, intermediate result tables, and the scripts that generate every figure are openly available in Zenodo (https://doi.org/10.5281/zenodo.22184899) and mirrored on GitHub (https://github.com/luizdione/Paper_slick_ancestry).

References

1. Thornton, P., Nelson, G., Mayberry, D. & Herrero, M. Impacts of heat stress on global cattle production during the 21st century: A modelling study. *The Lancet Planetary Health* **6**, e192–e201 (2022).

2. Littlejohn, M. D. *et al.* Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nature Communications* **5**, 5861 (2014).

3. Olson, T. A., Lucena, C., Chase, C. C. & Hammond, A. C. Evidence of a major gene influencing hair length and heat tolerance in Bos taurus cattle. *Journal of Animal Science* **81**, 80–90 (2003).

4. Landaeta-Hernández, A. J. *et al.* Heat stress response in slick vs normal-haired Criollo Limonero heifers in a tropical environment. *Tropical Animal Health and Production* **53**, 445 (2021).

5. Dikmen, S. *et al.* Differences in thermoregulatory ability between slick-haired and wild-type lactating Holstein cows in response to acute heat stress. *Journal of Dairy Science* **91**, 3395–3402 (2008).

6. Dikmen, S. *et al.* The SLICK hair locus derived from Senepol cattle confers thermotolerance to intensively managed lactating Holstein cows. *Journal of Dairy Science* **97**, 5508–5520 (2014).

7. Carmickle, A. T. *et al.* Physiological responses of Holstein calves and heifers carrying the SLICK1 allele to heat stress in California and Florida dairy farms. *Journal of Dairy Science* **105**, 9216–9225 (2022).

8. Sosa, F. *et al.* Effects of the bovine SLICK1 mutation in PRLR on sweat gland area, FOXA1 abundance, and global gene expression in skin. *Journal of Dairy Science* **105**, 9206–9215 (2022).

9. Hansen, P. J. Prospects for gene introgression or gene editing as a strategy for reduction of the impact of heat stress on production and reproduction in cattle. *Theriogenology* **154**, 190–202 (2020).

10. Zayas, G. A., Dikmen, S., Mateescu, R. G. & Hansen, P. J. Maintaining breed integrity: Successful introgression of the SLICK1 allele into the Holstein breed. *Journal of Heredity* **116**, 216–224 (2025).

11. Davis, S. R., Spelman, R. J. & Littlejohn, M. D. Breeding heat tolerant dairy cattle: The case for introgression of the "slick" prolactin receptor variant into Bos taurus dairy breeds. *Journal of Animal Science* **95**, 1788–1800 (2017).

12. Cuellar, C. J. *et al.* Consequences of gene editing of PRLR on thermotolerance, growth, and male reproduction in cattle. *FASEB BioAdvances* **6**, 223–234 (2024).

13. Porto-Neto, L. R. *et al.* Convergent evolution of slick coat in cattle through truncation mutations in the prolactin receptor. *Frontiers in Genetics* **9**, 57 (2018).

14. Flórez Murillo, J. M. *et al.* Three novel nonsense mutations of prolactin receptor found in heat-tolerant Bos taurus breeds of the Caribbean Basin. *Animal Genetics* **52**, 132–134 (2021).

15. Flórez Murillo, J. M. *et al.* Allele frequencies of PRLR truncating variants in heat-tolerant taurine cattle native to Latin America. *Animal Genetics* **55**, 910–913 (2024).

16. McTavish, E. J., Decker, J. E., Schnabel, R. D., Taylor, J. F. & Hillis, D. M. New World cattle show ancestry from multiple independent domestication events. *Proceedings of the National Academy of Sciences* **110**, E1398–E1406 (2013).

17. Ward, J. A. *et al.* Genomic insights into the population history and adaptive traits of Latin American Criollo cattle. *Royal Society Open Science* **11**, 231388 (2024).

18. Beja-Pereira, A. *et al.* The origin of European cattle: Evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences* **103**, 8113–8118 (2006).

19. Decker, J. E. *et al.* Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genetics* **10**, e1004254 (2014).

20. Flori, L. *et al.* A quasi-exclusive European ancestry in the Senepol tropical cattle breed highlights the importance of the slick locus in tropical adaptation. *PLoS ONE* **7**, e36133 (2012).

21. Huson, H. J. *et al.* Genome-wide association study and ancestral origins of the slick-hair coat in tropically adapted cattle. *Frontiers in Genetics* **5**, 101 (2014).

22. Gebeyehu, S. *et al.* Analysis of SLICK allele in African taurine and Zebu cattle breeds. *Animal Genetics* **56**, e13499 (2025).

23. dos Santos, C. A. *et al.* Genomic analysis of taurine and indicine ancestry in the Montana Tropical composite population. *Journal of Animal Breeding and Genetics* **143**, 273–282 (2025).

24. Pezet, A., Ferrag, F., Kelly, P. A. & Edery, M. Tyrosine docking sites of the rat prolactin receptor required for association and activation of Stat5. *Journal of Biological Chemistry* **272**, 25043–25050 (1997).

25. Melzer, N., Qanbari, S., Ding, X. & Wittenburg, D. CLARITY: A Shiny app for interactive visualisation of the bovine physical-genetic map. *Frontiers in Genetics* **14**, 1082782 (2023).

26. Cuellar, C. J. *et al.* Impact of mutations in PRLR for hair weight, circulating prolactin concentrations, feed intake, and growth of beef cattle in a hot and humid climate. *Frontiers in Endocrinology* **17**, 1907038 (2026).

27. Wei, J. *et al.* Cytoplasmic injection of zygotes to genome edit naturally occurring sequence variants into bovine embryos. *Frontiers in Genetics* **13**, 925913 (2022).

28. Cheng, X. *et al.* Regulation of the homeostasis of early embryo development in dairy cows by targeted editing of the PRLR gene-mediated activation of the anti-heat stress pathway. *Cells* **14**, 1856 (2025).

29. Browning, S. R., Temple, S. D. & Browning, B. L. Local ancestry inference with poorly-matched reference panels. *PLOS Genetics* **22**, e1011919 (2026).

30. Hayes, B. J. & Daetwyler, H. D. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annual Review of Animal Biosciences* **7**, 89–102 (2019).

31. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* **108**, 1880–1890 (2021).

32. Danecek, P. & McCarthy, S. A. BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics* **33**, 2037–2039 (2017).

33. Rosenberg, N. A., Li, L. M., Ward, R. & Pritchard, J. K. Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics* **73**, 1402–1422 (2003).

34. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–1664 (2009).

35. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

36. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).

37. Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* **6**, 925–929 (2015).

38. Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience* **9**, giaa021 (2020).

39. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *The American Journal of Human Genetics* **93**, 278–288 (2013).

Acknowledgements

I thank the 1000 Bull Genomes consortium and the authors who deposited the sequence data analysed here. I also thank Sheila Albert dos Reis for reviewing earlier versions of this manuscript. She is listed as a co-creator of the Zenodo archive described under Data availability.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. It was carried out with the institutional support of the Federal Institute of Education, Science and Technology of Rio de Janeiro (IFRJ).

Author contributions

L.D.B.D.: conceptualization, methodology, software, formal analysis, investigation, data curation, validation, writing (original draft), and writing (review and editing).

Competing interests

The author declares no competing interests.

Figures

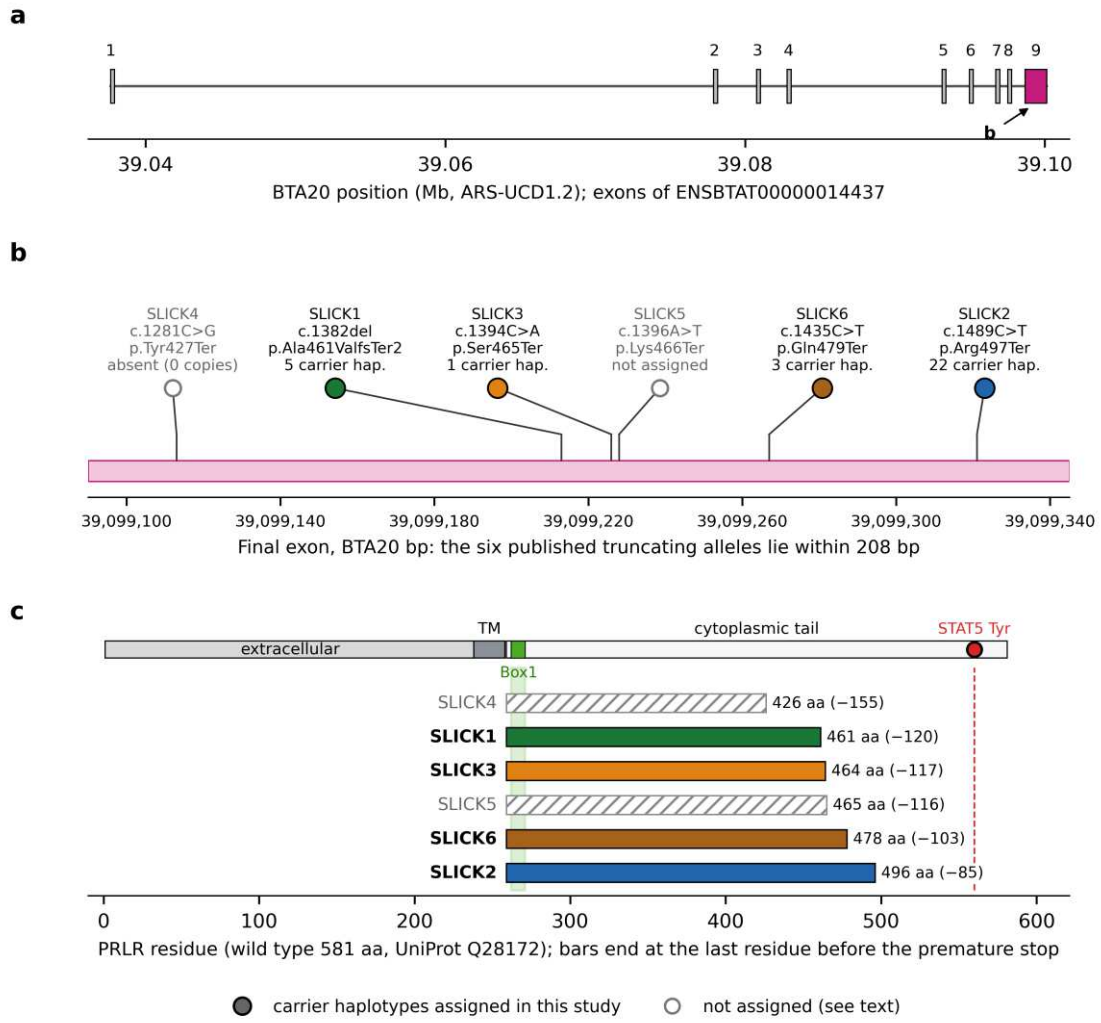


Figure 1. *The PRLR locus and the six published slick alleles. (**a**) Gene model of ENSBTAT00000014437 on BTA20 (ARS-UCD1.2); the final coding exon is highlighted and the window of panel **b** is arrowed. (**b**) The six published truncating alleles, all within 208 bp of the final exon, with HGVS nomenclature and the number of carrier haplotypes recovered here. SLICK4 is present as a site but monomorphic; SLICK5 also occurs outside the breeds expressing the phenotype and is not assigned. (**c**) The 581-residue receptor, with the transmembrane segment, the Box1 motif that engages JAK2 and the distal tyrosine that docks STAT5; bars give the cytoplasmic tail retained by each allele. Box1 and the STAT5 tyrosine are placed by homology with the human receptor and are not UniProt annotations.*

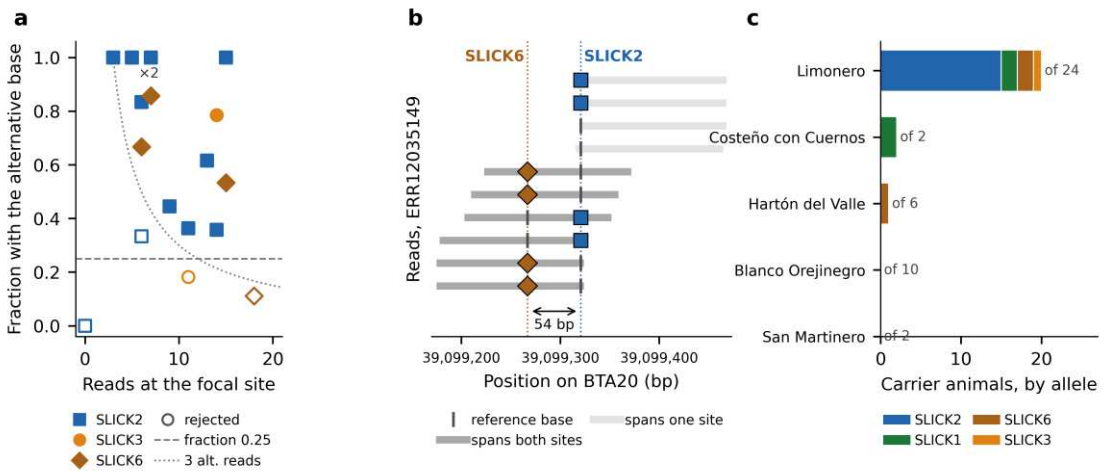


Figure 2. *Evidence from the reads. (**a**) Read support for the 18 carrier calls with alignments: fraction of reads carrying the alternative base against the number of reads at the focal site. Open symbols are the four calls rejected; the dashed line marks the minimum alternative fraction of 0.25 and the dotted curve the minimum of three alternative reads; two SLICK2 calls coincide (×2). (**b**) Reads of ERR12035149 over the SLICK6 and SLICK2 positions, 54 bp apart. Symbols mark the alternative base and ticks the reference base. Of the six reads that span both positions, four carry SLICK6 alone and two SLICK2 alone, so the two alleles lie on opposite chromosomes. (**c**) Carrier animals of each allele by Criollo breed, beside the number of animals of that breed in the phased panel; an animal carrying two alleles is counted under each.*

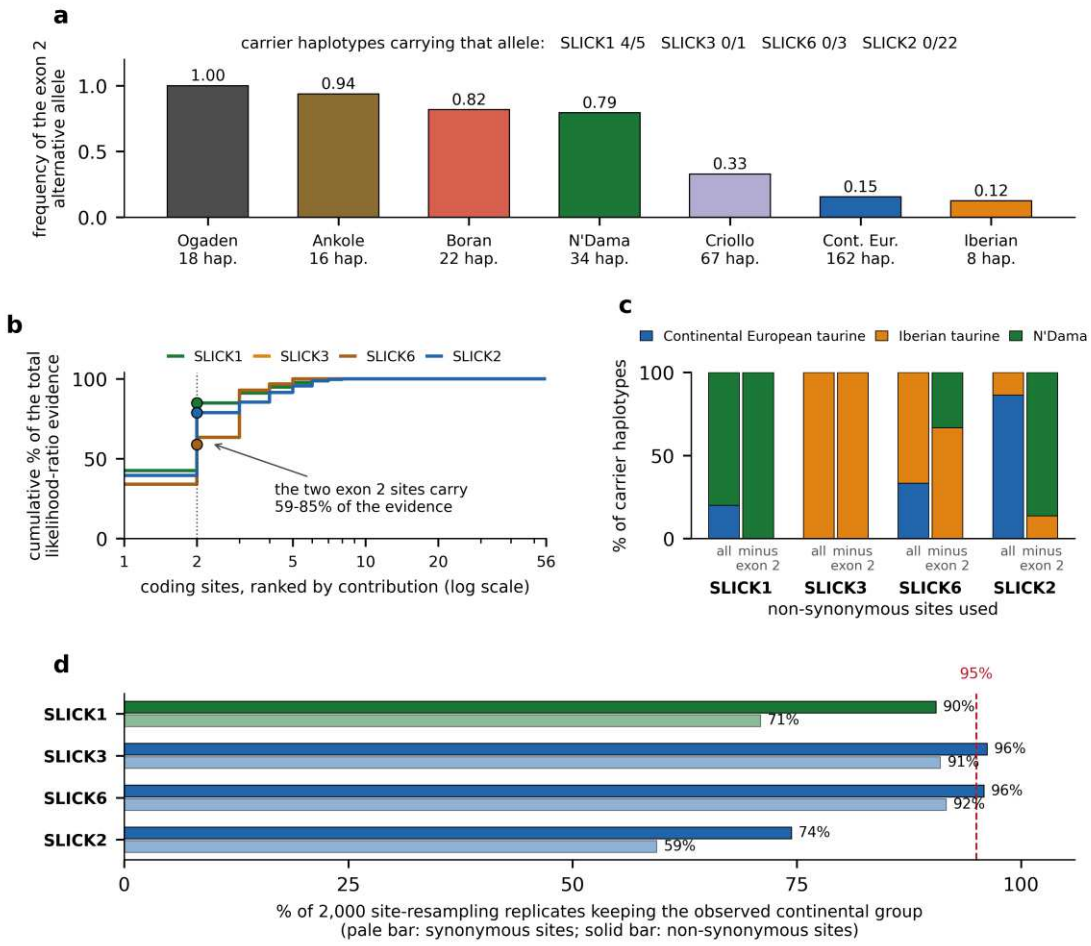


Figure 3. *The coding classifier rests on one pair of linked sites 21 kb from the locus. (**a**) Frequency of the alternative allele at chr20:39,077,999 in each reference panel, excluding carriers, with the number of haplotypes per panel; above, the number of carrier haplotypes of each slick allele that carry it. (**b**) Coding sites ranked by their contribution to the European-over-African log-likelihood ratio, as a cumulative share of the total; the two exon 2 sites, which are adjacent bases of one codon inherited together, carry 59 to 85% of the evidence for every allele. (**c**) Panel assigned to each allele at non-synonymous sites, with all coding sites and with the exon 2 pair removed. (**d**) Share of 2,000 site-resampling replicates that keep the observed continental group.*

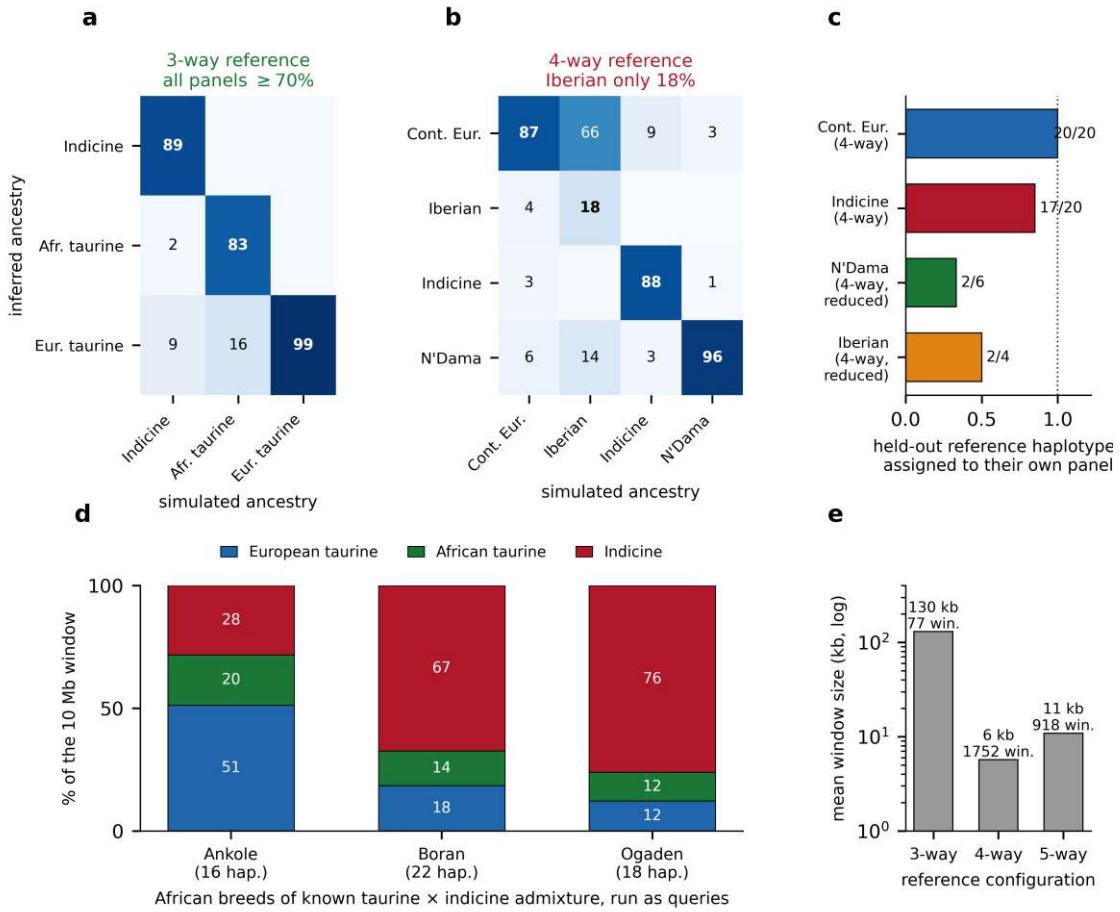


Figure 4. *What the ancestry inference can and cannot resolve, before any carrier is read. (**a**, **b**) Internal simulation of RFMix: haplotypes are simulated from the reference panels and reclassified, so the column gives the simulated ancestry and the row the inferred one. The three-way configuration recovers every panel; in the four-way configuration simulated Iberian haplotypes are recovered as Iberian 18% of the time and as continental European 66%. (**c**) Reference animals held out of the panels and run as queries, and the share of their haplotypes returned to their own panel at the locus. (**d**) Positive control: three African breeds of known taurine by indicine admixture, run as queries, return as mixtures. (**e**) Mean window size of each configuration over the 10 Mb interval.*

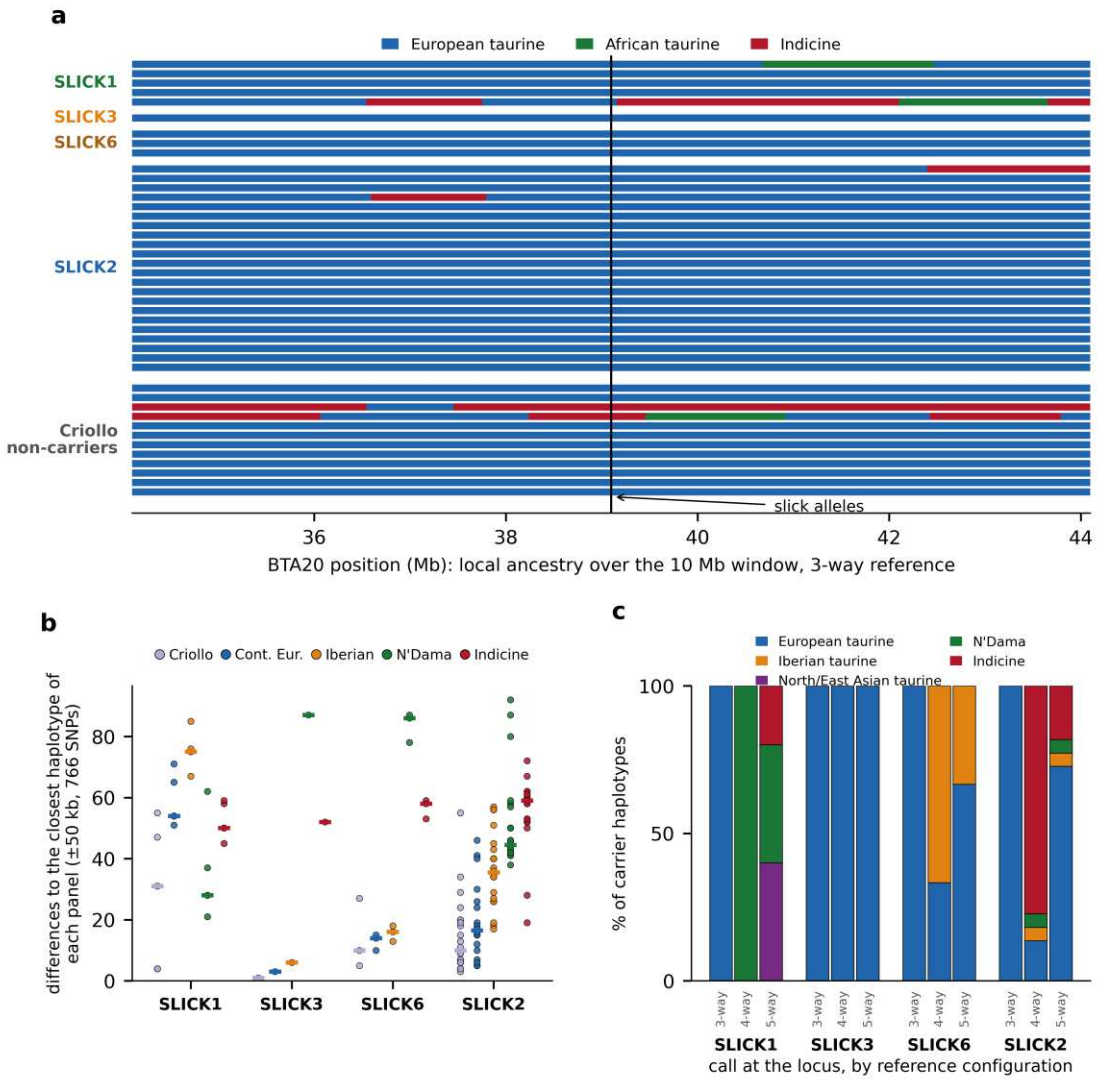


Figure 5. *Haplotype background over a window wide enough to escape any single variant. (**a**) Local ancestry across the 10 Mb interval for every carrier haplotype and for twelve Criollo haplotypes carrying none of the four segregating alleles, in the three-way configuration, the one that passes validation in Fig. 4a. (**b**) Differences between each carrier haplotype and the closest haplotype of each panel within 50 kb of the locus, over 766 SNPs; bars give the median. (**c**) Ancestry called at the locus under the three reference configurations; the configurations that fail validation are also the ones that disagree.*

Tables

Table 1. *Source of each dataset and the analysis it supports. The two halves of the study rest on different resources, and the reference panels are built only from animals whose breed is declared in the BioSample database.*

Source	Data type	Contribution	Analysis it supports
1000 Bull Genomes, Run 8 (PRJEB42783)	Whole-genome sequence, ARS-UCD1.2	1,842 animals	*PRLR* phasing and haplotype assignment; local ancestry and tract length over a 10 Mb window on BTA20
ENA ERP150979 (PRJEB65887)	Whole-genome sequence	49 runs, 31 animals after technical duplicates were collapsed	Criollo carriers of SLICK2, SLICK3 and SLICK6; Criollo reference panel
ENA PRJEB36894, PRJEB39353, PRJEB39924; SRA PRJNA604048	Whole-genome sequence	7 animals	N'Dama reference panel
ENA PRJEB38981	Whole-genome sequence	7 Asturiana animals genotyped from their alignments; 4 phased, 2 with at least 40% of sites called used as references	Iberian reference panel for local ancestry
CLARITY, doi:10.5281/zenodo.17909700 [25]	Physical-genetic map, BTA20	Genetic positions for the 10 Mb window	Phasing and local ancestry over 10 Mb
GEO GSE192471	Liver RNA-seq	7 animals	Senepol genotypes at the focal sites only; contributes no carrier haplotype
[19], Dryad `10.5061/dryad.th092`	BovineSNP50 array	753 animals, 29,869 markers after quality control	Genome-wide ancestry: supervised ADMIXTURE and sNMF

Reference panels drawn from the phased dataset, by declared breed: European taurine 81 animals (Simmental, Charolais, Limousin, Salers, Shorthorn, two of them Limousin crosses), Criollo 50, N'Dama 17, Boran 11, Ogaden 9, Ankole 8, Iberian 4 (Sayaguesa,

Maronesa, Limia, Pajuna). Of the 1,880 phased animals, 180 carry a panel label; the remainder provide phasing context and carrier screening only. Breed assignments inferred from accession ranges were discarded, because they disagree with both the declared breed and the genetic pole.

Table 2. *The four segregating alleles: carriers, the number of calls that could be checked against the aligned reads and the number of those that were rejected (calls from Run 8 have no alignments to check), the number of distinct coding configurations they occupy (the effective number of independent observations), and the haplotype background returned by three independent approaches. Panel names follow Table 1.*

Allele	Position (BTA20)	HGVS	Carrier haplotypes / animals	Calls checked on reads / rejected	Distinct coding configurations	39 kb coding classifier	Local ancestry, 10 Mb (3-way)	Closest panel (±50 kb), by haplotype
SLICK1	39,099,213	c.1382del, p.Ala461ValfsTer2	5 / 4	0 of 4 / –	2	N'Dama 4, Continental European 1	European taurine 5	Criollo 3 (4-55 diff.), N'Dama 2 (21-37 diff.)
SLICK3	39,099,226	c.1394C>A, p.Ser465Ter	1 / 1	2 of 2 / 1	1	Iberian 1	European taurine 1	Criollo 1 (1 diff.)
SLICK6	39,099,267	c.1435C>T, p.Gln479Ter	3 / 3	4 of 4 / 1	2	Iberian 2, Continental European 1	European taurine 3	Criollo 2 (5-10 diff.), Continental European 1 (14 diff.)
SLICK2	39,099,321	c.1489C>T, p.Arg497Ter	22 / 15	12 of 17 / 2	5	Continental European 19, Iberian 3	European taurine 22	Criollo 12 (3-34 diff.), Continental European 8 (5-46 diff.), Indicine 2 (19-28 diff.)

Table 3. *Validation of local ancestry inference. Simulation accuracy is the share of haplotypes simulated from each reference panel that RFMix returns to that panel. Only the three-way configuration recovers every panel; the Iberian panel, six animals from five breeds, does not recover itself.*

Reference configuration	Internal simulation accuracy per panel (%)	Weakest panel	Held-out reference haplotypes recovered	Mean window
3-way (European taurine / African taurine / indicine)	Indicine 89; African taurine 83; European taurine 99	African taurine 83%, mistaken for European taurine	Continental European 20/20; Indicine 19/20	130 kb
4-way (continental European / Iberian / N'Dama / indicine)	Continental European 87; Iberian 18; Indicine 88; N'Dama 96	Iberian 18%, mistaken for Continental European	Continental European 20/20; Indicine 17/20	6 kb
5-way (4-way plus North and East Asian taurine)	Asian taurine 97; Continental European 80; Iberian 13; Indicine 92; N'Dama 92	Iberian 13%, mistaken for Continental European	n.a.	11 kb

Supplementary information

Supplementary Note on SLICK5. Distribution of SLICK5 across the panel and what is needed before it enters a marker panel.

Supplementary Methods. Validation of carrier calls against the aligned reads, the criteria applied, and the read-backed phase of the compound carrier.

Supplementary Note on Senepol ancestry. Genome-wide ancestry of the Senepol against lineage-matched controls, with the detection limit of the estimator.

Supplementary Note on classifier robustness. Informativeness test, assignment by distance to the causal site, panel rarefaction and batch separation.

Supplementary Tables S1–S6. Read support for each carrier call (S1), every carrier haplotype by method (S2), panel rarefaction and batch separation (S3), intronic assignment by distance to the causal site (S4), frequency of SLICK1 in N'Dama compatible with non-detection (S5), and inbreeding control for tract length (S6).

Supplementary Figs. S1–S3. Distribution of SLICK5 across populations (S1), intronic assignment by distance to the causal site (S2), and inbreeding control, tract length at the locus against random positions (S3).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supportinginformationSREN.pdf
- FigureS1.tiff
- FigureS2.tiff
- FigureS3.tiff