GMR
The Original

ASSEMBLY-AWARE EVALUATION OF CATTLE GROWTH GWAS: REPRODUCIBILITY, REGIONAL RECURRENCE, AND PRIORITIES FOR YAK GENOMICS

Albert Iribagiza[1,2]*, Constantin Nimbona[2], Vanessa Nibigira[3], Wondossen Ayalew[1,5], Alfred Niyokwishimira[2], Liang Chunnian[1,4]

[1]Lanzhou Institute of Husbandry and Pharmaceutical Sciences, Chinese Academy of Agricultural Sciences; Key Laboratory of Yak Breeding Engineering of Gansu Province, No. 335 Jianguoyan, Xiaoxihu, Qilihe District, Lanzhou 730050, Gansu Province, China
[2]Research Center in Animal, Crop and Environmental Sciences (CRAVE), University of Burundi, B.P. 1550, Bujumbura, Burundi
[3]Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, No. 12 Zhongguancun South Street, Haidian District, Beijing 100081, China
[4]Key Laboratory of Livestock and Poultry Genetics and Breeding on the Qinghai-Tibet Plateau, Ministry of Agriculture and Rural Affairs, No. 335 Jianguoyan, Xiaoxihu, Qilihe District, Lanzhou 730050, Gansu Province, China
[5]South China Agricultural University, Guangzhou, Guangdong, China
***Corresponding author**: Albert Iribagiza; iribagizaalbert@gmail.com

ABSTRACT

Genome-wide association studies (GWAS) have identified many cattle growth loci, but differences in phenotype, population, statistical model, reference assembly, and result availability can make apparent replication misleading. We assessed whether published cattle growth GWAS could support quantitative synthesis, identified defensible regional recurrence, and defined priorities for translating cattle evidence to yak. The audit drew on a 318-record livestock-genomics library, a PubMed horizon search updated to 28 August 2026, and direct source verification. A 13-study primary cattle inventory framed the audit: ten studies supplied complete extractable locus-level data, two supplied partial main-text coordinates, and one was retained at study and chromosome-summary levels. Two additional contemporary cattle GWAS were appraised narratively, together with seven yak growth GWAS. Coordinates were compared only within verified assemblies and were never pooled across species. The ten complete-source studies supplied 1,163 marker-test records after duplicate removal. Within the original nine-study extraction, 399 of 450 records had verified assemblies and formed the primary coordinate-analysis subset; 33 coordinates recovered from main texts were retained for sensitivity analysis. Two independent populations supported a phenotype-compatible BTA1 region near VEPH1, indicating regional rather than causal-variant recurrence. Heterogeneous designs and incomplete effect reporting precluded quantitative meta-analysis of the seven yak GWAS. An assembly-aware, phenotype-compatible synthesis is therefore more defensible than cross-species effect pooling. Cattle loci should guide yak validation only when supported by orthology or conserved synteny, compatible phenotypes, and independent functional evidence. This framework links regional prioritization to functional annotation, fine-mapping, metabolomic integration, and precision yak breeding.

KEYWORDS: cattle; yak; growth traits; genome-wide association study; evidence synthesis

1. INTRODUCTION

Growth is a central component of biological fitness and production efficiency in cattle. Birth weight, pre- and post-weaning gain, mature weight, skeletal dimensions, muscularity, and carcass composition influence feed requirements, age at slaughter, reproductive management, calving difficulty, and the economic value of dairy and beef animals. These phenotypes are related but not interchangeable: body weight reflects the combined development of bone, muscle, adipose tissue, viscera, and gut fill, whereas height, body length, chest girth, muscling scores, and carcass measurements capture more specific developmental components. Age, sex, maternal effects, nutrition, management, climate, and breed type further modify their expression (Hozakova et al., 2020; Seabury et al., 2017). A genomic synthesis must therefore preserve phenotype definition rather than treat every measurement labelled "growth" as the same outcome.

Cattle growth traits are moderately to highly heritable in many populations but have a predominantly polygenic architecture. Linkage mapping and candidate-gene studies initially highlighted major biological regulators, while dense SNP arrays, imputed whole-genome sequence, high-coverage resequencing, and single-step genomic methods subsequently enabled genome-wide discovery. Early cattle GWAS demonstrated genome-wide associations for birth and postnatal growth in crossbred beef cattle, feedlot growth and intake traits, and longitudinal growth-curve parameters (Snelling et al., 2010; Bolormaa et al., 2011; Crispim et al., 2015). Later studies extended this work to dairy, beef, indigenous, and composite populations, including Holstein, Charolais, Simmental, Nellore, Braunvieh, Belgian Blue, Red Angus, Blanco Orejinegro, Romosinuano, and African indigenous cattle (Jahuey-Martinez et al., 2016; Seabury et al., 2017; Terakado et al., 2018; An et al., 2020; Zhuang et al., 2020; Vanvanhossou et al., 2020; Zepeda-Batista et al., 2021; Smith et al., 2022; Bejarano et al.,

2023; Gualdron Duarte et al., 2023; Zhang et al., 2025a). However, the associated markers reported by one population frequently fail to recur as exact variants in another because allele frequencies, linkage disequilibrium, breed history, phenotype measurement, power, and genomic resolution differ.

Several loci nevertheless have strong biological and cross-population support. The PLAG1 region on BTA14 has pleiotropic effects on stature, body weight, reproduction, and production traits, with functional work supporting altered PLAG1 regulation as an underlying mechanism (Fink et al., 2017; Takasuga, 2016). The NCAPG-LCORL interval on BTA6 is repeatedly associated with fetal growth, mature stature, lean growth, carcass weight, and feed efficiency. Recent cattle analyses further implicate LCORL and STC2 in body size and growth-rate variation (Majeres et al., 2024; Bai et al., 2025). These examples show that credible gene prioritization requires convergence of positional, statistical, functional, and phenotypic evidence rather than recurrence of a gene name alone.

The interpretation of cattle GWAS has also moved beyond nearest-gene annotation. Most associated variants and QTL fall in noncoding regions, where the regulated gene may be distant from the lead marker. AnimalGWASAtlas integrates livestock GWAS/QTL evidence with chromatin interaction, epigenomic, conservation, transcription-factor, and tissue information to improve regulatory annotation (Gou et al., 2025). Integrative cattle studies similarly combine GWAS with expression, splicing, transcriptomic, or multi-omics information to prioritize causal genes and distinguish positional candidates from biologically supported targets (Liu et al., 2025; Ghoreishifar et al., 2026). These resources strengthen downstream interpretation, but they do not correct an erroneous chromosome label, an unspecified reference assembly, or a phenotype mismatch in the source literature.

Earlier reviews established the importance and limitations of livestock GWAS. Sharma et al. (2015) summarized early livestock GWAS successes and challenges, while Sahana et al. (2023) provided cattle-specific guidance for genotype quality control, relatedness, population structure, significance testing, and post-GWAS interpretation. These reviews provide methodological context, but neither performs the source-level, assembly-aware audit of cattle growth loci undertaken here.

Previous cattle growth syntheses answer different questions. Candidate-gene reviews and meta-analyses evaluate predefined polymorphisms, whereas coordinated GWAS meta-analyses combine harmonized genome-wide results. Bouwman et al. (2018) analysed stature in 58,265 cattle from 17 populations, and Sanchez et al. (2023) analysed growth, morphology, and carcass traits in 54,782 animals from 15 populations. Multitrait methods have also combined correlated traits or effect directions to characterize pleiotropy (Bolormaa et al., 2014; Xiang et al., 2020; Ma et al., 2023). These are genuine meta-analytic achievements, but they are not literature-level audits of heterogeneous published locus tables.

Additional cattle meta-GWAS illustrate why the term meta-analysis should be reserved for harmonized statistical inputs. Pausch et al. (2017) standardized breed-specific allelic effects and standard errors and applied inverse-variance meta-analysis, while Marete et al. (2018) combined P values, effect directions, and sample-size weights across three dairy populations. Although their principal outcomes were not identical to the present growth synthesis, both demonstrate the need for complete marker-level inputs and explicit heterogeneity assessment. Cross-breed, cross-study, and cross-trait synthesis are not interchangeable designs.

Formal SNP-effect meta-analysis has strict data requirements. The same variant, or a defensible linkage-disequilibrium proxy, must be represented across independent studies; effect alleles and genomic builds must be aligned; phenotypes and units must be compatible; and an effect estimate with uncertainty, or another prespecified summary-statistic framework, must be available. Bouwman et al. (2018), Xiang et al. (2020), and Sanchez et al. (2023) could satisfy these conditions because they had coordinated access to genome-wide study results. In contrast, the published cattle literature often supplies only significant lead markers, genomic windows, Manhattan plots, candidate genes, or incomplete supplementary tables. Pooling such records as if they were commensurate effects can create artificial replication and false-positive biological conclusions.

A second challenge is coordinate provenance. Cattle studies report positions on UMD3.1, ARS-UCD1.2, ARS-UCD2.0, or an unstated assembly; the same numerical position is not comparable across builds. PDF extraction can shift rows or chromosome labels, candidate-gene assignments may use different flanking windows, and repeated rows can represent multiple traits, ages, breeds, or analytical models rather than independent loci. Consequently, native supplementary files, explicit source locations, assembly verification, phenotype harmonization, and marker-level deduplication are prerequisites for regional comparison. Without these safeguards, expanding the analysis from breeds to different species would compound rather than resolve the problem.

Our targeted verification of previous cattle reviews and meta-analyses did not identify a synthesis that combined five features: source-level validation of reported growth-associated loci; explicit genome-assembly provenance; harmonization of weight, body-dimension, and muscling/carcass phenotype families; sensitivity analysis across alternative regional windows; and a graded distinction among exact-variant replication, phenotype-compatible regional recurrence, gene-name recurrence, and cross-trait proximity. This feature-based gap, rather than a claim that cattle growth genomics has never been reviewed or meta-analyzed, defines the contribution of the present study.

Accordingly, cattle were retained as the primary locus-level evidence base, while yak was incorporated as the focal translational species. The objectives were to (i) determine whether eligible cattle studies provide the information required for conventional SNP-effect meta-analysis; (ii) verify reported markers, coordinates, genome assemblies, phenotype definitions, and dataset independence; (iii) identify phenotype-compatible recurrent cattle

regions using conservative and sensitivity windows; (iv) evaluate which cattle findings have defensible orthologous, syntenic, or functional relevance to yak; and (v) define evidence and reporting priorities for future yak GWAS and metabolomic GWAS. Species-specific coordinates were never pooled.

2. MATERIALS AND METHODS

2.1 Evidence-synthesis design, information sources, and search provenance

This study is reported as a focused critical evidence synthesis rather than a PRISMA systematic review. The available project archive did not contain a complete database-specific search log or a mutually exclusive exclusion ledger; therefore, a retrospective PRISMA flow diagram was not reconstructed. Claims are restricted to the source-audited cattle-yak corpus described below.

The supplied file Documents exportés_318paper.csv contained 318 Zotero records spanning livestock GWAS, candidate-gene studies, genomic prediction, reviews, methodological papers, metabolomics, and contextual biology. It was treated as a mixed reference library for evidence discovery and bibliographic verification, not as 318 independently screened eligible reports. Duplicate copies and off-topic records were not converted into invented exclusion counts.

The focused corpus was established by assessing the curated cattle set, classifying yak-related records in the supplied library, and updating verification through 28 August 2026. The primary cattle inventory contained 13 GWAS: ten with complete extractable locus-level data, two with partial main-text coordinates, and Ojo et al. (2026) at study and chromosome-summary levels because exact coordinate overlap was not testable from the publication package. Korkuć et al. (2025) and Santana et al. (2025) were eligible contemporary cattle growth GWAS but were appraised narratively because they were outside the frozen coordinate dataset. Mota et al. (2025) was retained separately as contextual genomic-prediction evidence. Seven peer-reviewed yak growth GWAS met the updated cutoff, including Wang et al. (2026).

The analytic period was 1 January 2016 to 28 August 2026. Full texts and accessible supplements were checked for species, phenotype, genome-wide design, source availability, and extractable study characteristics. Because the 318-record export was a mixed reference library rather than a screening ledger, this focused reproducibility audit does not estimate study prevalence or claim exhaustive database coverage.

A reproducible PubMed horizon search was run on 28 August 2026 using cattle, bovine, yak, or Bos grunniens terms combined with growth, body-weight, body-size, conformation, average-daily-gain, or yearling terms and GWAS or genome-wide-association terms. It returned 320 records before relevance assessment. The broad result set contained many off-topic records because the keywords appeared contextually; it was therefore used to detect potentially missed or newly published studies, not as a retrospective PRISMA denominator. The complete query and its permitted interpretation are reported in Supplementary Table S10.

2.2 Prior-review verification and novelty assessment

Before defining the contribution of the review, targeted searches were used to identify previous cattle systematic reviews, narrative genomic reviews, single-gene meta-analyses, coordinated GWAS meta-analyses, and multitrait association analyses concerning growth, stature, morphology, carcass traits, or named growth-related polymorphisms. Comparator publications were classified by evidence unit: genome-wide summary statistics, literature-derived SNP/QTL coordinates, single-gene genotype associations, genetic-parameter estimates, or narrative/systematic review. For each comparator, the phenotype and population scope, input data, assembly handling, allele harmonization, and quantitative method were recorded.

The novelty claim was prespecified as a feature comparison rather than an absence claim. The manuscript therefore does not claim that cattle growth genomics has not been reviewed or meta-analyzed. Instead, it asks whether a prior synthesis combined all five features used here: source-level verification of reported loci, explicit genome-assembly provenance, phenotype-family harmonization, distance-window sensitivity, and an evidence hierarchy separating exact-variant, regional, gene, and pathway recurrence.

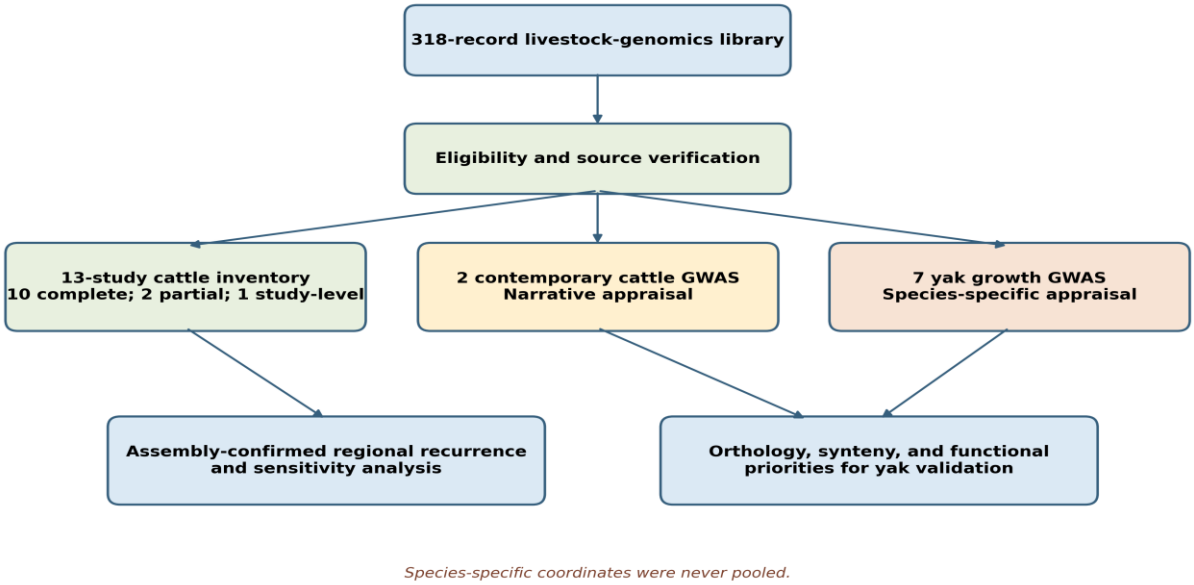

Figure 1. Evidence provenance and focused cattle-yak synthesis
The figure reports only information supported by the supplied library, verified supplementary searches, and focused evidence audit. The 13-study cattle inventory comprises ten complete-source datasets, two partial-source datasets, and one study-level/chromosome-summary dataset. This is an evidence-provenance diagram, not a reconstructed PRISMA flow.

2.3 Study scope and evidence eligibility
The review question was structured using a PECO/SD framework appropriate for observational genomic association evidence: Population, Bos taurus cattle populations; Exposure, inherited SNP or QTL variation evaluated genome-wide; Comparator, alternative alleles, genotypes, haplotypes, or genomic regions as defined by each study; Outcomes, growth, weight, linear body dimensions, conformation, muscling, feed-efficiency, and carcass phenotypes; and Study design, GWAS, sequence-based GWAS, or single-step genome-wide association analysis. The review protocol was not prospectively registered. This limitation is disclosed because the cattle-only revision developed from an archived seven-species project rather than from a newly registered protocol.
The focused primary synthesis included cattle and yak GWAS addressing body weight, average daily gain, linear body dimensions, conformation, muscular development, feed-efficiency-associated growth, or carcass growth phenotypes. Study-level eligibility required an on-target measured phenotype, a genome-wide or single-step genome-wide association design, adequate reporting of association testing, and an accessible full report. Locus-level coordinate analysis additionally required extractable chromosome positions or intervals and a documented reference assembly. Candidate-gene association studies, gene-editing experiments, selection scans without direct phenotype association, reviews, differential metabolomics, and prediction-only reports did not contribute primary loci.
2.4 Study-characteristics variables and data extraction
Data were recorded in a 41-field dictionary spanning bibliographic information, population and sampling, phenotype definition, genotyping and quality control, association-model specification, locus-level results, and extraction provenance. Study-level fields included breed, country or study region, sample size, sex, age or time point, verbatim trait, harmonized trait family, genotyping platform, post-quality-control marker count, imputation, association model, covariates, population-structure correction, multiple-testing threshold, genome assembly, and summary-statistics availability. Country/region was retained only to describe the geographic and population context of the evidence, not as an analytical moderator; it is reported with breed in Supplementary Table S1. Locus-level fields included marker, chromosome, coordinate or interval, alleles, effect estimate, standard error, allele frequency, P value, variance explained, candidate gene, source table, and extraction notes. Original coordinates and source locations were preserved. The four evidence-dataset denominators and their permitted analytical uses are summarized in Supplementary Table S3.
2.5 Assembly verification and coordinate quality control
Coordinates were assigned to UMD3.1, ARS-UCD1.2, or ARS-UCD2.0 only when the source explicitly identified the assembly. The Charolais whole-genome resequencing study explicitly aligned reads to ARS-UCD1.2, and the Belgian Blue and Beninese studies also reported ARS-UCD1.2/ARS1.2 coordinates. The Creole ssGWAS paper identified the SNP arrays but did not state the coordinate assembly in the main text; its windows were therefore excluded from coordinate comparisons pending manifest verification.

Source-level checking of the Belgian Blue results identified chromosome-label shifts caused by stateful parsing of a wrapped PDF table. The native Excel supplement was treated as authoritative. Fifteen of 37 extracted chromosome labels were corrected, after which all regional analyses were rerun. This audit step prevented false BTA5 and BTA19 overlaps with Charolais loci.

2.6 Cattle-genomics database triangulation

Primary publications and native supplements remained the authoritative sources for inclusion, phenotype, effect, and coordinate extraction. Animal QTLdb Release 59 (20 April 2026), AnimalGWASAtlas, NCBI Assembly, and Ensembl were used only for contextual triangulation of trait terminology, marker/gene identifiers, assembly metadata, and candidate-region interpretation (Hu et al., 2022; Gou et al., 2025; Dyer et al., 2025). Release 59 contained 212,372 cattle QTL/eQTL/association records curated from 1,231 publications. Database occurrence was not counted as an independent replication because a database record may reproduce the same primary publication included in this review. Coordinates were never transferred between UMD3.1, ARS-UCD1.2, ARS-UCD1.3, and ARS-UCD2.0 solely because a database displayed an updated position; any future lift-over must retain the original coordinate, target assembly, mapping tool or chain, mapping status, and ambiguity flag. Ensembl release 116 uses ARS-UCD2.0 as its current cattle reference, whereas much of the included evidence was reported on ARS-UCD1.2 or UMD3.1.

2.7 Yak evidence and cattle-yak translation

Primary yak evidence required a genome-wide association design, an explicitly measured growth or body-size phenotype, an accessible peer-reviewed report, and sufficient population and analytical information for study-level appraisal. Seven reports met these criteria through 28 August 2026: Jia et al. (2020), Wang et al. (2022a), Jiang et al. (2022), Liu et al. (2023), Zhang et al. (2025b), Zhao et al. (2025), and Wang et al. (2026). Candidate-gene studies, selection scans without direct phenotype association, genomic-prediction-only reports, and non-peer-reviewed preprints were excluded from the primary yak GWAS count.

Cross-species comparison was performed through verified one-to-one orthologues, conserved syntenic regions, compatible phenotype families, and supporting functional evidence. Numerical cattle and yak coordinates were not compared directly. Evidence was graded from same or confidently aligned orthologous loci with compatible phenotypes to pathway-level similarity. Candidate-name recurrence without verified orthology or locus correspondence was considered exploratory.

Possible population reuse was evaluated from breed, location, recruitment period, authorship, institutional provenance, and available cohort descriptions. Recurrence within the same animals, across correlated phenotypes, or among multiple models was not treated as independent replication. Potentially overlapping studies were assigned to the same provisional dataset family until independence could be demonstrated.

2.8 Phenotype harmonization

Phenotypes were assigned to three broad families: growth/weight, body size/conformation, and muscling/carcass. Regional evidence was considered phenotype-compatible only when both studies belonged to the same family. This broad grouping was used for screening rather than to imply equivalence between specific measurements or ages. Exact-effect pooling would require substantially narrower definitions, such as the same weight trait at a comparable age.

2.9 Regional overlap and sensitivity analyses

Within each confirmed assembly and chromosome, a symmetric 250-kb flank was added to every reported SNP or QTL interval. Overlapping intervals from independent studies were identified and pairwise hits were collapsed into independent regions by chromosome and study pair. The procedure was repeated with 500-kb and 1-Mb flanks to evaluate dependence on the window definition. The 250-kb result was designated the primary analysis because it was the most conservative of the prespecified windows. The source-validated primary regions are reported in Table 1, and the alternative-window results are shown in Figure 4.

2.10 Eligibility for conventional SNP-effect meta-analysis

A variant was eligible for inverse-variance meta-analysis only if it was reported in at least two independent studies with a compatible phenotype, the same reference/alternate context or an alignable effect allele, beta, and standard error. No variant satisfied all criteria; therefore, pooled beta estimates, heterogeneity statistics, and forest plots were not calculated.

This decision rule is consistent with established GWAS meta-analysis practice, in which study-specific association results are harmonized by marker identity, genomic position, coded allele, effect direction, effect estimate, and uncertainty before pooling (Willer et al., 2010; Evangelou and Ioannidis, 2013). Pausch et al. (2017) provide an inverse-variance cattle example, whereas Marete et al. (2018) demonstrate a weighted-Z design using P values, directions, and per-breed sample sizes. None of these inputs can be reconstructed reliably from a catalogue restricted to significant lead loci; regional proximity or recurrence of a candidate-gene name was therefore not converted into a pooled effect size.

2.11 Study-level quality appraisal

Each study was appraised descriptively across eight domains: population definition and effective sample size; phenotype measurement and adjustment; genotyping/sequencing quality control; relatedness and population structure; covariate specification; multiple-testing control; genome-assembly and coordinate reporting; and completeness of locus-level results. No numerical total score was calculated because no universally accepted livestock-GWAS risk-of-bias instrument exists and the domains are not exchangeable. The appraisal evaluated

whether each design was adequate for its stated objective and sufficiently reported for verification; a study was not penalized merely for finding few loci.

2.12 Data completeness and missing information

Missing information was treated explicitly. A blank beta, standard error, allele frequency, marker identifier, or assembly was coded as unavailable rather than inferred from a P value, Manhattan plot, nearby database record, or another publication. Analysis eligibility was determined from the observed field completeness; the resulting source strata and analytical restrictions are reported in Results and Supplementary Table S3. Corresponding authors or repositories should be contacted for genome-wide summary statistics before any future effect-size meta-analysis.

2.13 Locus deduplication, independence, and evidence hierarchy

Within-study duplicate records were checked using study identifier, trait, chromosome, coordinate or interval, marker identifier, and source table. Multiple traits associated with the same marker were preserved as separate evidence rows because they represent distinct phenotype tests. Multiple markers within a study were not assumed to be independent; consequently, the number of extracted rows was never used as a statistical weight. Regional matches were accepted only across different studies, and pairwise locus matches collapsed into an independent regional cluster to prevent dense reporting by one study from producing multiple apparent replications.

Evidence was classified hierarchically. Level 1 required the same variant, compatible phenotype, aligned effect allele, beta, and standard error in independent studies. Level 2 required the same assembly and a phenotype-compatible regional overlap under the prespecified window. Level 3 comprised gene recurrence or a broader-window regional overlap without exact-variant evidence. Level 4 comprised cross-trait or cross-species proximity and was considered exploratory. Only Level 1 would justify a conventional pooled effect; Level 2 supported regional prioritization; Levels 3 and 4 were hypothesis-generating.

2.14 Visualization strategy and interpretation safeguards

The main figures were prespecified to answer distinct questions. The evidence-provenance diagram reports source identification and analytic inclusion without implying a reconstructed PRISMA screening ledger. A study-contribution plot shows the number of extracted locus records per study but is not interpreted as study precision or weight. A reported-locus landscape places the available significant or suggestive loci along the 29 bovine autosomes and plots minus log10(P) when a numerical P value is available. This landscape is explicitly not labelled as a conventional Manhattan plot because the extracted literature dataset lacks the complete set of tested variants, including null and non-significant results. The interpretation limits for each genomic figure are stated in the corresponding caption and Results section.

A conventional quantile-quantile plot was not produced. QQ plots compare the observed distribution of genome-wide test statistics or P values with their expected null distribution and are used to evaluate calibration and residual inflation. Constructing one from a significance-selected subset would guarantee severe truncation and create a misleading impression of genomic inflation. Valid study-specific Manhattan and QQ plots require full GWAS summary statistics, or at minimum all tested marker P values, from each study. If those files are obtained, plots should be generated separately by study and phenotype before any harmonized meta-analysis.

2.15 Reproducibility and sensitivity analyses

Every extracted row retained a study identifier, source file, source table or supplement, reported build, original coordinate, correction note, and evidence status. Corrections were logged rather than silently overwriting provenance. Regional analysis was rerun after the Belgian Blue chromosome audit, and results were compared across 250-kb, 500-kb, and 1-Mb flanking windows. The 250-kb rule was designated primary before interpreting recurrence, while the broader windows quantified sensitivity to an uncertain linkage-disequilibrium scale.

The analysis package separates raw extraction, harmonized loci, source-row validation, corrected coordinates, region clusters, and manuscript-ready summaries. This layered structure permits an auditor to trace every highlighted region back to its source and to rerun the overlap analysis after new supplementary data are obtained. Reference-assembly reporting follows the principle that genomic coordinates are meaningful only within a named assembly; cattle loci were therefore stratified among UMD3.1, ARS-UCD1.2, ARS-UCD2.0, and unreported builds rather than pooled across coordinate systems (Rosen et al., 2020).

3. RESULTS

3.1 Study and locus coverage

The 13-study primary cattle inventory comprised Liu et al. (2025), An et al. (2020), Zepeda-Batista et al. (2021), Zhang et al. (2017), Zhang et al. (2025a), Bila et al. (2026), Seabury et al. (2017), Bejarano et al. (2023), Ma et al. (2023), Gualdrón Duarte et al. (2023), Zhuang et al. (2020), Vanvanhossou et al. (2020), and Ojo et al. (2026). Mota et al. (2025) was retained only as contextual genomic-prediction evidence. Ten previously extracted studies supplied complete numerical results. The original nine-study extraction contained 450 records, of which 399 had sufficiently verified assemblies for coordinate analysis. Addition of the audited Seabury et al. workbook produced 1,163 complete-source marker-test records. Thirty-three coordinates from incomplete-source publications were retained separately for sensitivity analysis.

The contribution of locus records was uneven: the Charolais resequencing study contributed 173 records, followed by the Simmental body-size GWAS (63), Chinese Simmental weighted single-step GWAS (60), Creole cattle (51), Belgian Blue (37), Beninese cattle (28), Holstein (27), Sussex (8), and Braunvieh (3). These counts reflect

reporting granularity and extraction availability and were not used as analytical weights. The distribution is shown in Figure 2; Supplementary Table S3 defines the corresponding evidence-dataset denominators.

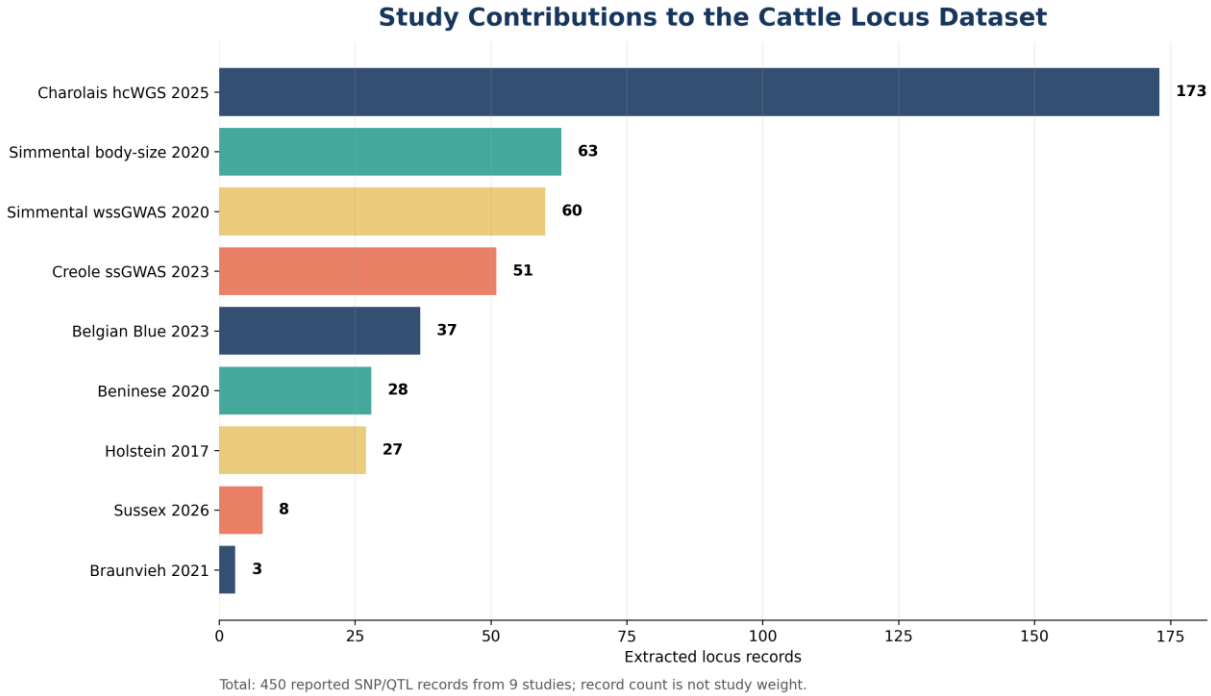

Figure 2. Study contributions to the extracted cattle locus dataset
Bars represent extracted SNP/QTL records, not sample size, precision, or meta-analytic weight.

3.1.1 Audit of outstanding supplementary evidence

The second Simmental body-size study was recovered from main Table 1. Sixty-three marker-by-analysis or trait rows representing 52 unique marker-position pairs were extracted on the explicitly reported UMD3.1 assembly. Repeated markers were retained when reported under different single-trait, multitrait, or longitudinal analyses, but overlapping evidence within a study was collapsed before cross-study comparison. Adding these rows generated no new cross-study regional match at +/-250 kb, +/-500 kb, or +/-1 Mb (An et al., 2020).

The U.S. beef supplementary workbook was obtained and audited. It contained 717 association rows across Angus, Hereford, and SimAngus analyses of residual feed intake, dry-matter intake, average daily gain, and mid-test metabolic weight. Four exact duplicate copies were removed, leaving 713 valid marker-test records and 702 unique marker-position pairs. Fifteen records were located in the pseudoautosomal region. All retained rows met the study's reported P-value, variance-explained, and minor-allele-frequency criteria. Because reference-assembly provenance could not be established with sufficient confidence from the available source, these coordinates were retained for descriptive synthesis but not treated as assembly-confirmed cross-study replication (Seabury et al., 2017).

The Huaxi cattle study reported 99 significant associations across seven growth or carcass phenotypes, but the complete association table was unavailable because the underlying commercial-farm data were restricted. Eight explicit ARS-UCD2.0 coordinates were recoverable from the main article. Only BTA8:65,120,621 was unambiguously connected to body weight in the available text; the remaining loci were retained with broader or unresolved phenotype labels in the partial-source sensitivity table (Liu et al., 2025).

The Holstein-heifer study cited two Figshare supplements, but the cited DOI landing pages did not provide retrievable files during the audit. Twenty-five unique ARS-UCD1.2 coordinates were recovered from the article: 23 could be assigned to specific growth or body-dimension phenotypes and two retained unresolved phenotype categories. These records were analysed as partial-source sensitivity evidence and not as a complete representation of the study (Ma et al., 2023).

Candidate-gene frequency and regional recurrence were interpreted by source-completeness tier. The 1,163-record complete-source dataset supports descriptive extraction, and the 1,196-record sensitivity dataset adds 33 recoverable main-text coordinates. The highlighted VEPH1 region is based on the 399-record assembly-confirmed coordinate subset and is reported as phenotype-compatible regional recurrence, not exact-variant replication. The larger datasets are not allowed to create coordinate matches until their reference assemblies and phenotype mappings meet the same prespecified rules.

3.2 Distribution of reported association evidence

The published-locus landscape describes the original 450-record extraction from nine studies. Within that extraction, 399 records had verified assembly provenance and formed the primary coordinate-analysis subset. This subset was retained for regional recurrence because the additional Seabury records lacked sufficiently verified assembly provenance and the partial-source records did not represent complete association outputs. The

updated 1,163-record dataset is therefore used for descriptive source accounting, while coordinate recurrence is restricted to the 399 assembly-confirmed records. This separation prevents unresolved coordinates from creating artificial overlap.

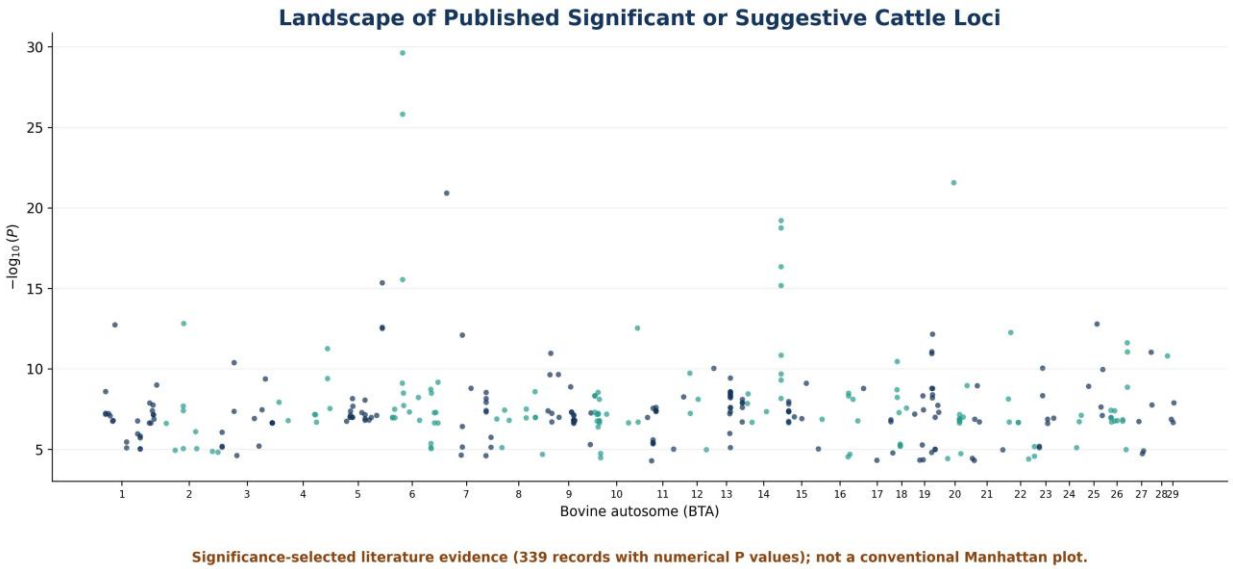

Figure 3. Landscape of published significant or suggestive cattle loci.

The in-figure warning distinguishes this significance-selected evidence display from a conventional Manhattan plot based on all tested variants.

This figure is not a conventional Manhattan plot because non-significant genome-wide variants were unavailable. A QQ plot was not generated for the significance-selected literature dataset.

3.3 Corrected conservative regional synthesis

After correcting the Belgian Blue chromosome labels, the 250-kb analysis identified three independent cross-study regions (Table 1). Only the BTA1 region connected studies within the same broad phenotype family. The Beninese study reported rs109126926 at 110,160,486 bp for height at withers (P = 1.12 x 10^-6), and the Charolais study reported a position at 110,115,447 bp for body length at three months (P = 1.76 x 10^-7). The lead coordinates were separated by 45,039 bp. Both studies annotated VEPH1, whereas the Charolais record also included PTX3.

The BTA7 overlap linked a chromosome-wide suggestive Beninese ear-length SNP with a cluster of significant Charolais backfat-thickness variants near RIOK2 and LIX1. The BTA19 overlap linked a chromosome-wide suggestive Beninese heart-girth SNP with a Belgian Blue MRC2 frameshift QTL associated with top muscling, rib shape, and buttock rear. Because these two regions involved different trait families and lacked a shared reported gene, they were retained as exploratory observations rather than replicated growth loci.

Table 1. Source-validated cross-study regions detected with the primary +/-250-kb rule.

Priority	BTA	Study 1	Study 2	Distance	Decision
Primary	1	Beninese: height at withers; rs109126926; VEPH1	Charolais: 3-month body length; VEPH1/PTX3	45,039 bp	Retain
Exploratory	7	Beninese: ear length; suggestive SNP	Charolais: 18-month backfat thickness; RIOK2/LIX1	~436-442 kb	Cross-trait only

Exploratory	19	Beninese: heart girth; Belgian Blue: muscling traits; suggestive SNP	MRC2	365,572 bp	Cross-trait only

Sources: Vanvanhossou et al. (2020), Gualdron Duarte et al. (2023), and Zhang et al. (2025a), with coordinates checked against the cited main and supplementary tables.

3.4 Sensitivity to window width

The corrected analysis identified five independent regions at +/-500 kb and six at +/-1 Mb, compared with three at +/-250 kb. Same family regions numbered one, three, and five, respectively. The expected increase at broader windows demonstrates that overlap counts are sensitive to the distance rule. Accordingly, broad-window-only findings should not be elevated to primary replication evidence. The full window-width comparison is presented in Figure 4.

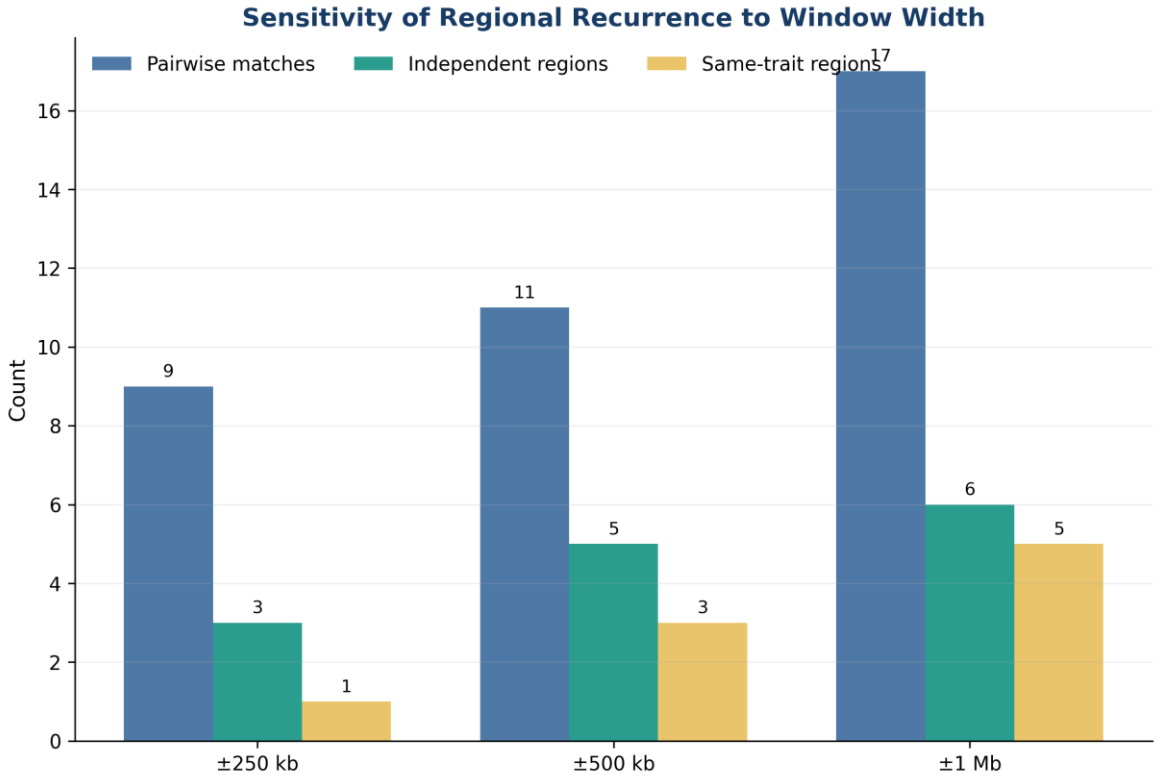

Figure 4. Sensitivity of regional recurrence to the flank width.

Broader windows increased pairwise matches and same-trait regions, demonstrating dependence on the spatial rule.

3.5 Effect-size meta-analysis feasibility

No exact marker recurred with the information required for quantitative pooling. The Braunvieh study directly reported beta and standard error, and selected Belgian Blue and Beninese results contained effect information, but the variants and phenotypes were not independently replicated in a compatible form. The statistically appropriate result was therefore a decision not to pool SNP effects. Source strata and their permitted analytical uses are documented in Supplementary Table S3, while the appraisal domains are reported in Supplementary Table S5.

3.6 Yak growth GWAS evidence

The August 2026 horizon search identified two eligible contemporary cattle studies outside the frozen coordinate dataset. Korkuć et al. (2025) analysed birth weight, body weight, and weight gain in 669 German Black Pied bulls and reported 14 loci using the ARS-UCD1.2 assembly. Santana et al. (2025) analysed yearling weight at approximately 378 days in 5,070 Nellore males using a single-step genomic reaction-norm GWAS; pasture- and feedlot-response windows were reported on ARS-UCD1.3. These studies reinforce the review's central argument that environment, phenotype, assembly, and reporting unit must be harmonized before recurrence or effect pooling is attempted. They were not retrofitted into the frozen coordinate analysis because doing so without reconstructing the complete source-level extraction would change its denominator after analysis.

Seven primary yak growth GWAS met the updated cutoff of 28 August 2026. Six were identified in the earlier verified corpus: Jia et al. (2020), Wang et al. (2022a), Jiang et al. (2022), Liu et al. (2023), Zhang et al. (2025b), and Zhao et al. (2025). Wang et al. (2026) added a peer-reviewed GWAS and genomic-selection analysis of average daily gain in 474 Ashidan yaks using high-depth whole-genome resequencing. The seven studies are summarized in Supplementary Table S2.

The seven yak studies differed substantially in population, sex, age, genomic platform, assembly, association model, and significance threshold. Agreement among several models applied to the same animals was interpreted as within-dataset analytical robustness rather than independent replication. Potential cohort overlap was assessed from breed, location, recruitment period, authorship, and institutional provenance. The yak literature therefore provides a growing discovery base, but independently replicated growth loci remain limited.

3.7 Cattle-yak transferability and mGWAS gap

No cattle-yak association was considered replicated solely because similar gene names or chromosome numbers were reported. Transferability required a verified orthologous gene or conserved syntenic interval, a compatible growth phenotype, and explicit consideration of study independence. On the presently verified evidence, no Level 1 cross-species replication based on an aligned causal or lead variant could be established. Cattle candidates should therefore be used to prioritize yak regions for validation rather than to define a universal bovine growth architecture.

The targeted evidence review did not identify an eligible peer-reviewed direct metabolomic GWAS in which genome-wide variants were tested against metabolite phenotypes specifically in relation to yak growth. Differential metabolomics and integrative genomic-metabolomic studies remain biologically informative but were not classified as direct mGWAS. This gap provides a focused rationale for future yak studies integrating genome-wide variants, quality-controlled metabolite intensities, growth phenotypes, fine-mapping, and independent validation.

4. DISCUSSION

4.1 Principal findings and strength of evidence

The principal result is methodological as well as biological. Among the 12 cattle studies represented in the existing locus-level datasets (within the 13-study eligible inventory), no exact variant was reported in at least two independent studies with a compatible phenotype, an aligned effect allele, a beta coefficient, and a standard error. Ojo et al. (2026) could not be tested for exact overlap because the publication package omits the significant-marker identifiers and coordinates. A conventional inverse-variance SNP-effect meta-analysis was therefore not estimable for the updated inventory. This is not a negative result: it identifies the precise reporting and harmonization barrier that must be overcome before quantitative pooling can be justified.

Within the source-validated 450-record dataset, the conservative +/-250-kb analysis identified three cross-study regions, but only one connected compatible phenotype families. The BTA1 region near 110.1 Mb linked a Beninese height-at-withers association with a Charolais three-month body-length association. The lead positions were separated by 45,039 bp, were reported on ARS-UCD1.2, and both source studies annotated VEPH1. This convergence is stronger than gene-name recurrence alone because it combines independent populations, positional agreement, assembly agreement, and related body-dimension phenotypes. It should nevertheless be described as a recurrent region, not replication of one causal SNP.

The BTA7 and BTA19 overlaps remain exploratory. They connected different trait families, depended partly on suggestive evidence, and did not provide a shared reported gene. Their presence demonstrates why proximity alone cannot establish biological replication. The increase from three regions at +/-250 kb to five at +/-500 kb and six at +/-1 Mb further shows that recurrence is sensitive to the spatial rule; broad-window findings require stronger external evidence before they are promoted.

4.2 Position relative to previous cattle reviews and meta-analyses

The present contribution must be positioned beside, rather than above, established cattle meta-analyses. Bouwman et al. (2018) identified 163 stature-associated regions after harmonizing approximately 25.4 million sequence variants across 17 populations. Sanchez et al. (2023) used 54,782 animals from 15 populations and aligned sequence-level summary statistics for growth, morphology, and carcass traits. Xiang et al. (2020) used effect-direction agreement in more than 44,000 cattle to quantify pleiotropy, while Ma et al. (2023) used monthly Holstein growth measurements to detect pleiotropic markers across traits and ages. Their access to genome-wide results enabled statistical operations that cannot be reconstructed from published significant-only tables.

The comparator studies reinforce, rather than weaken, this distinction. Pausch et al. (2017) demonstrated both the additional power and the imputation-related heterogeneity possible in across-breed pooling. Marete et al. (2018) detected stature QTL from complete marker-level results but also showed that population-specific signals can lose power when heterogeneous datasets are combined. These papers are methodological comparators, not additional primary growth-locus inputs. A comparator can enter the analytic dataset only if it satisfies the same date, phenotype, study-design, source-file, assembly, and locus-extraction criteria applied to every included paper.

Candidate-gene syntheses answer a different question from GWAS evidence synthesis. Reviews of predefined GH or LEP polymorphisms evaluate selected genotype contrasts rather than the genome-wide distribution and assembly provenance of reported loci (Bila et al., 2024; Wang et al., 2022b). General GWAS reviews remain useful for methodological context, but they do not replace verification of each marker against its native source (Sharma et al., 2015; Sahana et al., 2023).

The novelty of the present study is deliberately narrow and reproducible: it tests whether heterogeneous published cattle growth-GWAS results are combinable; corrects extraction errors against native supplements; stratifies coordinates by assembly; harmonizes phenotype families; evaluates dependence on regional window size; and grades exact-variant, regional, gene, pathway, and cross-trait evidence separately. This approach directly

addresses the artifact risk created by unverified coordinate aggregation. It should be described as a focused, assembly-aware critical evidence synthesis and regional sensitivity analysis, not as a substitute for a consortium-scale GWAS meta-analysis.

4.3 Source validation, reporting quality, and implications for future meta-analysis

Returning to the Belgian Blue native spreadsheet corrected 15 chromosome-label shifts introduced during PDF-table extraction. Without that correction, false overlaps with Charolais loci would have entered the results. This finding demonstrates that source verification is part of statistical validity: a wrong chromosome or assembly can generate a completely artificial recurrent region. Reviews of genomic associations should therefore preserve the verbatim marker, coordinate, assembly, trait, analytical model, significance threshold, and source-table location for every extracted record.

The evidence hierarchy used here also limits overinterpretation. Exact-variant replication with aligned effects is the strongest level and can support quantitative meta-analysis. Phenotype-compatible regional recurrence can prioritize a QTL but cannot establish the same causal allele. Gene-name recurrence without positional agreement is hypothesis-generating, because different studies may use different annotation windows or assign the same broad gene to distinct signals. Pathway recurrence is useful for biological integration but is vulnerable to gene-set size and annotation bias. Cross-trait or cross-species recurrence requires an explicit pleiotropic or orthology-based hypothesis.

Future quantitative synthesis should prioritize acquisition of complete summary statistics from the eligible studies. Harmonization should include genome-build conversion with documented chain files, reference/alternate and effect-allele alignment, allele-frequency checks, phenotype and age matching, sample-overlap assessment, and breed-stratified sensitivity analyses. Fixed-effect estimates may be appropriate for closely harmonized effects, whereas random-effects or meta-regression approaches may be required when breed, age, production system, or phenotype definitions introduce genuine heterogeneity. Study-specific Manhattan and QQ plots should be regenerated only from complete genome-wide results; a significance-selected literature dataset cannot provide a valid genomic-inflation diagnostic.

Cattle-genomics databases should support, but not replace, this source-level workflow. Animal QTLdb provides broad trait and positional coverage and supports comparison across studies; AnimalGWASAtlas adds regulatory and multi-omics prioritization; NCBI and Ensembl provide assembly and annotation provenance. However, database scale is not evidence of independent recurrence, and database-remapped coordinates should not be mixed with author-reported coordinates without explicit build conversion. For the present synthesis, database evidence is therefore classified as external annotation support rather than an additional study or locus count.

4.4 Biological interpretation and cattle-to-yak transferability

VEPH1 is the most defensible candidate emerging from the current regional synthesis, but it remains a prioritization target rather than a confirmed growth gene. The two contributing studies support a shared body-dimension region, yet the causal variant, local linkage-disequilibrium structure, and regulatory target remain unresolved. PTX3 was additionally annotated at the Charolais position but not in the Beninese record. Functional follow-up should therefore compare VEPH1, PTX3, and other genes within the shared interval using breed-specific linkage disequilibrium, relevant tissue expression, chromatin interactions, and regulatory annotations. AnimalGWASAtlas and cattle multi-omics frameworks provide useful resources for this next step (Gou et al., 2025; Ghoreishifar et al., 2026).

The limited recurrence of extracted loci does not contradict the established roles of PLAG1, NCAPG-LCORL, and STC2. Those loci are supported by coordinated GWAS, functional studies, or fine-mapped cattle haplotypes (Fink et al., 2017; Majeres et al., 2024; Bai et al., 2025). Their incomplete recurrence in a literature-derived dataset can reflect phenotype and age differences, population-specific allele frequencies, significance thresholds, limited power, and incomplete supplements. Absence from the present regional result is therefore not evidence against their biological importance.

Yak is the only translational species retained because it is the biological focus of the doctoral project. Seven yak GWAS met the formal search cutoff and addressed weaning growth, body weight, linear dimensions, and image-derived body-size traits (Jia et al., 2020; Wang et al., 2022a; Jiang et al., 2022; Liu et al., 2023; Zhang et al., 2025b; Zhao et al., 2025). These studies constitute an independent yak evidence base; their coordinates were not pooled numerically with cattle coordinates.

Cattle-to-yak transferability should be evaluated hierarchically. The strongest claim requires an orthologous gene or conserved syntenic interval associated with a compatible phenotype in both species. Functional or pathway agreement without positional support is weaker and should be described as biological plausibility. Gene-name recurrence alone is insufficient because annotation windows, genome assemblies, linkage disequilibrium, and population history differ between cattle and yak. The focused comparison therefore prioritizes one-to-one orthology, assembly provenance, phenotype compatibility, and independent functional evidence.

The current yak literature remains limited in sample size, cohort diversity, and locus-level reporting. A genomic-prediction study provides useful context for breeding applications but is not treated as primary association evidence (Fu et al., 2019). The 2026 Ashidan-yak average-daily-gain GWAS is included in the updated seven-study inventory, but its signals require validation in independent yak populations (Wang et al., 2026).

The metabolomic-GWAS gap remains directly relevant to the project. Current livestock applications show how metabolite traits can connect association signals to biochemical pathways, but a credible yak mGWAS must use

species-specific genotype quality control, metabolite identification and normalization, covariate adjustment, multiple-testing control, and independent functional interpretation (Ren et al., 2024). Cattle evidence can guide candidate pathways and annotation, but it cannot substitute for a properly powered yak discovery design.

Taken together, the findings support a focused sequence of work: retain the 13-study primary cattle inventory with its explicit source-completeness strata; appraise Korkuć et al. (2025) and Santana et al. (2025) narratively until their complete result files can be harmonized; retain seven formal yak GWAS; keep species-specific coordinates separate; interpret only source-validated cattle recurrence; and evaluate transfer through orthology, synteny, compatible phenotypes, and pathways.

5. Supplementary material

The accompanying file contains Supplementary Tables S1-S10: cattle and yak study characteristics, evidence-dataset definitions, dataset-family independence, methodological appraisal domains, priority genes and regions, pathway themes, metabolomic evidence, evidence classification, and provenance rules. Restricted or unavailable publisher supplements were not reconstructed.

6. Limitations

- Ten cattle GWAS contributed complete extractable locus-level data, two contributed partial main-text coordinates, and one contributed study-level and chromosome-summary evidence without testable exact coordinate overlap. Source incompleteness may change the regional landscape and is handled through separate primary and sensitivity datasets.
- The Creole study was excluded from coordinate matching because the coordinate assembly was not explicitly stated in the main paper.
- The phenotype families are intentionally broad and cannot substitute for exact age- and unit-specific phenotype harmonization.
- Fixed flanking windows approximate linkage regions and do not model breed-specific linkage disequilibrium.
- Published significant loci are subject to winner's curse, selective reporting, and heterogeneous multiple-testing thresholds.
- The recurrent VEPH1 region is supported by two studies only and requires replication in additional populations.

7. CONCLUSIONS

The accessible literature does not support a conventional inverse-variance meta-analysis of cattle and yak growth loci. Numerous associations have been reported, but phenotype heterogeneity, incomplete effect reporting, uncertain assemblies, unavailable supplements, and possible population reuse substantially reduce the number of independently reproducible findings. The BTA1 region near VEPH1 remains a phenotype-compatible cattle signal for follow-up, not proof of a shared causal variant. Cattle discoveries can inform yak research when translation is based on verified orthology, conserved synteny, compatible phenotypes, and independent evidence rather than raw coordinates or repeated candidate-gene names. The absence of an eligible direct yak growth mGWAS identifies a concrete opportunity for the planned doctoral research. Future studies should publish complete summary statistics, document cohorts and assemblies, and include independent validation.

Data availability

All data analysed in this evidence synthesis were derived from the cited publications and their accessible supplementary materials. Study-level extraction, evidence classifications, source-availability decisions, and analytical limitations are reported in the accompanying supplementary material. Publisher-restricted or unavailable supplementary data were not reconstructed.

Author contributions

Conceptualization: Albert Iribagiza. Methodology: Albert Iribagiza. Literature search and screening: Albert Iribagiza, Vanessa Nibigira. Data extraction and curation: Albert Iribagiza, Vanessa Nibigira. Formal analysis: Albert Iribagiza, Vanessa Nibigira. Writing-original draft: Albert Iribagiza, Vanessa Nibigira. Writing-review and editing: Albert Iribagiza, Constantin Nimbona, Alfred Niyokwishimira, Wondossen Ayalew, Liang Chunnian. Supervision: Liang Chunnian.

Funding

The authors received no financial support for the research, authorship, or publication of this article.

Conflict of interest

The authors declare no conflict of interest.

Ethics approval: Not applicable; this review analysed published literature and did not involve new animal experimentation.

AI-assisted editing: Generative AI tools were used only to support language, structure, and consistency checks. The authors verified the searches, eligibility decisions, numerical results, interpretations, and references and accept full responsibility for the final manuscript.

REFERENCES

1. An, B., Xu, L., Xia, J., Wang, X., Miao, J., Chang, T., Song, M., Ni, J., Xu, L., Zhang, L., Li, J., & Gao, H. (2020). Multiple association analysis of loci and candidate genes that regulate body size at three growth stages in Simmental beef cattle. BMC Genetics, 21, 32. https://doi.org/10.1186/s12863-020-0837-6
2. Bai, F., et al. (2025). LCORL and STC2 variants increase body size and growth rate in cattle and other animals. Genomics, Proteomics & Bioinformatics, qzaf025. https://doi.org/10.1093/gpbjnl/qzaf025
3. Bejarano, D. H., Martinez, R. A., & Rocha, J. F. (2023). Genome-wide association study for growth traits in Blanco Orejinegro and Romosinuano cattle. Tropical Animal Health and Production, 55, 358. https://doi.org/10.1007/s11250-023-03743-9
4. Bila, L., Malatji, D. P., Putra, W. P. B., Sanarana, Y. P., & Tyasi, T. L. (2026). Genome-wide association study (GWAS) for linear body measurements of Sussex cattle (Bos taurus) at weaning and yearling age in South Africa. Veterinary Medicine and Science, 12, e70717. https://doi.org/10.1002/vms3.70717
5. Bila, L., Malatji, D. P., & Tyasi, T. L. (2024). Single nucleotide polymorphisms of growth hormone gene in cattle and their association with growth traits: a systematic review. Tropical Animal Health and Production, 56, 141. https://doi.org/10.1007/s11250-024-03985-1
6. Bolormaa, S., Hayes, B. J., Savin, K., et al. (2011). Genome-wide association studies for feedlot and growth traits in cattle. Journal of Animal Science, 89, 1684-1697. https://doi.org/10.2527/jas.2010-3079
7. Bolormaa, S., Pryce, J. E., Reverter, A., et al. (2014). A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. PLOS Genetics, 10, e1004198. https://doi.org/10.1371/journal.pgen.1004198
8. Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., et al. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics, 50, 362-367. https://doi.org/10.1038/s41588-018-0056-5
9. Bovo, S., et al. (2025). Merging metabolomics and genomics provides a catalog of genetic factors that influence molecular phenotypes in pigs linking relevant metabolic pathways. Genetics Selection Evolution, 57, 11. https://doi.org/10.1186/s12711-025-00960-8
10. Chen, B., et al. (2022). Transcriptome-metabolome analysis reveals how sires affect meat quality in hybrid sheep populations. Frontiers in Nutrition, 9, 967985. https://doi.org/10.3389/fnut.2022.967985
11. Chen, J., et al. (2020). Metabolite-based genome-wide association study enables dissection of the flavonoid decoration pathway of wheat kernels. Plant Biotechnology Journal, 18, 1722-1735. https://doi.org/10.1111/pbi.13335
12. Chen, W., et al. (2016). Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. Nature Communications, 7, 12767. https://doi.org/10.1038/ncomms12767
13. Crispim, A. C., Kelly, M. J., Guimarães, S. E. F., et al. (2015). Multi-trait GWAS and new candidate genes annotation for growth curve parameters in Brahman cattle. PLOS ONE, 10, e0139906. https://doi.org/10.1371/journal.pone.0139906
14. Dyer, S. C., Austine-Orimoloye, O., Azov, A. G., et al. (2025). Ensembl 2025. Nucleic Acids Research, 53, D948-D957. https://doi.org/10.1093/nar/gkae1071
15. Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. Nature Reviews Genetics, 14, 379-389. https://doi.org/10.1038/nrg3472
16. Fink, T., Tiplady, K., Lopdell, T., et al. (2017). Functional confirmation of PLAG1 as the candidate causative gene underlying major pleiotropic effects on body weight and milk characteristics. Scientific Reports, 7, 44793. https://doi.org/10.1038/srep44793
17. Fu, D., et al. (2019). A study of genomic prediction of 12 important traits in domesticated yak. Animals, 9, 927. https://doi.org/10.3390/ani9110927
18. Ghoreishifar, S. M., et al. (2026). Bridging GWAS to genes: an integrative multi-omics approach using cattle data. BMC Genomics, 27, 171. https://doi.org/10.1186/s12864-026-12525-0
19. Gou, Y., Jing, Y., Wang, Y., et al. (2025). AnimalGWASAtlas: Annotation and prioritization of GWAS loci and quantitative trait loci for animal complex traits. Journal of Biological Chemistry, 301, 108267. https://doi.org/10.1016/j.jbc.2025.108267
20. Gualdrón Duarte, J. L., Yuan, C., Gori, A.-S., Moreira, G. C. M., Takeda, H., Coppieters, W., Charlier, C., Georges, M., & Druet, T. (2023). Sequenced-based GWAS for linear classification traits in Belgian Blue beef cattle reveals new coding variants in genes regulating body size in mammals. Genetics Selection Evolution, 55, 83. https://doi.org/10.1186/s12711-023-00857-4
21. Hozáková, K., Vavrišínová, K., Neirurerová, P., & Bujko, J. (2020). Growth of beef cattle as prediction for meat production: A review. Acta Fytotechnica et Zootechnica, 23, 58-69. https://doi.org/10.15414/afz.2020.23.02.58-69
22. Hu, Z.-L., Park, C. A., & Reecy, J. M. (2022). Bringing the Animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. Nucleic Acids Research, 50, D956-D961. https://doi.org/10.1093/nar/gkab1116
23. Jahuey-Martínez, F. J., Parra-Bracamonte, G. M., Sifuentes-Rincón, A. M., et al. (2016). Genomewide association analysis of growth traits in Charolais beef cattle. Journal of Animal Science, 94, 4570-4582. https://doi.org/10.2527/jas.2016-0359

24. Jia, C., et al. (2020). Identification of genetic loci associated with growth traits at weaning in yak through GWAS. Animal Genetics, 51, 300-305. https://doi.org/10.1111/age.12897
25. Jiang, H., Chai, Z.-X., Cao, H.-W., Zhang, C.-F., Zhu, Y., Zhang, Q., et al. (2022). Genome-wide identification of SNPs associated with body weight in yak. BMC Genomics, 23, 833. https://doi.org/10.1186/s12864-022-09077-4
26. Karjalainen, M. K., et al. (2024). Genome-wide characterization of circulating metabolic biomarkers. Nature, 628, 130-138. https://doi.org/10.1038/s41586-024-07148-y
27. Korkuć, P., Reißmann, M., & Brockmann, G. A. (2025). Genomic insights into growth traits in German Black Pied cattle: A dual-purpose breed at risk. Animal, 19(6), 101540. https://doi.org/10.1016/j.animal.2025.101540
28. Li, J., et al. (2020). Genomic heritability and genome-wide association studies of plasma metabolites in crossbred beef cattle. Frontiers in Genetics, 11, 538600. https://doi.org/10.3389/fgene.2020.538600
29. Lin, Y., et al. (2024). Transcriptomic and metabolomic dissection of skeletal muscle of crossbred Chongming white goats with different meat production performance. BMC Genomics, 25, 443. https://doi.org/10.1186/s12864-024-10304-3
30. Li, Q., et al. (2024). Comparative metabolomics reveals serum metabolite changes in goats during different developmental stages. Scientific Reports, 14, 7291. https://doi.org/10.1038/s41598-024-57803-7
31. Liu, C.-L., Ren, T., Ruan, P.-C., et al. (2025). Genome-wide association integrating a transcriptomic meta-analysis suggests that genes related to fat deposition and muscle development are closely associated with growth in Huaxi cattle. Veterinary Sciences, 12, 109. https://doi.org/10.3390/vetsci12020109
32. Liu, X., Wang, M., Qin, J., Liu, Y., Chai, Z., Peng, W., et al. (2023). Identification of candidate genes associated with yak body size using a genome-wide association study and multiple populations of information. Animals, 13, 1470. https://doi.org/10.3390/ani13091470
33. Majeres, L. E., Dilger, A. C., Shike, D. W., McCann, J. C., & Beever, J. E. (2024). Defining a haplotype encompassing the LCORL-NCAPG locus associated with increased lean growth in beef cattle. Genes, 15, 576. https://doi.org/10.3390/genes15050576
34. Marete, A. G., Guldbrandtsen, B., Lund, M. S., & Fritz, S. (2018). A meta-analysis including pre-selected sequence variants associated with seven traits in three French dairy cattle populations. Frontiers in Genetics, 9, 522. https://doi.org/10.3389/fgene.2018.00522
35. Ma, Z., Chang, Y., Brito, L. F., et al. (2023). Multitrait meta-analyses identify potential candidate genes for growth-related traits in Holstein heifers. Journal of Dairy Science, 106, 9055-9070. https://doi.org/10.3168/jds.2023-23462
36. Mota, L. F. M., Arikawa, L. M., Valente, J. P. S., Fonseca, L. F. S., Mercadante, M. E. Z., Cyrillo, J. N. S. G., et al. (2025). Variable selection strategies for genomic prediction of growth and carcass related traits in experimental Nellore cattle herds under different selection criteria. Scientific Reports, 15, 22266. https://doi.org/10.1038/s41598-025-06949-z
37. Ojo, A. O., Mulim, H. A., Campos, M. A. F., Garcia, A., Retallick-Riley, K., Fonseca, P. A. S., & Oliveira, H. R. (2026). Genome-wide association study of mature cow size traits in American Angus cattle. Mammalian Genome, 37, 53. https://doi.org/10.1007/s00335-026-10226-3
38. Pausch, H., MacLeod, I. M., Fries, R., Emmerling, R., Bowman, P. J., Daetwyler, H. D., & Goddard, M. E. (2017). Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. BMC Genomics, 18, 853. https://doi.org/10.1186/s12864-017-4263-8
39. Ren, J., et al. (2024). Application of GWAS and mGWAS in livestock and poultry breeding. Animals, 14, 2382. https://doi.org/10.3390/ani14162382
40. Rosen, B. D., Bickhart, D. M., Schnabel, R. D., et al. (2020). De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience, 9, giaa021. https://doi.org/10.1093/gigascience/giaa021
41. Sahana, G., Cai, Z., Sanchez, M. P., Bouwman, A. C., & Boichard, D. (2023). Invited review: Good practices in genome-wide association studies to identify candidate sequence variants in dairy cattle. Journal of Dairy Science, 106, 5218-5241. https://doi.org/10.3168/jds.2022-22694
42. Sanchez, M.-P., Tribout, T., Kadri, N. K., et al. (2023). Sequence-based GWAS meta-analyses for beef production traits. Genetics Selection Evolution, 55, 70. https://doi.org/10.1186/s12711-023-00848-5
43. Santana, M. L., Bignardi, A. B., Rodrigues, G. R. D., Cyrillo, J. N. S. G., Brito, L. F., & Mercadante, M. E. Z. (2025). Genotype-by-environment interaction for yearling weight of Nellore cattle in pasture and feedlot conditions using a double genomic reaction norm model. Journal of Animal Science, 103, skaf169. https://doi.org/10.1093/jas/skaf169
44. Seabury, C. M., Oldeschulte, D. L., Saatchi, M., et al. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. BMC Genomics, 18, 386. https://doi.org/10.1186/s12864-017-3754-y
45. Sharma, A., Lee, J. S., Dang, C. G., et al. (2015). Stories and challenges of genome-wide association studies in livestock: A review. Asian-Australasian Journal of Animal Sciences, 28, 1371-1379. https://doi.org/10.5713/ajas.14.0715

46. Shin, S.-Y., et al. (2014). An atlas of genetic influences on human blood metabolites. Nature Genetics, 46, 543-550. https://doi.org/10.1038/ng.2982
47. Smith, J. L., Wilson, M. L., Nilson, S. M., et al. (2022). Genome-wide association and genotype by environment interactions for growth traits in U.S. Red Angus cattle. BMC Genomics, 23, 517. https://doi.org/10.1186/s12864-022-08667-6
48. Snelling, W. M., Allan, M. F., Keele, J. W., et al. (2010). Genome-wide association study of growth in crossbred beef cattle. Journal of Animal Science, 88, 837-848. https://doi.org/10.2527/jas.2009-2257
49. Takasuga, A. (2016). PLAG1 and NCAPG-LCORL in livestock. Animal Science Journal, 87, 159-167. https://doi.org/10.1111/asj.12417
50. Terakado, A. P. N., Costa, R. B., de Camargo, G. M. F., et al. (2018). Genome-wide association study for growth traits in Nelore cattle. Animal, 12, 1358-1362. https://doi.org/10.1017/S1751731117003068
51. Tian, J., et al. (2023). Serum metabolic profile and metabolome genome-wide association study in chicken. Journal of Animal Science and Biotechnology, 14, 69. https://doi.org/10.1186/s40104-023-00868-7
52. Vanvanhossou, S. F. U., Scheper, C., Dossa, L. H., Yin, T., Brügemann, K., & König, S. (2020). A multi-breed GWAS for morphometric traits in four Beninese indigenous cattle breeds reveals loci associated with conformation, carcass and adaptive traits. BMC Genomics, 21, 783. https://doi.org/10.1186/s12864-020-07170-0
53. Wang, J., Li, X., Peng, W., Zhong, J., & Jiang, M. (2022a). Genome-wide association study of body weight trait in yaks. Animals, 12, 1855. https://doi.org/10.3390/ani12141855
54. Wang, L., Raza, S. H. A., Gui, L., et al. (2022b). Associations between UASMS2 polymorphism in leptin gene and growth, carcass and meat quality traits of cattle: a meta-analysis. Animal Biotechnology, 33, 279-288. https://doi.org/10.1080/10495398.2020.1805327
55. Wang, X., & Kadarmideen, H. N. (2020). Metabolite genome-wide association study (mGWAS) and gene-metabolite interaction network analysis reveal potential biomarkers for feed efficiency in pigs. Metabolites, 10, 201. https://doi.org/10.3390/metabo10050201
56. Wang, Z., Ma, X., Hu, G., Jing, J., La, Y., et al. (2026). Genome-Wide Association Study and Genomic Selection for Average Daily Gain in Ashidan Yak. Animals, 16, 2452. https://doi.org/10.3390/ani16162452
57. Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 26, 2190-2191. https://doi.org/10.1093/bioinformatics/btq340
58. Xiang, R., van den Berg, I., MacLeod, I. M., Daetwyler, H. D., & Goddard, M. E. (2020). Effect direction meta-analysis of GWAS identifies extreme, prevalent and shared pleiotropy in a large mammal. Communications Biology, 3, 88. https://doi.org/10.1038/s42003-020-0823-6
59. Yang, X., et al. (2026). Integrated genomics and metabolomics analysis revealed the metabolic mechanisms by which genetic variations affect the growth performance of local chickens. Poultry Science, 105, 106495. https://doi.org/10.1016/j.psj.2026.106495
60. Zepeda-Batista, J. L., Núñez-Domínguez, R., Ramírez-Valverde, R., et al. (2021). Discovering of genomic variations associated to growth traits by GWAS in Braunvieh cattle. Genes, 12, 1666. https://doi.org/10.3390/genes12111666
61. Zhang, F., Wang, C., Shangguan, A., et al. (2025a). Genome-wide association studies of growth and carcass traits in Charolais cattle based on high-coverage whole-genome resequencing. International Journal of Molecular Sciences, 26, 11411. https://doi.org/10.3390/ijms262311411
62. Zhang, M., Huang, C., Bao, Q., Wang, T., Ma, X., Meng, G., et al. (2025b). Whole-genome resequencing revealed the genetic diversity and body weight-related genes of Gannan yak. Frontiers in Veterinary Science, 12, 1540523. https://doi.org/10.3389/fvets.2025.1540523
63. Zhang, X., Chu, Q., Guo, G., Dong, G., Li, X., Zhang, Q., et al. (2017). Genome-wide association studies identified multiple genetic loci for body size at four growth stages in Chinese Holstein cattle. PLOS ONE, 12, e0175971. https://doi.org/10.1371/journal.pone.0175971
64. Zhao, J., Li, Z., Liu, X., Liu, Y., Yue, B., Wang, H., et al. (2025). Large-scale genome-wide association analysis reveals candidate genes in yak body size traits. BMC Genomics, 26, 844. https://doi.org/10.1186/s12864-025-12017-7
65. Zhuang, Z., Xu, L., Yang, J., et al. (2020). Weighted single-step genome-wide association study for growth traits in Chinese Simmental beef cattle. Genes, 11, 189. https://doi.org/10.3390/genes11020189
66. Zhu, G., et al. (2018). Rewiring of the fruit metabolome in tomato breeding. Cell, 172, 249-261.e12. https://doi.org/10.1016/j.cell.2017.12.019
67.