ACAVIS PUBLISHERS

Current Trends in Biological Science

Review Article

Volume 2, Issue 1

Genomic Tools Conceived for Temperate Livestock Breeds: A Systematic Diagnosis of Their Failure to Translate to Tropical Production Systems and a Roadmap for Structural Reform

Bhushan B*

Former Head and Principal Scientist, Division of Animal Genetics, ICAR–Indian Veterinary Research Institute (IVRI), India

ABSTRACT

Over the past two decades, genomic technologies encompassing genome-wide association studies (GWAS), genomic selection (GS), high-density single nucleotide polymorphism (SNP) arrays, and sequence imputation pipelines have fundamentally transformed livestock improvement programmes across temperate production environments. The genetic gains realised for milk yield, feed conversion efficiency, and disease resistance in Holstein-Friesian dairy cattle, Duroc swine, and Merino sheep stand as compelling demonstrations of what a well-resourced, strategically coordinated genomic programme can achieve. Yet when the same constellation of tools has been extended to tropical livestock like *Bos indicus* cattle, Murrah and Nili-Ravi buffalo, indigenous small ruminants, and native pig breeds across South Asia, Sub-Saharan Africa, and Southeast Asia, the outcomes have been consistently, and in several documented instances severely, inferior to expectations. The translation gap is wide, the consequences for millions of smallholder households are profound, and its causes have not received sufficient systematic scrutiny in the scientific literature. This comprehensive review undertakes a structured diagnosis of five principal categories of translational failure. First, fundamental biological and population-genetic incompatibilities; chiefly the substantially more rapid linkage disequilibrium (LD) decay and divergent minor allele frequency spectra characteristic of indicine populations, render the marker densities and reference architectures of temperate chip platforms inadequate for tropical breeds. Second, chronic statistical insufficiency arising from severely underpowered tropical reference populations causes genomic prediction accuracies to collapse, often below the threshold of practical utility. Third, phenotypic recording deficits and pronounced genotype-by-environment (G×E) interactions invalidate trait architectures derived from temperate training data when deployed in thermally stressed, nutritionally variable, and disease-endemic tropical contexts. Fourth, a complex of infrastructural and institutional barriers, encompassing the prohibitive cost of commercial genotyping, cold-chain logistics constraints, bioinformatics capacity deficits, and the absence of a data-governance framework for sharing national genetic resources inhibits programme-level implementation. Fifth, and most structurally consequential, SNP chip ascertainment bias rooted in Holstein-centric panel design, compounded by the inadequacy of a single Hereford-anchored taurine reference genome, systematically misrepresents genomic diversity in indicine and buffalo populations. The review examines detailed case material from Indian Murrah buffalo,

***Correspondence:**
Bharat Bhushan
Former Head and Principal Scientist, Division of Animal Genetics, ICAR–Indian Veterinary Research Institute (IVRI), Izatnagar, Bareilly 243 122, India.
Email: bhushan.drbharat@gmail.com

Article Info:
Received Date: April 29, 2026
Published Date: June 09, 2026

Citation:
Bhushan B. Genomic Tools Conceived for Temperate Livestock Breeds: A Systematic Diagnosis of Their Failure to Translate to Tropical Production Systems and a Roadmap for Structural Reform. Curr Trends Biol Sci. 2026;2(1):1-12.

Garole sheep of the Sundarbans, East African Shorthorn Zebu, and West African N'Dama cattle to illustrate how these barriers interact in practice. A forward-looking agenda centred on the construction of tropical-specific reference pangenomes incorporating long-read assemblies from diverse indicine breeds, the adoption of genotyping-by-sequencing as a cost-effective and ascertainment-unbiased alternative, the implementation of multi-environment reaction-norm genomic models, and the establishment of South-South scientific consortia with the mandate to build the infrastructure tropical livestock genomics requires is proposed and justified. The central argument of this review is that the translational deficit is not primarily a resource shortfall; it reflects a foundational paradigm mismatch a tool architecture conceived, validated, and globalised for a narrow set of temperate breeds, without adequate interrogation of its fitness for the biological, ecological, and institutional realities of tropical livestock systems. Structural reform, rather than incremental adaptation, is the appropriate response.

KEYWORDS: Genomic selection, Tropical livestock, *Bos indicus*, Linkage disequilibrium, SNP ascertainment bias, GWAS, Genotype × environment interaction, Smallholder systems, Indigenous breeds, Pangenome, Genomic prediction accuracy.

INTRODUCTION

The genomic revolution in livestock science, catalysed by the publication of the bovine genome sequence in 2004 and the commercial release of high-density SNP arrays from 2008 onwards, constitutes one of the most consequential technological inflections in the history of animal breeding. Within a remarkably compressed timeframe, genomic estimated breeding values (GEBVs) displaced pedigree-based estimated breeding values (EBVs) as the primary selection criterion in dairy cattle programmes across North America, Western Europe, and Australasia. Annual genetic gain for traits such as milk yield, longevity, and fertility accelerated by factors of two to three in leading Holstein populations.1-3 The Holstein-Friesian cow became, in effect, both the biological substrate and the commercial template through which the genomic selection paradigm was refined, validated, and subsequently propagated across the global livestock science community.

Beneath this narrative of technological triumph, however, lies a geography of benefit that is profoundly and structurally unequal. Tropical and subtropical livestock systems, which sustain the livelihoods of over 600 million smallholder farming households across South Asia, Sub-Saharan Africa, and Southeast Asia, and which collectively harbour approximately 60 to 70 percent of the world's ruminant population have been conspicuously and largely excluded from the genomic dividend.[5,6] The animals central to these systems are extraordinary in their own right. *Bos indicus* breeds such as Gir, Kankrej, Sahiwal, Ongole, and Tharparkar; riverine buffalo breeds (*Bubalus bubalis*) including Murrah, Surti, Nili-Ravi, and Jaffarabadi; indigenous small ruminants (*Capra hircus*) such as Garole, Marwari, Bannur, and Attappady Black goat; and trypanotolerant taurine cattle such as N'Dama and Orma Boran represent an irreplaceable reservoir of adaptive genetic variation, shaped over millennia by selection pressures radically different from those acting on temperate breeds, extreme heat and humidity, unpredictable and often severe nutritional constraints, endemic multi-species parasitic burdens, and sustained exposure to vector-borne pathogens. These animals encode adaptive solutions that temperate breeds neither possess nor require. Yet when investigators have sought to apply the genomic tools developed for temperate systems to these populations, the results have been, at best, modest, and at worst, systematically misleading. Genomic prediction accuracies collapse when reference populations are cross-breed or cross-subspecies. GWAS signals identified in Holstein cattle fail to replicate at equivalent chromosomal positions in Zebu or buffalo populations. SNP panels designed around *Bos taurus* haplotypes capture only a fraction of the segregating variation in *Bos indicus* and *Bubalus* genomes. Phenotypic recording systems adapted to high-input, intensively managed temperate dairy enterprises bear little structural resemblance to the dispersed smallholder systems in which tropical animals are raised and selected. These failures are not incidental or correctable by marginal adjustment, they reflect deep biological, statistical, infrastructural, and institutional fault lines that have received insufficient systematic attention in the scientific literature.

This review undertakes that systematic diagnosis. Drawing on genomic, quantitative genetic, evolutionary, and policy evidence, it constructs a comprehensive account of why the temperate genomic paradigm has not transferred effectively to tropical livestock systems, and what a genuinely tropical genomic science would need to look like. The analysis is structured around five interlocking categories of failure, followed by illustrative case studies from four distinct tropical breed-system contexts and a forward agenda proposing structural not merely incremental corrective actions. The review is intended to serve both as a scholarly synthesis and as an evidence-grounded platform for researchers, policy-makers, and national livestock programme leaders who are seeking to redirect genomic investment toward the production systems where its transformative potential remains overwhelmingly unrealised.

BACKGROUND: THE ARCHITECTURE OF TEMPERATE GENOMIC TOOL DEVELOPMENT

Historical and Intellectual Foundations

The intellectual foundation of genomic selection rests on a landmark theoretical paper by Meuwissen, Hayes, and Goddard (2001), which demonstrated that, given a sufficiently dense marker map and a training population of adequate size equipped with both genotype and phenotype data, the effects of all genome-wide SNPs could be estimated simultaneously to generate accurate predictions of total genetic value the GEBV. The architecture of this proposal carried embedded assumptions that were, at the time, neither stated as limitations nor subsequently subjected to systematic re-examination: a single-breed reference population of approximately one thousand individuals; markers in sufficient density to capture LD across the genome of the population being studied; and trait architectures that were relatively constant within the population and environment of training. These assumptions were confirmed by subsequent experimental validation almost exclusively in large, intensively phenotyped, structurally uniform commercial populations

of European origin, thereby establishing a theoretical and empirical canon whose conditions of validity were essentially coextensive with temperate breed genomics.

The Bovine HapMap Consortium (2009) provided the foundational map of LD structure and haplotype diversity across the bovine genome, drawing on 19 breeds that were predominantly *Bos taurus*, with indicine representation that was geographically restricted and numerically modest. The subsequent development of commercial genotyping arrays, the Illumina BovineSNP50 (50K) and BovineHD (770K) platforms, used SNP discovery panels derived substantially from Holstein, Angus, Limousin, and Hereford cattle. The primary bovine reference genome assembly (UMD3.1, subsequently updated to ARS-UCD1.2) was constructed from a single Hereford cow. Each of these decisions was scientifically defensible given the then-available resources, yet collectively they embedded a systematic *Bos taurus* ascertainment bias into the foundational infrastructure of bovine genomics before tropical production systems were meaningfully consulted.

Scale and Statistical Power of Temperate Reference Populations

The effectiveness of genomic selection scales with reference population size in a mathematically well-defined manner. Daetwyler et al.[6] demonstrated analytically that prediction accuracy is approximately proportional to $h^2N / (N + Me)$, where h^2 is trait heritability, N is the effective reference population size, and Me is the number of independently segregating chromosome segments. For Holstein cattle, reference populations assembled through international collaboration, EuroGenomics, the North American CDN reference set, and the USDA national reference population, now encompass tens of thousands of genotyped bulls with multi-daughter phenotypic records spanning multiple countries and environments, providing statistical power commensurate with the precision that genomic selection achieves in practice. No remotely comparable infrastructure exists for any tropical breed.

The Phenotyping Infrastructure Advantage

Genomic tools are, at their core, sophisticated phenotype-prediction machines whose precision is ultimately bounded by the quality, volume, and longitudinal consistency of the phenotypic records used in their training. The milk-recording infrastructure of the Netherlands, Germany, Canada, and the United States, maintained over decades with harmonised protocols, electronically integrated identification systems, and mandatory participation by commercial operators, has provided the essential substrate from which temperate genomic tools derive their predictive power. Tropical livestock systems are characterised, in contrast, by phenotypic recording that is frequently heterogeneous, episodic, institutionally fragmented, and not designed to meet the data quality standards required for genomic reference population construction. Individual animal performance data in many tropical countries are irregularly collected, incompatibly coded across institutions, and rarely integrated into unified national databases available for genomic analysis.[5,7] This phenotypic infrastructure asymmetry is as consequential as any biological difference between temperate and tropical breeds.

BARRIER I: BIOLOGICAL AND POPULATION-GENETIC INCOMPATIBILITIES

Linkage Disequilibrium Structure, Extent, and Decay Kinetics

Linkage disequilibrium, the non-random statistical association between alleles at different genomic loci, is the essential bridge between marker genotypes and causal variants that makes genomic prediction possible. The marker density required to adequately tag LD across the genome is determined by the rate at which LD decays with physical distance, which is itself a deterministic function of a population's effective size (Ne) and demographic history. Empirical studies have documented substantial and reproducible differences in LD architecture between *Bos taurus* and *Bos indicus* populations. In Holstein-Friesian cattle, mean r^2 values typically remain above 0.20 at inter-marker distances approaching 100 kilobases (kb) and above 0.15 at 200 kb, reflecting the demographic bottlenecks and intense artificial selection associated with commercial breed formation.[8,9] In *Bos indicus* breeds, LD decays substantially more rapidly: r^2 values frequently fall below 0.20 within 50 kb and below 0.10 within 100 kb, with considerable variation among breeds reflecting their distinct demographic histories.[10,11] In Murrah buffalo (*Bubalus bubalis*), preliminary estimates suggest LD decay is yet more pronounced, with r^2 approaching baseline at distances below 60 kb in some genome-wide analyses.[12] This accelerated decay reflects the larger effective population sizes historically characteristic of South Asian and African livestock populations, which have been maintained under substantially less intense artificial selection and have consequently preserved greater genomic diversity.

The practical implication is stark and quantifiable. A SNP density that adequately captures the LD architecture of the Holstein genome is systematically inadequate to tag the majority of quantitative trait loci (QTL) in Sahiwal, Gir, or Murrah buffalo populations. The commercially dominant 50K chip provides on average approximately one marker per 50 kb across the bovine genome, a density broadly sufficient for *Bos taurus* application but one that leaves large genomic intervals in indicine populations without meaningful marker coverage of segregating causal variation. Theoretical and empirical estimates suggest that adequate LD coverage of *Bos indicus* genomes would require marker densities in the range of 150,000 to 300,000 SNPs, approaching the information content of shallow whole-genome sequencing for some breeds.[3,10]

Allele Frequency Divergence and Population-Specific Variation

Beyond LD architecture, the allele frequency spectra at causative loci differ substantially between temperate and tropical breeds, reflecting the divergent histories of selection, genetic drift, and local adaptation that have acted on these lineages for millennia. Alleles that are segregating and causally important for a quantitative trait in Holstein populations may be nearly fixed in either the favourable or unfavourable direction in Zebu or indigenous small ruminant populations, rendering corresponding SNPs uninformative for within-population genomic analysis.[11,13] This problem is not merely statistical; it reflects the genuine biological reality that the trait architectures of tropical breeds are partially distinct from those of their temperate counterparts, having been shaped by different selective environments over many thousands of generations.

Positive selection for local adaptation encompassing thermotolerance mediated by the SLICK prolactin receptor variant, tick and gastrointestinal nematode resistance governed by partially characterised immune loci in the BoLA region and elsewhere, trypanotolerance in West and East African taurine breeds, and water use efficiency in dryland Zebu populations has driven selective sweeps in tropical breeds that are absent from temperate genomes. These sweeps reduce genetic diversity in chromosomal flanking regions, restructure haplotype blocks, and shift allele frequencies at linked markers in ways that confound the application of temperate reference panels.[14,15] The genomic regions harbouring these adaptive variants represent precisely the most economically relevant targets for tropical livestock genomic research, yet they are the regions where temperate-derived tools are least effective.

Subspecific Genetic Distance and Divergent Ancestral Linkage Phase

Bos indicus and *Bos taurus* diverged from a common domesticated ancestor approximately 200,000 to 250,000 years before present, with independent domestication events occurring in the Indian subcontinent and the Near East, a divergence time roughly equivalent to that separating some human continental populations.[16,17] This deep divergence has produced systematic differences in genome-wide LD phase, the specific combination of alleles on shared haplotype blocks that substantially impair cross-subspecies genomic prediction. The accuracy of genomic prediction across breed boundaries depends critically on the consistency of LD phase between markers and QTL across the populations involved; when phase consistency is low, as it necessarily is between *Bos taurus* and *Bos indicus* given their evolutionary distance, marker effects estimated in a reference population provide minimal predictive information for animals in the target population.[18,19]

Admixed animals, which are numerically dominant in many tropical production systems where indicine cows are crossed with taurine sires for dairy improvement, present a further layer of genomic complexity. F_1 and backcross individuals carry mosaic genomes in which taurine and indicine chromosome segments alternate at frequencies that vary among individuals and across chromosomal regions. No single-breed reference population whether taurine or indicine provides an adequate statistical framework for predicting GEBVs in such admixed animals, and yet admixed genotypes constitute a large fraction of the commercially important dairy cattle populations of Brazil, India, East Africa, and Southeast Asia.[20,21]

BARRIER II: STATISTICAL INADEQUACY AND REFERENCE POPULATION FAILURE

Reference Population Size, Composition, and the Minimum Threshold for Utility

The quantitative relationship between reference population size and genomic prediction accuracy is theoretically well established and empirically corroborated. For traits of moderate heritability ($h^2 \approx 0.30$), minimum reference populations of 2,000 to 4,000 genotyped and phenotyped individuals are generally considered necessary to achieve prediction accuracies of practical utility in single-breed scenarios, with substantially larger reference sets required when heritability is low, the number of QTL is large, or across-breed prediction is attempted.[9,22] In practice, reference populations for major temperate breeds have expanded far beyond these minima through decades of sustained international collaboration.

For tropical breeds, the situation is dramatically and disconcertingly different. A systematic survey of the genomic selection and GWAS literature for *Bos indicus* cattle conducted for this review found that the majority of published studies employed reference populations of fewer than 500 animals, with a significant proportion falling below 200 individuals, well beneath the threshold at which marker effect estimates can be expected to be either precise or stable across validation sets. For Murrah buffalo, the largest published GWAS reference to date involves approximately 292 animals.[12] For indigenous small ruminant breeds across South Asia and Sub-Saharan Africa, sample sizes in genomic studies rarely exceed 150 individuals. Under these conditions, marker effect estimates are inevitably dominated by sampling variance rather than signal, and the resulting GEBVs cannot be expected to provide predictions of reliably positive economic value.

Non-Replication of GWAS Signals and Population Stratification Artefacts

A consistent and troubling pattern in the tropical livestock genomic literature is the failure of QTL associations originally detected in *Bos taurus* breeds to replicate at equivalent chromosomal positions when the same or related traits are studied in *Bos indicus* or indigenous small ruminant populations. While a proportion of this non-replication reflects genuine biological differences in trait genetic architecture between subspecies differences that are real, ecologically grounded, and scientifically interesting much of it is attributable to statistical artefacts. Inadequate reference population size generates false-positive signals through inflated test statistics; insufficient correction for population stratification produces spurious associations at the subspecific rather than the locus level; and the use of *Bos taurus*-derived imputation reference panels reduces genotyping accuracy in indicine animals in ways that bias association statistics.[13,20]

Population stratification is particularly acute in tropical livestock GWAS, where management system, agro-ecological zone, breed composition, and production system are frequently confounded in ways that are difficult to disentangle using standard principal-component correction. Genomic studies of East African cattle must navigate an admixture landscape of exceptional complexity, encompassing ancient indigenous taurine breeds (N'Dama, Lagune), *Bos indicus* populations (Boran, East African Zebu), and relatively recent taurine introductions introduced during colonial and post-colonial development programmes. In such populations, inadequately controlled stratification can generate phantom associations reflecting continental ancestry rather than locus-level causal variation, rendering published GWAS findings ambiguous in their biological interpretation.[11]

The Collapse of Cross-Breed Genomic Prediction Accuracy

Attempts to use Holstein or other temperate breed reference populations to predict GEBVs in *Bos indicus* animals have produced accuracy estimates that are, in the most optimistic assessments,

modest and, in realistic appraisals, insufficient to justify programmatic implementation. Multiple independent evaluations have placed cross-subspecies genomic prediction accuracies for traits such as milk yield, fat percentage, and reproductive efficiency in the range of 0.10 to 0.35, compared to within-breed accuracies of 0.60 to 0.80 achievable for Holstein in well-resourced national programmes.[3,18,19] For adaptive traits like heat stress response, parasite resistance, reproductive efficiency under nutritional constraint, for which temperate reference populations are by definition phenotypically uninformative, cross-breed prediction accuracy approaches zero.

The consequences for tropical breeding programmes have been substantive and underappreciated. National dairy improvement initiatives in India, Brazil, East Africa, and parts of Southeast Asia that have attempted to implement genomic selection using Holstein-derived reference populations have reported systematic overestimation of genetic merit in indicine animals, producing selection decisions that failed to deliver the expected phenotypic gains at the farm level and in some cases selecting against locally adaptive traits of genuine economic importance.[23,24] This is not merely an academic concern; it represents a misdirection of public investment in genetic improvement with concrete consequences for food security and farmer livelihoods.

BARRIER III: PHENOTYPIC GAPS AND GENOTYPE × ENVIRONMENT INTERACTIONS

The Uncharacterised Trait Architecture of Tropical Production Systems

Many of the traits that define economic and adaptive value in tropical livestock systems are absent from the training datasets that underpin temperate genomic tools, because they are either not measurable in temperate environments, not recorded by temperate performance schemes, or simply not recognised as breeding objectives by institutions designing genomic programmes for commercial dairy or meat production. Heat tolerance, defined operationally as the maintenance of productive performance across a gradient of increasing temperature-humidity index is a quantitative trait with substantial and well-documented genetic variance in indicine and synthetic tropical cattle populations, yet virtually no temperate reference population dataset incorporates direct or even indirect measures of this trait. Resistance to *Theileria parva*, *Babesia bovis*, *Trypanosoma brucei* and *Trypanosoma congolense*, and *Haemonchus contortus* parasites collectively responsible for enormous productive losses across tropical livestock systems requires specialised pathogen-challenge phenotyping protocols that are logistically demanding, costly, and currently absent from routine performance recording in virtually all tropical countries.

Water use efficiency under dryland grazing conditions; feed conversion from fibrous roughage-based diets; milk let-down reflex and total milk yield under hand-milking systems without calf-at-foot stimulus; maternal behaviour and calf survival under semi-intensive or extensive management; traction performance and working endurance in draught cattle; and survival and reproductive resilience under seasonal nutritional stress are additional traits of substantial economic significance in tropical smallholder systems that are entirely absent from the temperate genomic reference architecture. A selection tool optimised for Holstein traits measured under total mixed ration feeding and machine milking provides no guidance whatsoever for the selection decisions facing a Zebu-cross cattle owner managing animals on communal rangeland.

The Magnitude and Breeding Consequences of G×E Interactions

Genotype-by-environment interaction, the phenomenon whereby the relative genetic merit of individuals or genotypes changes as a function of the environment in which they are expressed, is a fundamental biological reality with particularly severe consequences when genomic tools developed in one environment are applied to a markedly different one. For livestock, the principal environmental axes that generate G×E are thermal load, nutritional plane, infectious disease pressure, and management intensity. The transition from a temperate high-input dairy environment to a tropical smallholder grazing system represents simultaneous and substantial displacement along all four axes at once.

Reaction norm approaches to modelling G×E in livestock have demonstrated that the genetic correlation for the same trait measured across widely differing environments can fall substantially below 0.50 in certain breed-environment combinations, indicating that the genetic determinism of performance is genuinely and materially environment-specific.[25,26] When across-environment genetic correlations are substantially below unity, breeding values estimated in one environment provide not merely imprecise but potentially biased and directionally misleading predictions for performance in a second environment. This has direct and serious implications for the practice of using Holstein GEBVs as benchmark reference points in tropical dairy improvement initiatives.

Phenotypic Recording Capacity in Smallholder Production Systems

The phenotypic infrastructure available to support genomic tool development in tropical smallholder livestock systems is independently formidable in its inadequacy. The dispersed, small-herd or small-flock structure of the predominant production system; the absence of systematic individual animal identification that can be reliably traced through multiple production cycles; the low and geographically heterogeneous density of extension and veterinary services capable of conducting standardised performance measurements; the irregular and seasonally interrupted nature of milk production; and the prevalence of subsistence-oriented management in which economic record-keeping is neither customary nor perceived as practically justified, all of these factors combine to produce a phenotypic data environment that is inadequate for genomic reference population construction by any credible standard.

National livestock recording schemes in South Asia and Sub-Saharan Africa, where functional examples exist, typically encompass only a small, geographically concentrated, and institutionally selected fraction of the breeding population. Records are frequently limited to pedigree registration and reproductive events rather than economically critical production traits. Data quality is inconsistent in ways that are difficult to correct post hoc and that limit utility for genomic analyses requiring precise, repeatedly measured phenotypes

on large numbers of genotyped animals. The contrast with the comprehensive, longitudinal, harmonised milk recording data available through INTERBULL for Holstein populations in fifteen or more countries is stark and, in the short to medium term, not readily bridgeable by goodwill alone.

BARRIER IV: INFRASTRUCTURAL AND INSTITUTIONAL CONSTRAINTS

Genotyping Cost as a Binding Structural Constraint

The market cost of high-density SNP genotyping has fallen considerably since the introduction of the first commercial bovine arrays but remains a binding constraint to genomic programme development in low- and middle-income country contexts. At prices prevailing over the period 2018 to 2024, the Illumina BovineHD 770K chip cost approximately USD 80 to 120 per animal, while the widely used 50K platform cost USD 35 to 60. For a reference population of even the most modest size consistent with statistical utility 2,000 animals genotyping costs alone amount to USD 70,000 to 240,000, exclusive of laboratory infrastructure, bioinformatics analysis, personnel, and phenotypic recording costs. In production systems where a smallholder household's annual gross margin from a dairy cow may be USD 100 to 300, these investment requirements are prohibitive in the absence of substantial external subsidy or sustained international partnership.[12,20]

Lower-density chip options in the 3K to 10K SNP range reduce per-animal genotyping costs substantially but require subsequent imputation to higher density for analyses requiring adequate genome coverage. Imputation accuracy for tropical breeds is itself compromised by the absence of breed-matched, breed-specific high-density reference panels creating a circularity in which the cost-reduction strategy depends on infrastructure investments that have not yet been made for the vast majority of tropical breeds. This structural circularity has no resolution within the existing temperate-derived tool paradigm.

Bioinformatics Capacity and Computational Infrastructure

The analytical pipelines required for rigorous GWAS, genomic selection, population genomic characterisation, and comparative genomic analysis demand computational resources, software proficiency, and specialised domain knowledge that are unevenly and inequitably distributed globally. The bioinformatics capacity required to manage and quality-control whole-genome sequence datasets, implement population-stratification-corrected GWAS at genome-wide significance thresholds, deploy GBLUP or Bayesian genomic prediction models, conduct cross-validation and independent validation, and communicate results with appropriate statistical transparency, this capacity is concentrated in a relatively small number of research universities and government research institutes, predominantly in the northern hemisphere. In South Asia, Sub-Saharan Africa, and Southeast Asia, the institutions responsible for national livestock genetic improvement, animal breeding divisions of agricultural universities, state livestock development boards, and national agricultural research systems have typically limited access to high-performance computing facilities, are structurally understaffed with trained bioinformaticians, and depend on collaborative relationships with overseas institutions that are fragile, grant-contingent, and not designed to build durable in-country capacity.

Sample Logistics, Data Governance, and Institutional Coordination

Biological sample collection for genomic analysis under field conditions in tropical environments presents logistical challenges that are qualitatively and quantitatively greater than those encountered in the controlled settings for which standard genomic laboratory protocols were developed. Blood, semen, ear tissue, and hair follicle samples require cold-chain management throughout transport from collection sites to genotyping laboratories that may be located in a different city, country, or continent. Sample integrity degrades more rapidly under tropical ambient temperatures and humidity. Import and export regulations governing the international transfer of biological materials from livestock motivated by legitimate biosecurity concerns but implemented through bureaucratic processes that are often slow, unpredictable, and inconsistent across jurisdictions add delays and compliance costs that impede timely programme implementation.

Data sharing norms in tropical livestock genomic research are further constrained by concerns about national genetic resource sovereignty, intellectual property over genomic discoveries in locally adapted breeds, and competitive dynamics among research groups operating in chronically resource-limited environments. The collaborative trust and data-standardisation infrastructure that enabled the construction of large, multi-country Holstein reference populations through INTERBULL over four decades was built incrementally through sustained institutional investment and peer-driven norm-setting. No comparable collaborative architecture yet exists for tropical breeds. The African Livestock Genomics Consortium, the ICAR working group on buffalo genomics, and analogous regional initiatives represent important early steps, but remain far short of the scale, data harmonisation, and governance sophistication required to address the reference population deficit.

BARRIER V: SNP ASCERTAINMENT BIAS AND REFERENCE GENOME DEFICIENCY

The Mechanism and Consequences of SNP Ascertainment Bias

Ascertainment bias in genomic studies arises when the SNPs selected for inclusion on a genotyping array are identified from a non-representative discovery panel, producing systematic distortions in allele frequency estimation, LD pattern characterisation, and population differentiation statistics in populations excluded from that discovery panel. The Illumina BovineSNP50 array, the workhorse of bovine genomic selection and GWAS globally was developed using variants identified through sequencing of a small number of *Bos taurus* individuals, primarily from Holstein, Angus, and Limousin populations. The consequences for *Bos indicus* genomic research are multiple, documented, and operationally significant.

Most directly, SNPs that are robustly polymorphic in *Bos taurus* populations but nearly fixed in *Bos indicus* populations provide zero within-breed information for genomic analysis, effectively reducing the usable marker density below the chip's nominal specification. Gautier et al.[10] documented that approximately 20 to 35 percent of

50K chip markers are monomorphic in various *Bos indicus* breeds, compared to fewer than five percent in Holstein, a differential that translates directly into reduced genome coverage and reduced statistical power for any analysis requiring adequate marker representation. Furthermore, the SNPs that are polymorphic in both subspecies tend, by the mathematical properties of ascertainment, to be those segregating at intermediate frequencies in both not necessarily the variants most informative for detecting QTL specific to indicine trait architectures. Finally, the haplotype block structure defined by 50K chip SNPs in *Bos taurus* populations does not correspond to the haplotype block boundaries in *Bos indicus*, impairing the effectiveness of haplotype-based imputation and comparative analyses.

Reference Genome Limitations and Structural Variant Misrepresentation

The reference genome is the coordinate framework upon which all comparative and population genomic analyses depend. The current primary bovine reference genome, ARS-UCD1.2, was constructed from a single Hereford cow and consequently represents a single haplotype from a single individual of a single *Bos taurus* breed. While the assembly achieves high continuity and annotation quality for the applications for which it was designed, it exhibits documented gaps and systematic misassemblies in chromosomal regions that are highly divergent between *Bos taurus* and *Bos indicus* lineages. Structural variants including insertions, deletions, inversions, and copy number variations that are characteristic of the indicine genome may be altogether absent from, or seriously misrepresented in, the *Bos taurus* reference, producing read-mapping errors, variant miscalling, and systematic information loss when whole-genome sequence data from indicine animals are aligned to the taurine coordinate system.

For buffalo (*Bubalus bubalis*), the reference genome situation is considerably more acute. Although assemblies for the Murrah buffalo genome have been published[12,27], these remain substantially less complete, less annotated, and less functionally characterised than the bovine reference. SNP arrays specifically designed with buffalo-specific ascertainment panels remain at an early stage of development. The functional annotation of the buffalo genome, the systematic identification of gene regulatory elements, cis-regulatory modules, alternative transcripts, tissue-specific chromatin states, and non-coding functional sequences lags considerably behind the bovine genome annotation, limiting the biological interpretability of any genomic discoveries made in this species. Millions of Murrah buffalo farmers make selection decisions daily in a genomic darkness that is maintained, in significant part, by this infrastructural deficit.

The Pangenome Paradigm as a Structural Solution to Reference Genome Limitations

The pangenome concept, a reference structure that captures the full genomic diversity of a species or subspecies by representing multiple individuals as a graph or ensemble of sequences, rather than the single linear sequence of a single representative has emerged as a principled and technically feasible solution to the fundamental limitations of single-reference-genome genomics. A livestock pangenome incorporating high-quality long-read assemblies from multiple individuals spanning major *Bos indicus* breeds across Asia and Africa, alongside existing *Bos taurus* assemblies, would provide a coordinate system within which the structural variation characteristic of tropical breeds could be accurately mapped, annotated, and functionally interpreted. Pilot pangenome projects for cattle have demonstrated both the technical feasibility and the substantial genomic insights accessible through this approach.[18,29] However, comprehensive and proportionate tropical breed representation remains a future aspiration rather than a current programmatic reality, one whose realisation requires coordinated international investment rather than the spontaneous initiatives of individual research groups.

ILLUSTRATIVE CASE STUDIES FROM TROPICAL PRODUCTION SYSTEMS

Murrah Buffalo in India: Extraordinary Economic Significance, Neglected Genomics

The Murrah buffalo (*Bubalus bubalis*) contributes approximately 58 percent of India's total milk production, positioning India as the world's largest milk producer and making the Murrah breed the single most economically important dairy animal in any tropical country.[30] Originating in the districts of Rohtak and Hisar in Haryana, the Murrah is capable of producing 2,200 to 3,500 litres of milk per lactation at a fat content of 6.5 to 8.5 percent; figures that render it broadly competitive with Holstein in fat-corrected milk yield under optimal management conditions, while maintaining substantially superior adaptation to heat, humidity, and fibrous roughage-based feeding. Despite this extraordinary economic centrality, the genomic characterisation of the Murrah breed stands at a developmental stage that bovine genomics had essentially surpassed by 2010.

The largest published GWAS for milk production traits in Murrah buffalo to date involved 292 animals genotyped with a custom 90K buffalo SNP chip[12]; a reference population size substantially below the minimum thresholds established by quantitative genetics theory for reliable marker effect estimation. Genomic heritability estimates for milk yield, fat percentage, and protein concentration are broadly consistent with those published for dairy cattle, establishing that a genomic improvement programme is biologically feasible and that adequate genetic variance exists to sustain meaningful selection response. The statistical infrastructure to implement such a programme reliably, however, does not yet exist. The buffalo SNP chip itself, though an important advance over the practice of applying bovine arrays to buffalo DNA, was developed from a limited ascertainment panel and achieves substantially less genomic coverage of the buffalo LD landscape than a purpose-designed, population-calibrated high-density array would provide. The consequence is that hundreds of millions of buffalo farmers across South Asia continue to select replacement animals based on phenotypic appearance and anecdotal maternal performance records, forgoing genetic gains that a well-resourced genomic programme could, conservative estimates suggest, triple or quadruple within a decade.

Garole Sheep: Globally Unique Adaptive Traits and the Invisibility of Causative Loci

The Garole sheep of the Sundarbans delta region of West Bengal and Bangladesh is, by any measure, a genetically remarkable breed. Small-bodied ewes typically weigh 12 to 18 kg at maturity and

adapted to the humid, warm, and seasonally flooded mangrove-fringe environment of the Bengal delta, Garole ewes routinely produce litter sizes of 1.8 to 2.5 lambs per lambing. This hyper prolificacy, which exceeds that of many purpose-selected prolific commercial breeds, is combined with documented resistance to gastrointestinal nematode challenge, particularly *Haemonchus contortus*, that is severe in the humid subtropical environment. Both traits i.e., hyper prolificacy and endoparasite resistance are of substantial international interest, having attracted attention from breeding programmes in Australia, New Zealand, and Western Europe seeking to introgress these characteristics into commercial flocks.

The genetic basis of Garole prolificacy has been partially resolved at the molecular level: a point mutation in the bone morphogenetic protein receptor 1B gene (BMPR-1B, the FecB or Booroola mutation) is segregating at high frequency in the Garole population and makes a substantial contribution to the elevated ovulation rate underlying hyper prolificacy. However, GWAS conducted using commercial ovine SNP arrays (OvineSNP50, Illumina) have demonstrated substantially lower power to detect additional contributing loci in Garole than equivalent analyses in Merino-derived breeds, reflecting both the breed-specific LD structure and the allele frequency distortions introduced by ascertainment bias. The nematode resistance phenotype, which appears to be polygenic with a genomic architecture distinct from that characterised in temperate sheep populations, has proved even more refractory to genomic investigation with available tools.[31] A breed with globally unique adaptive genetics of confirmed international value consequently remains substantially uncharacterised at the genomic level, a consequence not of scientific disinterest but of tool inadequacy.

East African Shorthorn Zebu: Navigating Admixture in a Complex Genomic Landscape

The East African Shorthorn Zebu (EASZ) cattle distributed across Kenya, Tanzania, Uganda, Rwanda, and adjacent countries represent one of the most genomically complex admixture landscapes in domestic livestock. Phylogenomic studies have documented ancient taurine introgression events, likely from West African taurine populations (N'Dama, Ankole ancestors), overlaid on a zebu genomic background that itself encompasses multiple ancestral lineages (Boran, Rendille, Samburu, and other regional variants), with more recent admixture from European taurine breeds (Ayrshire, Friesian, Guernsey, Brown Swiss) introduced through colonial and post-colonial livestock development programmes. This layered admixture history produces population structure that is highly variable among geographic and ethnic communities, temporally dynamic as crossbreeding continues, and formally confounded with the agro-ecological, management, and cultural variables that influence phenotypic outcomes. Disentangling breed-level from locus-level associations in such populations demands sophisticated and computationally intensive genomic modelling approaches that are rarely available within the institutions responsible for national livestock improvement.

Trypanotolerance, the capacity of certain East African cattle to survive, reproduce, and maintain acceptable productive performance in tsetse-endemic areas where *Trypanosoma brucei*, *Trypanosoma brucei* rhodesiense, and *Trypanosoma congolense* are enzootic among the most economically and epidemiologically important adaptive traits in the region, determining the viability of livestock-based agriculture across millions of square kilometres of productive land that would otherwise be unsuitable for cattle. GWAS for trypanotolerance in EASZ populations and neighbouring N'Dama taurine breeds have identified candidate genomic regions on multiple chromosomes, but signals have proved inconsistent across studies, a pattern attributable to the compound effects of phenotypic complexity (tolerance involves multiple systems including haematopoietic response, innate immune function, and metabolic regulation), population stratification confounding, and the statistical underpowering that affects all tropical breed genomic studies.[11,32] The genomic basis of this trait of critical importance for food security across a substantial portion of the African continent thus remains inadequately characterised.

A PATH FORWARD: STRUCTURAL REFORM OF GENOMIC SCIENCE FOR TROPICAL LIVESTOCK SYSTEMS

Construction of Tropical-Specific Reference Pangenomes

The most structurally consequential investment required to address the translation failure documented in this review is the deliberate construction of reference pangenomes that authentically and proportionately represent the genomic diversity of tropical livestock species and breeds. For cattle, this would necessitate the generation of high-quality long-read genome assemblies using PacBio HiFi and Oxford Nanopore Technologies from a minimum of 30 to 50 individuals spanning the major *Bos indicus* breeds of South Asia and Africa (Sahiwal, Gir, Kankrej, Tharparkar, Nelore, Boran, Ankole, and East African Zebu populations), alongside geographically diverse EASZ genotypes and economically important admixed breeds, integrated with existing *Bos taurus* assemblies in a graph-pangenome framework. Equivalent efforts for buffalo would encompass Murrah, Nili-Ravi, Surti, Jaffarabadi, and Swamp buffalo, representing both the riverine and swamp subspecies across their geographic ranges. Comparable programmes for indigenous small ruminants of South Asia and Sub-Saharan Africa would require analogous coordinated efforts at the small ruminant level.

Long-read sequencing now achieves the read lengths and per-base accuracy required to resolve complex repetitive regions, large structural variants, and copy number polymorphisms that are systematically underrepresented in short-read assemblies and that are likely to harbour substantial biologically and economically important variation in tropical breeds. The cost trajectory for generating chromosome-level genome assemblies has declined dramatically, with high-quality reference-quality assemblies now achievable at costs below USD 5,000 per individual. The primary constraint to pangenome construction is no longer technological; it is organisational, the assembly of international collaboration, biological material repositories, bioinformatics analytical capacity, and data governance frameworks capable of delivering a multi-breed tropical pangenome within a timeframe relevant to current breeding decisions.[28,29]

Genotyping-by-Sequencing as a Cost-Effective and Ascertainment-Unbiased Alternative

Genotyping-by-sequencing (GBS) and closely related reduced-representation sequencing approaches including restriction-site associated DNA sequencing (RADseq) and its variants offer a potentially transformative alternative to commercial SNP arrays for tropical breed populations in which no breed-matched array exists or in which array genotyping costs exceed available programme budgets. GBS uses restriction enzyme digestion to reduce genomic complexity systematically, followed by sequencing of the resulting representative fragments, to generate genotype data at thousands to hundreds of thousands of SNPs per individual at per-sample costs that can be 25 to 40 percent of equivalent commercial array costs. Critically, and in sharp contrast to fixed-content arrays, GBS-derived SNP discovery is unbiased with respect to breed of origin: the variants identified reflect the actual segregating diversity of the population under study rather than the diversity ascertained in a different breed for a different purpose.

Imputation from GBS-level genotyping to whole-genome sequence coverage, using breed-matched imputation reference panels, can further increase information density at modest marginal cost and in ways that are compatible with subsequent genomic prediction analyses. The development of such breed-matched imputation reference panels is a natural and synergistic complement to the pangenome programme described above. Research groups in India, Brazil, and Kenya have begun exploring GBS-based genomic selection frameworks for indigenous breeds, with preliminary results that are encouraging regarding both genotyping cost-effectiveness and within-breed prediction accuracy.[33,34] Scaling these initiatives from proof-of-concept demonstrations to programme-level implementation requires standardisation of laboratory protocols across institutions, development of quality-controlled tropical breed-specific imputation panels, and systematic cross-validation of genomic prediction accuracy across multiple trait-breed-environment combinations before widespread adoption.

Multi-Environment Reaction Norm Genomic Models and Machine Learning Approaches

Addressing the G×E challenge in tropical livestock genomics requires statistical modelling frameworks capable of estimating breeding values that are explicitly environment-specific rather than environment-averaged frameworks that acknowledge the biological reality that the genetic determinism of trait expression changes meaningfully across the environmental gradients that define tropical production system diversity. Reaction norm genomic models, which parameterise the phenotypic response of a genotype across an environmental gradient using a linear or polynomial function of a measured environmental covariate, provide a theoretically principled approach to this problem. In the tropical livestock context, candidate environmental covariates include temperature-humidity index integrated over the production period, seasonal pasture quality indices, disease incidence rates, and composite management intensity scores that distinguish subsistence from market-oriented production. Genomic reaction norm models have been implemented for heat stress tolerance in dairy cattle under temperate conditions[35,36] but have rarely been deployed in tropical breed-environment contexts where their utility would be greatest and where their environmental covariates are most meaningful.

Machine learning approaches including deep learning architectures, gradient boosting algorithms, and random forest ensemble methods have been explored as complements or alternatives to conventional linear genomic prediction in contexts where marker-trait relationships are non-linear or involve complex epistatic interactions that linear models cannot capture. Published evidence suggests that these approaches can outperform GBLUP for specific trait-population combinations, particularly when interaction effects are biologically important. Their general advantage over linear models, however, is inconsistent across traits and populations, and they typically require substantially larger training datasets than are currently available for any tropical breed. Their effective deployment at scale in tropical systems consequently awaits the reference population development in terms of both sample size and phenotypic depth that remains the primary statistical bottleneck.

South–South Scientific Consortia and Structural Policy Redirection

The biological, statistical, infrastructural, and institutional barriers documented in this review are interconnected and mutually reinforcing in ways that make them resistant to resolution through the efforts of individual research groups or national programmes acting in isolation, however well motivated. What is required is a coordinated, sustained, and strategically governed international investment in tropical livestock genomics analogous in its ambition to the collaborative infrastructure that has enabled temperate breed genomic progress, but designed explicitly and from the outset for the biological and institutional realities of tropical production systems. The BovINDIA initiative in India, embryonic elements of a Pan-African Livestock Genomics network, the CGIAR Livestock Programme, and the SAPLING project represent early examples of the collaborative architecture needed. Strengthening these consortia, extending them to incorporate systematic large-scale genotyping campaigns, standardised and validated phenotyping protocols for tropical-specific traits, shared high-performance computing infrastructure, and South Asian and African bioinformatics training programmes constitutes a high-priority actionable recommendation of this review.

International funding agencies including the Bill and Melinda Gates Foundation, the International Fund for Agricultural Development, the Asian Development Bank, and bilateral development cooperation programmes have historically channelled livestock genomics investments predominantly toward technology transfer from established temperate reference centres to developing-country partners: transferring tools, not building capacity to create new ones. The evidence assembled in this review argues compellingly that this model has fundamental structural limitations for tropical systems: it perpetuates dependence on tools that are, by design and by validation history, not fit for purpose. Redirecting a meaningful proportion of international investment toward the construction of genuinely tropical genomic infrastructure reference pangenomes, GBS pipelines, tropical trait phenotyping systems, and bioinformatics

capacity embedded within tropical institutions would represent a structural reorientation with proportionately and demonstrably greater long-term impact than continued incremental transfer of temperate tools.[38-40]

SUMMARY TABLES

Table 1: Summary of key barriers and proposed solutions for translating genomic tools to tropical livestock systems

Barrier Category	Specific Challenge	Effect on Genomic Tool Performance	Proposed Solution
I. Biological & Population Genetic	Rapid LD decay in *Bos indicus* (< 50 kb)	50K chip marker density inadequate; QTL tagging fails	Higher-density tropical-specific chips; GBS
	Allele frequency divergence at causative loci	GWAS signals shift or disappear across subspecies	Within-breed discovery panels; pangenome reference
II. Statistical & Reference Population	Reference population n < 500	Genomic prediction accuracy < 0.30	South–South consortium-based genotyping campaigns
	Cross-breed LD phase inconsistency	Cross-breed GEBV accuracy 0.10–0.35	Breed-matched reference panels; multi-breed models
III. Phenotypic & G×E	Tropical adaptive traits unmeasured	No training data for heat tolerance or parasite resistance	Standardised tropical trait phenotyping protocols
	G×E across environments	GEBVs estimated in one environment biased in another	Reaction-norm genomic models; multi-environment trials
IV. Infrastructural & Institutional	Genotyping cost prohibitive	Reference populations remain underpowered	GBS; subsidised genotyping; pooled procurement
	Bioinformatics capacity deficit	Analyses outsourced; local capacity not built	South–South capacity building; open-source training
V. Ascertainment Bias & Reference Genome	SNP chip designed on *Bos taurus*	20–35% SNPs monomorphic in indicine breeds	Indicine-specific chip development; pangenome
	Single Hereford reference genome	Indicine structural variants misrepresented	Multi-breed tropical pangenome programme

Table 2: Linkage disequilibrium decay characteristics in representative temperate and tropical livestock breeds

Breed / Population	Subspecies	$r^2 > 0.20$ (distance)	$r^2 > 0.10$ (distance)	Key Reference
Holstein-Friesian	*Bos taurus*	~100 kb	~200–300 kb	8
Angus	*Bos taurus*	~80 kb	~180 kb	9
Limousin	*Bos taurus*	~70 kb	~150 kb	10
Sahiwal	*Bos indicus*	~25–40 kb	~60–80 kb	13
Gir	*Bos indicus*	~20–35 kb	~50–70 kb	10
Nelore	*Bos indicus*	~30–50 kb	~70–100 kb	37
East African Zebu	*Bos indicus*	~20–40 kb	~50–80 kb	11
Murrah Buffalo	*Bubalus bubalis (riverine)*	~15–30 kb	~40–60 kb	12
Garole Sheep	*Ovis aries*	~30–50 kb	~80–120 kb	31

CONCLUSION

The failure of genomic tools developed in temperate breeds to translate effectively to tropical livestock systems is neither a peripheral limitation nor a temporary inconvenience that will resolve itself as genomic technologies mature. It is a foundational structural problem that reflects the conditions under which the technology was conceived, validated, and globalised conditions defined by the biology, demography, production economics, and institutional infrastructure of a small number of intensively managed breeds in a small number of high-income countries. Holstein-centric SNP arrays, a Hereford-anchored taurine reference genome, Holstein-derived training populations, and Holstein-calibrated trait architectures collectively constitute a genomic infrastructure that is genuinely and demonstrably fit for purpose within the narrow set of breeds and environments for which it was designed and systematically inadequate for the far larger set of breeds, environments, farmers, and food systems that it was never designed to serve.

The five categories of barrier documented in this review biological and population-genetic incompatibility, statistical inadequacy of reference populations, phenotypic gaps and G×E interaction, infrastructural and institutional constraints, and SNP ascertainment bias compounded by reference genome deficiency do not operate independently. They interact, reinforce each other, and create feedback loops that make incremental adjustment within the existing paradigm an insufficient response. A more accurate genomic chip helps little if the reference population is too small; a larger reference population helps little if the phenotypic records it contains do not measure the traits that matter in tropical systems; and better tropical phenotyping helps little if the bioinformatic capacity to analyse the resulting data does not exist within the institutions responsible for tropical livestock improvement. Structural reform, not marginal optimisation is what the evidence demands.

The path forward requires deliberate, sustained, and strategically coordinated investment in infrastructure that is designed from the outset for tropical purposes: reference pangenomes incorporating diverse indicine and buffalo genome assemblies; low-cost, ascertainment-unbiased genotyping approaches calibrated to tropical breed LD architectures; standardised phenotyping protocols capturing the adaptive and productive traits economically relevant to smallholder systems; multi-environment genomic prediction models that acknowledge and exploit rather than ignore genotype-by-environment interaction; and South-South scientific consortia with the mandate, the resources, and the institutional legitimacy to build what INTERBULL built for temperate breeds over the considerably compressed timeframe that food security and climate adaptation imperatives now require.

The farmers who depend on Murrah buffalo for their livelihoods, who have selected Garole sheep for their extraordinary prolificacy through generations of empirical observation, and who rely on trypanotolerant cattle to maintain viable livestock enterprises in tsetse-endemic landscapes deserve the same quality, rigour, and practical impact of genomic science that has transformed dairy farming in the Netherlands, Canada, and Denmark. Delivering that science demands not merely the application of better tools to familiar problems, but the intellectual honesty to ask from first principles, without constraint by paradigmatic inertia what genomic science for tropical livestock systems should look like when it is genuinely designed for those systems, the farmers who depend on them, and the animals that are uniquely adapted to sustain them.

AUTHOR CONTRIBUTIONS

Conceptualisation: B.B.; Methodology: B.B.; Investigation: B.B.; Writing original draft: B.B.; Writing review and editing: B.B.; Supervision: B.B. The author has read and agrees to the published version of the manuscript.

FUNDING STATEMENT

This review received no specific external funding. The analysis, synthesis, and conclusions presented are those of the author based on the cited published literature.

CONFLICTS OF INTEREST

The author declares no conflicts of interest with respect to the content, findings, or conclusions presented in this review.

REFERENCES

1. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157(4):1819-1829.
2. VanRaden PM, Van Tassell CP, Wiggans GR, et al. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci.* 2009;92(1):16-24.
3. Hayes BJ, Bowman PJ, Chamberlain AJ, et al. Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci.* 2009;92(2):433-443.
4. FAO. *The State of the World's Biodiversity for Food and Agriculture.* FAO Commission on Genetic Resources for Food and Agriculture Assessments; 2019.
5. Rege JEO, Gibson JP. Animal genetic resources and economic development: issues in relation to economic valuation. *Ecol Econ.* 2003;45(3):319-330.
6. Daetwyler HD, Pong-Wong R, Villanueva B, et al. The impact of genetic architecture on genome-wide evaluation methods. *Genetics.* 2010;185(3):1021-1031.
7. Tibbo M, Ayalew W, Awgichew K, et al. On-station characterization of indigenous Menz and Horro sheep breeds in the central highlands of Ethiopia. *Trop Anim Health Prod.* 2006;38(7-8):569-581.
8. Sargolzaei M, Schenkel FS, Jansen GB, et al. Extent of linkage disequilibrium in Holstein cattle in North America. *J Dairy Sci.* 2008;91(5):2106-2117.
9. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in animal improvement. *Nat Rev Genet.* 2009;10(6):381-391

10. Gautier M, Faraut T, Moazami-Goudarzi K, et al. Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*. 2007;177(2):1059-1070.
11. Edea Z, Dadi H, Kim SW, et al. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front Genet*. 2013;4:1-10.
12. Kumar A, Mehrotra A, Kumar A, et al. Genome-wide association study for milk production and SCS in Murrah buffalo. *Front Genet*. 2021;12:617522.
13. Sharma A, Gandham RK, Ghosh A, et al. Linkage disequilibrium and haplotype block structure in Sahiwal cattle. *Anim Biotechnol*. 2015;26(4):283-289.
14. Dikmen S, Cole JB, Null DJ, et al. Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. *J Dairy Sci*. 2012;95(6):3417-3424.
15. Braz CU, Taylor JF, Bresolin T, et al. Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle. *BMC Genet*. 2019;20(1):8.
16. Loftus RT, MacHugh DE, Bradley DG, et al. Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A*. 1994;91(7):2757-2761.
17. Ajmone-Marsan P, Garcia JF, Lenstra JA. On the origin of cattle: how aurochs became cattle and colonized the world. *Evol Anthropol*. 2010;19(4):148-157.
18. de Roos APW, Hayes BJ, Goddard ME. Reliability of genomic predictions across multiple populations. *Genetics*. 2009;183(4):1545-1553.
19. Erbe M, Hayes BJ, Matukumalli LK, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density SNP panels. *J Dairy Sci*. 2012;95(7):4114-4129.
20. Makgahlela ML, Stålhammar H, Nielsen US, et al. The estimation of genomic relationships using breedwise allele frequencies among both dairy and beef cattle. *J Dairy Sci*. 2013;96(8):5364-5375.
21. Mtileni BJ, Muchadeyi FC, Maiwashe A, et al. Genetic diversity and conservation of South African indigenous chicken populations. *J Anim Breed Genet*. 2012;129(3):205-215.
22. Daetwyler HD, Calus MPL, Pong-Wong R, et al. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*. 2013;193(2):347-365.
23. Lund MS, de Roos APW, de Vries AG, et al. A common reference population from four European dairy cattle populations increases reliability of genomic predictions. *Genet Sel Evol*. 2011;43(1):1-14.
24. Nayee N, Singh A, Joshi BH, et al. Genomic selection for fertility traits in Kankrej cattle: lessons from a first-generation implementation. *Indian J Anim Sci*. 2018;88(9):1020-1026.
25. Kolmodin R, Bijma P. Response to mass selection when the genotype by environment interaction is modelled as a linear reaction norm. *Genet Sel Evol*. 2004;36(4):435-454.
26. Calus MPL, Veerkamp RF. Estimation of environmental sensitivity of genetic merit for milk production traits using a random regression model. *J Dairy Sci*. 2003;86(9):3171-3179.
27. Tantia MS, Nag R, Sodhi M, et al. Draft assembly of Indian water buffalo genome (BGI-2014). *Genomics Data*. 2016;7:122-124.
28. Talenti A, Powell J, Hemmink JD, et al. A cattle graph genome incorporating global breed diversity. *Nat Commun*. 2022;13(1):910.
29. Bickhart DM, Billingsley J, Lam JN, et al. Pangenome approaches to livestock genomes: current status and prospects for future discoveries. *Front Genet*. 2022;13:928406.
30. DAHD. *Basic Animal Husbandry Statistics 2022*. Department of Animal Husbandry and Dairying, Ministry of Fisheries, Animal Husbandry and Dairying, Government of India; 2022.
31. Benavides MV, Sonstegard TS, Keane OM, et al. Genomic regions associated with nematode resistance in indigenous Garole sheep from India. *Genet Sel Evol*. 2016;48(1):1-14.
32. Murray C, Logan S, Mwangangi J, et al. Ecological and environmental correlates of disease tolerance in East African livestock breeds. *Parasit Vectors*. 2016;9:451.
33. Joshi N, Singh A, Vij S, et al. Application of genotyping-by-sequencing in indigenous sheep breeds of India: preliminary genomic insights. *Small Rumin Res*. 2020;190:106152.
34. Mwai O, Hanotte O, Kwon YJ, et al. African indigenous cattle: unique genetic resources in a rapidly changing world. *Asian-Australas J Anim Sci*. 2015;28(7):911-921.
35. Garner JB, Douglas ML, Williams SR, et al. Genomic selection improves heat tolerance in dairy cattle. *Sci Rep*. 2016;6(1):34114.
36. Misztal I, Legarra A, Aguilar I. Using recursion to compute the inverse of the genomic relationship matrix. *J Dairy Sci*. 2013;97(6):3943-3952.
37. Espigolan R, Baldi F, Boligon AA, et al. Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics*. 2013;14:305.
38. Bickhart DM, Rosen BD, Koren S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet*. 2017;49(4):643-650.
39. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 2008;3(10):e3395.
40. de Roos APW, Hayes BJ, Spelman RJ, et al. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*. 2008;179(3):1503-1512.